Year : 2022

Forensic Examination of Dynamic Signatures

Linden Jacques

UNIL | Université de Lausanne

FACULTÉ DE DROIT, DES SCIENCES CRIMINELLES ET DE L'ADMINISTRATION PUBLIQUE

ECOLE DES SCIENCES CRIMINELLES

Forensic Examination of Dynamic Signatures

THÈSE DE DOCTORAT

présentée à la

Faculté de droit, des sciences criminelles et de l'administration publique de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès science

par

Jacques Linden

Directeur de thèse

Prof. Franco Taroni

Jury

Prof. Christophe Champod
Dr. Carolyne Bird
Dr. Tomasz Dziedzic
Prof. Silvia Bozza
Dr. Raymond Marquis

LAUSANNE
2022

Unil
UNIL | Université de Lausanne
Ecole des sciences criminelles
bâtiment Batochime
CH-1015 Lausanne

IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de M. Jacques Linden, candidat au doctorat en science forensique, intitulée

« Forensic Examination of Dynamic signatures »

Le Président du Jury

Prof. Christophe Champod

Lausanne, le 24 juin 2022

Résumé

Durant les dernières décennies, des changements profonds, tels l'internet et la numérisation ont impacté la vie des citoyens profondément. La digitalisation de la société a provoqué une évolution rapide du concept de signature. La signature électronique, le fruit de ce développement, est un équivalent numérique de la signature manuscrite. La signature électronique manuscrite est un enregistrement de signatures, comprenant des données biométriques et kinématiques, telles la vitesse. Vu la simplicité de mise en place, le faible coût, la sécurité et le gain de temps occasionné, la signature dite dynamique connaît actuellement un grand essor. Les changements profonds de la signature manuscrite obligent les forensiciens à faire face à de nouveaux défis et problématiques. Ainsi, les litiges impliquant ces signatures sont inévitables et doivent être pris en charge. La nature logique et complexe des signatures électroniques manuscrites requiert des avancées méthodologiques pour garantir une examination scientifique et des résultats valides. Vu la nouveauté de ces signatures et la profondeur des changements, l'état des recherches est encore lacunaire. La complexité des données de signature impose l'intégration de nouvelles méthodes de traitement de données, ainsi que de banques de données de comparaison. Additionellement, l'interprétation de ces données est fondamentale dans le processus forensique et requiert l'usage de probabilités et de statistiques pour exprimer les résultats de façon transparente et scientifique. Cette thèse propose une approche théorique et statistique à la question des signatures imitées. Son but est de compléter la méthodologie d'examen de signatures dynamiques. D'abord, l'état de l'art des signatures dynamiques est réalisé. Ensuite, des méthodes techniques et statististiques pour l'analyse de signatures dynamiques sont proposés, en liant les connaissances des domaines de biométrie, de neuroscience et de science forensique. Le mémoire approfondit les connaissances actuelles à travers d'analyses exploratoires et de statistiques réalisés sur des données réelles, collectées dans le cadre du projet. Au coeur de la thèse est un modèle statistique Bayésien pour l'interprétation des données de signatures électroniques manuscrites. Ce dernier a été construit spécifiquement pour une application au problème de détection des signatures imitées. Ce modèle statistique et ses applications permettent d'étudier et de mieux comprendre le caractère dynamique des signatures. La validation du modèle avec des données de signatures réelles et d'imitations a permis de démontrer une performance et reproductibilité encourageante. Le modèle, dans une logique de transparence et de communication, permet à l'examinateur d'obtenir une évaluation quantifiée de la valeur de la preuve. Il permet ainsi la communication, en toute transparence, de résultats scientifiques justifiables appuyés sur des données empiriques.

Summary

In recent years, our society has rapidly evolved and been digitalized. These deep-rooted changes affect every citizen's life, be it through government and commerce, communication or contracts. With the increasing number of important electronic transactions (e.g. e-government, e-commerce), the need for numerical authentication procedures has emerged. One of these solutions is electronic signatures, a numerical equivalent to its analogue homonym. Handwritten signatures have also been affected by the ongoing change. Dynamic signatures, complex recordings of signature image and movements, as well as their biometric attributes have changed signature examination. These electronic solutions have many advantages for businesses, such as their simple set-up, time- and cost-effectiveness, simple workflow, inherent security and ecological sustainability. As a result, dynamic signatures are currently enjoying an increase in popularity. Just as 'physical' signatures are, dynamic signatures are important contractual elements. Conflict, and therefore, challenges to their authenticity, are inevitable. It falls to forensic experts to deal with the novel challenges related to dynamic signature authenticity. The complex handwritten electronic signature data requires a revision of existing methods in order to produce valid scientific evidence. Dynamic signatures are relatively novel and the state of their research is still in its beginning stages. The inherent data complexity calls for the integration of novel techniques, databases as well as adequate statistical models. A case assessment and interpretation method for signature forgery cases requires both statistics and probability to express results scientifically. Such models are currently missing from the forensic and dynamic signature literature. This thesis proffers a logical and theoretical approach to the problem of forged signatures. It aims to do so by providing a deeper understanding of dynamic signatures, as well as proposing an extension to existing examination methodologies. To this effect, it features a state of the art on the forensically relevant knowledge, analysis methods and statistical tools, by uniting elements from biometrics, neuroscience and forensic science literature. The information relevant for an update of signature examination methodology is structured into four research propositions. At the heart of the thesis is a statistical Bayesian model constructed specifically for questioned and forged (i.e. simulated) signature scenarios. This model and its applications enabled the study and comprehension of the dynamic and kinematic features of signatures. Further, the validation of the model on known-source and large-scale data managed to produce encouraging results for evidence interpretation. The proposed model helps the forensic examiners quantify the value of disputed signature evidence. It may be used for the transparent communication of justifiable assessments of evidential value backed up by empirical data.

Acknowledgments

This thesis was written at the Ecole des Sciences Criminelles, at the Faculty of Law, Criminal Sciences and Public Administration of the University of Lausanne. The project was supervised by Prof. Franco Taroni.

During my years at the University of Lausanne, many people have helped and contributed to this thesis in innumerable ways. I would like to express my heartfelt gratitude to every one of these people, even if I am unable to name all of them.

First and foremost, I wish to thank my *volunteers*. Three people, who shall not be named for confidentiality, gave their valuable time freely and frequently to provide the necessary data. Another special thank you goes out to the volunteer signers and forgers. There were 57 forgers and 23 signers, who've all contributed to this project tremendously. Without you, this thesis would not have been possible.

- *Williams David Mazzella*, who started what would eventually become this thesis.
- *Raymond Marquis* to have started this project, as well as for your advice, curiosity and enthusiasm.
- *Silvia Bozza*, for her never ending patience.
- *Franco Taroni*, for his understanding and his positivity.
- *Nick Mettyear* and *Uwe Wuerzberger* at Wacom Europe GmbH, who have supported this project by providing hardware, information and assistance.
- *Lorenzo Gaborini*, *Marco De Donno* and *Julien Furrer*, for saving me approximately 7 years of computations!
- *Matthieu Di Floriano*, *Florent Koenig*, *Cédric Grisoni*, *Denis Werner*, *Ilaria De March* and *Céline Burnier*, for their friendship and support.
- *Laurane Favre*, for being there.
- *My family*, for all their love and support.

Contents

Introduction

Research Context

Society is an aggregation of people living in an ordered community. In a society, many individuals coexist and interact with each other in a variety of ways. Some parts of these interactions are important enough to warrant being documented, in order to prove both the transaction's terms and the agreement of both parties. These interactions are very diverse and concern all parts of our daily lives, be it through commerce and purchases, loans, insurances, promises, liability and many others. Transactions such as these can have important consequences on the fortune, reputation or even the liberty of the people involved. In order to prevent abuse and fraud, especially important transactions are regulated through the law and made in a controllable and verifiable fashion, very often a contract or an official document. At first, informal and oral agreements had been the predominant form for these interactions, but written (and typed) contracts have since become the more popular alternatives. In order for documents to be valid and have value as proof, the signatories need to assert their identity and intent on the document. Various means have been used for this purpose, but doubtlessly the most long-lived, widespread and accepted one has been the handwritten signature. Handwritten signatures were first used by the Jewish (2^{nd} century AD) and Muslim Communities (7^{th} century AD), although they weren't widespread until the 16^{th} and 17^{th} century, when literacy was on the rise [520]. Social life and government have changed radically since then, especially during the last decades with the beginning of the age of information. An increase in administrative charges, formalism and contracts has been observed in modern society, leading to the creation of unimaginable quantities of contracts, signatures and data. These changes, along with the progressing digitization of our lives have been the heralds of the age of *big data*. The most extensive of these developments are due to computers, digital data and the internet. As a result, most of our transactions are either conducted or at least stored in digital form. Naturally, signatures have had to adapt to these new media and the dematerialization of contracts. The developments in electronic transactions have led to the dawn of electronic signatures. Within this new type of dematerialized signatures, the handwritten signature has shown that it is still relevant and trustworthy. The electronic handwritten signature is transformed from analogue signal to digital file by means of an electronic device. It can exist under two forms, called *static* and *dynamic* signatures, which are digitally recorded files. Dynamic signatures in particular present a fundamental difference to 'wet'[1] signatures. They include both timing and kinematic data, measured continuously while the signature is being made, all of which happens on a digitizing device. A variety of devices, ranging from smartphones, computers, gloves, glasses and cameras, to dedicated signature pads can be used to record dynamic signatures. Dynamic signatures have changed recording medium and form, from paper to file and from image to signal. The possibilities of their recording and usage continually evolve, but they already pose a number of challenges to the forensic scientist, such as the diversity in measurements or the data quality. Their relative novelty are the cause for the current absence of adequate methods and tools for their treatment, which is one of the major

[1] The term 'wet signature' has recently been used to describe signatures made on paper with ink. The ink is liquid and may be water-based, hence the 'wet' tag.

motivations for this project.

Dynamic signatures remain first and foremost handwritten products and therefore the project is at a crossroads between forensic science, signature examination, biometrics and statistics. The project is deeply rooted in forensic handwriting and forensic signature examination in particular. Forensic handwriting and signature examination are disciplines concerned with the question of authorship of a handwritten item. They are some of the first 'purely' forensic disciplines, created out of the court of law's necessity to determine the authenticity of important documents. As such, these disciplines and other pattern matching fields have often been criticized as pseudo-sciences for their lack of basis in natural sciences and subjective methods. Especially forensic handwriting examination has received much criticism. It has been exemplified for a lack of a scientific basis, method, knowledge of precision and performance, as well as empirical proof for claims made by examiners. As a reaction, forensic scientists have been working hard to validate methods and provide empirical evidence proving the worth and quality of their work. These developments have led to a rigorous framework for examination and reasoning based on evidence, which will be further described in the following paragraph. In forensic signature examination, two (or more) parties are in disagreement about the authenticity of a document and create a situation where the 'true' event is uncertain. When conflicting propositions of the 'truth' involving technical elements are presented, the court turns to forensic experts for technical counsel. In signature examination, this counselling amounts to providing evidence for either authenticity or spuriousness of the signature. The forensic scientist assists the court in building its opinion and taking its decision. The forensic examiner's main task involves two important aspects: source and originality[2]. The major challenge in forensic science is to deal with the uncertainty inherent to these assessments. The forensic scientist has to methodologically collect data, analyze samples and evaluate the probability of the evidence under each proposition. As evidence one considers the collection of characteristics from a trace object and a reference object, for comparison. He uses several tools, computerized, statistical and probabilistic, in order to coherently and scientifically deal with the uncertainty and the 'true' event. Recently, the legal and forensic communities have called for more transparency, reproducibility and rigour in these processes. They especially denounced the lack of empirical foundation and validity as a scientific discipline, as well as its inherently subjective[3] methods. Naturally, forensic scientists have reacted to these criticisms and addressed many of these concerns. Over the years, movement and neuroscience have provided a much needed theoretical basis for the generation of handwriting, while forensic and biometrics scholars have been building a foundation for pattern matching. Nowadays, there is sufficient empirical proof for the validity, reliability and thus the legitimacy of signature examination in the courtroom. Nevertheless, forensic signature examination is a very complex discipline, because it studies intentional and intricate human behavior. As a consequence, signatures often present high personal variation, and allegations often include intentional forgery and disguise. As previously mentioned, forensic handwriting examiners (FHE) have since mobilized to research and address many of the points made by the critics. As a result, the field has continually evolved and modernized through computer approaches, large-scale data collection and proficiency testing. In the process, researchers have found and articulated the necessary foundations, collected empirical data and validated claims. Even though the field has much progressed, there still remain many issues to research and address (see Izenmass' article in the Handbook of forensic statistics [29]). Most critics agree that the way forward for forensic science is the integration of statistical methods and rigorous procedures, as well as validation and transparency. Most forensic pattern matching fields have been slow in integrating statistics and inference into their methodologies. This reluctance stems from the practical difficulty of tailoring statistical approaches to the complex features used in these fields. Forensic handwriting examination in particular has been challenged vigorously. Rigorous statistical approaches are still rare in signature examination. By extension, dynamic signature examination currently

[2]Originality as in being an original, not a copy or reproduction

[3]Here, the term 'subjective' is synonymous with 'arbitrary'

suffers from the same failings. FHEs are faced with a multitude of decisional problems, such as sample collection, sampling strategy, population studies, feature selection, as well as difficult case assessment and interpretation. These problems are the main reasons for starting this project and will be examined in detail in the following chapters. The main interest has been the development of a probabilistic model to assign a value for data extracted from dynamic signatures.

Dynamic signatures are a novel addition to the already vast field of handwriting and signature examination. They are game-changers, as they are fundamentally different from paper-based (*wet*) signatures. They are entirely recorded on electronic sensors, which present technical and methodological challenges. Further, they are a complex type of logical data, with many variables and novel features yet to be explored. Dynamic signatures are a great novelty to signature examination and require exploration and adaptation. Nevertheless, they are closely related to their intellectual parent, wet signatures, and involve the same movement and neurological processes. Therefore, they present a challenge, but also offer an opportunity to modernize methods and gather relevant data on signatures. They are an ideal subject for study, as they are perfectly measurable and allow for many data analysis techniques and statistical analysis. The findings can however also largely be transposed to regular signature examination. This study will therefore focus on dynamic signatures, as they are novel and require exploration and rigorous examination and evaluation methodology. This thesis serves as starting point for the development of case assessment and interpretation in dynamic signature examination. To the best of our knowledge, no studies propose a data-driven statistical framework for evidence evaluation. Research in dynamic signature examination is still in its infancy. This thesis explores the dynamic signature problem on a theoretical and practical level and lays down a foundation for coherent source-level assessments. The goals are to explore dynamic signatures as a novel form of evidence, as well as describe their natural variation and kinematics. The thesis proposes actionable information through exploratory data analysis, as well as practical knowledge on dynamic signatures and sample collection. Additionally, it pioneers a probabilistic approach for their examination and evaluation, complementing existing methodologies. Its ultimate objective is to support examiners by providing them with methodology and tools to exploit dynamic signatures to their fullest potential, while integrating statistical modelling and empirical data to increase validity and credibility of their expertise.

Objectives

Dynamic signature examination is an emerging field of research, with many open subjects. As advanced by Flynn et al. [392] forensic scientists' current methods are applicable to dynamic signatures of sufficient quality. A review of biometric, forensic, medical and neuroscience literature has revealed a lack of a probabilistic framework as well as a transparent and reproducible examination methodology. Current methods for data evaluation do not make use of all of the data available in dynamic signatures, especially quantified features and measurements. The existing methodology should be extended by integrating this useful and often neglected information into the process.

The objectives of the thesis are two-fold: provide useable knowledge to examiners and propose a methodological and statistical framework for signature examination. The underlying goal is to propose and validate a statistical method within the Bayesian framework, that supports and complements methods currently used by forensic examiners for signature examination. The development of the methodological and statistical approach to case assessment and interpretation in dynamic signature examination is central to this document. Not only does it provide a strong statistical basis and justification for evidence strength, but it also respects the roles of participants in the justice system and guarantees scientific integrity and transparency while reporting.

One of the objectives is to provide practical and actionable information to forensic examiners, to ensure they are aware of the complexities of dynamic signatures. Dynamic signatures are novel to forensic examiners and pose a specific set of challenges due to digitalization, hardware and software components. They also include information on stroke kinematics and timing, which was previously unavailable for examination and data acquisition. Examiners should use sufficient information to allow for examination of dynamic signatures using a statistical approach. The primary objective of the thesis is to extend the examination methodology for dynamic signatures to incorporate the much needed statistical basis through the statistical evidence evaluation framework. Dynamic signature examination is concerned with providing evidence supporting authorship. Technically speaking, forensic examiners support 'source' level inferences, specifically verifying who made a questioned signature. A presumed source is assumed to have made the questioned signature, with the alternative being that someone else has made the signature. Forensic scientists need to consider multiple types of forgery and disguise behaviors, potentially obscuring the true source determinations. A probabilistic model, supporting inference of source, may help in dealing with these complex determinations. In this thesis, the groundwork for a signature evidence evaluation model is laid down, starting with the development of an extensible model, applicable to cases involving the suspicion of a signature forgery.

In the thesis, four main propositions will be addressed:

Proposition 1 Dynamic signature examination is novel for forensic scientists in terms of both technology and data. Dynamic signatures are complex constructs involving multiple research fields (e.g.

computer science, engineering, biometrics, cryptography). The technology used to record them is capable of measuring dynamic and kinematic signature data, which was previously inaccessible to forensic examiners. These developments have made dynamic signatures into complex multivariate data structures. The new developments and fundamental changes to dynamic signatures warrant a revision and extension of our knowledge on signature dynamics and statistical, data-driven evaluation of signature evidence. It is important to raise awareness, provide new insights and encourage research in the field.

Proposition 2 The analysis of the best evidence available is essential both for reliability and admissibility of evidence. Dynamic signatures are files containing large amounts of complex multivariate data. Traditional pattern matching techniques largely ignore the dynamic and kinematic measurements made during the recording of the signature. Previous methodology does not permit the use of dynamic signature to their fullest potential and therefore does not constitute the 'best available evidence'. A probabilistic evidence evaluation methodology for dynamic signatures provides a framework for the analysis of this additional information and may support examiners in quantifying evidential value, as well as in communicating reliable and justifiable conclusions. Its adaptation to dynamic signatures is an essential step to guarantee coherence and admissibility.

Proposition 3 Signature examination requires a different probabilistic approach than other forensic fields, like fingerprint or DNA evidence. The identity verification scheme is different from identification tasks. Further, simulated signatures are a type of mimicry of a 'target' signature. Dealing with these forms of 'impersonation' requires a probabilistic model tailored to the signature problem and its specific assumptions. A novel model for source-level signature evaluation is necessary to avoid misapprehension of the problem and avert erroneous interpretation of evidence.

Proposition 4 Contemporaneity is an important factor in forensic handwriting examination. Dynamic features are different from their static counterparts in terms of variation and stability in time. There is a need for long-term studies of dynamic signature variation to define a period of contemporaneity, as well as for validation of a probabilistic model in regard of template age.

Each of the propositions contributes to the exploration of dynamic signatures and the implementation of a data-driven evidence evaluation model for signature examination. Results found during our experiments and supporting the propositions are described and discussed in detail. Through the four proposed propositions, knowledge on both the fundamental aspects and the analysis of dynamic signatures is deepened. They address the theoretical foundation of forgery detection, its application to statistical model testing as well as the validation of the approach. Additionally, these pillars will form the basis and validation for statistical procedures in dynamic signature examination. They establish the necessary statistical foundation for signature source inference, providing legitimacy and credibility for dynamic signature evidence.

Originality of the Research

The thesis is original in four major aspects. A first point is the novelty of the subject area for forensic science. As a matter of fact, few monographs and scientific publications deal with the problem of forensic examination of dynamic signatures. The current state of the art shows a lack of methodological and fundamental research into the subject of dynamic signatures, which this research provides. A second point concerns the approach of the thesis. The research project innovates in the way the subject is approached, by including elements from data science, computer science, biometrics and forensic science. In this capacity, the thesis provides both an exploration of signature data, as well as a solid framework for inference and its use in forensic signature examination. In particular, it adapts statistical frameworks for use in forgery and signature examination. Third, the thesis features an original and specifically collected dataset. This data was collected on a state-of-the-art signature tablet with the help of over 60 volunteers. Further, all data treatment and inference-related code is original and freely available. Finally, the thesis adapts multivariate statistics for use on dynamic signatures. The evidence evaluation framework is described in detail for signature examination, as well as proposes a revised statistical model for multivariate data. It also features several original datasets, based on real signature data acquired in frequent sessions over an 18-month period, as well as skilled forgeries of the studied signatures. The thesis revises the evaluative methods used in forensic signature examination and extends them to incorporate empirical data into assessments.

The present research develops statistical components relating to dynamic signature examination. A revised approach, accounting for 'imitation' or forgery-type behavior is integrated into a general statistical framework and multivariate statistical models for source-level inference. This approach is, to the best of our knowledge, novel in forensic science and presents opportunities for application in other forensic domains. The method uses case-specific data, as well as collected background databases to elicit prior knowledge from empirical data, as well as enhance transparency, reproducibility and validity of the inference process. It further fits the specificity of signature examination by adapting the statistical models to forgery behavior from a theoretical point of view. The underlying dependencies incorporated into the novel approach can be used in other forensic fields and purposes, such dealing with masks and disguises in facial recognition, imitation in speaker recognition, spoofed fingerprints or other types of mimicry, for example.

Thesis Outline

This thesis is structured into three main parts: the first introduces theory and concepts for dynamic signatures and source-level inference, the second highlights the benefits of the application of exploratory data analysis and statistical frameworks to dynamic signature examination, and the third consists of the detailed original research articles illustrating the results of applying the case assessment and interpretation model to casework through computer simulations.

Part I: Theory and concepts summarizes the knowledge and concepts underlying the dynamic signatures, signature examination, probabilistic methodology. It contains four chapters:

- **Chapter 1: Inference and Reporting in Forensic Science** introduces the role of forensic science in the judicial process, its case assessment and interpretation model, as well as probabilistic inference and evaluative reporting.
- **Chapter 2: Handwritten Electronic Signatures** introduces the underlying legal and technical notions, as well as the development of dynamic signatures. It defines the technical recording, as well as the form of dynamic signatures.
- **Chapter 3: The Bayesian Framework and Bayesian Statistics** introduces the Bayesian framework and its many applications in forensic science, in particular Bayesian statistics for logical and coherent inference.
- **Chapter 4: Dynamic Signatures in Forensic Science: A State-of-the-Art** introduces the forensic examination of signatures and important advances in dynamic signature research and methodology.

Part II: Methodology, Experiments and Results explains the acquired data, experiments and major contributions of this thesis to the subject. It contains five chapters:

- **Chapter 5: Data Acquisition, Sampling and Databases** details the data acquired for the validation trials and experiments.
- **Chapter 6: Evaluation of Signature Evidence** briefly explains the applied multivariate statistics model and the Bayes Factor concept.
- **Chapter 7: Research Propositions** explains the main hypothesis examined in this thesis, which aims to improve our understanding and methods of dynamic signature examination.
- **Chapter 8: Results and Discussion** summarizes the major advances and conclusions made about our propositions.
- **Chapter 9: Conclusions** summarily describes current advances and limitations, as well as future avenues of research.

Part III: Original Research Articles contains several peer-reviewed publications. These explain the development, validation and application of the proposed statistical model and methodology. Its five chapters each contain a separate research paper, published in a peer reviewed journal, or in the process of being published.

Chapter 10: Dynamic signatures: A review of dynamic feature variation and forensic methodology. contains a thorough review of relevant forensic, methodological and technical literature on the subject of dynamic signatures, as well as time and kinematic signature features. *Published.*

Chapter 11: Bayesian multivariate models for case assessment in dynamic signature cases explains the statistical model, its theoretical foundation and the empirical data necessary to calculate Bayes Factors for simulated signatures. *Published.*

Chapter 12: Bayesian evaluation of dynamic signatures in operational conditions shows a large-scale study on feature selection and combining evidence. It also presents important information on statistical assumptions and variable selection for forensic examiners. *Published.*

Chapter 13: The influence of time on dynamic signatures: An exploratory data analysis is an exploratory study of dynamic signatures and the variation of dynamic features over an 18-month period. Several exploratory data analysis techniques are used to establish distributions, trends and multivariate visualizations of variation and change over time. *Unsubmitted.*

Chapter 14: The impact of time on probabilistic case assessment and interpretation of dynamic signatures is a sensitivity study on the time and template ageing aspects involved with forensic cases and dynamic signature analysis. It highlights the statistical model's (and BF's) sensitivity to using 'dated' signatures as reference materials. *Unsubmitted.*

Part I

Theory & Concepts

1 Inference and Reporting in Forensic Science

Forensic signature examination is a field focused on providing evidence regarding the source of a signature in a court of law. It is part of the *evaluative* disciplines of forensic science, as its core objective is to evaluate the support given to a hypothesis provided by the parties in the trial given the material evidence. Evaluative[1] forensic sciences have undergone vast changes in the past decades. Their foundation, as well as their way of expressing and communicating the results of the assessments has been sorely tested. As a result, profound changes in the forensic approach to inference and reporting have come to pass. The most important of these changes concern scientific testimony, transparency and limitations of conclusions and finally reporting. A shift in the general paradigm and approaches adopted all over the world can be noted, which increasingly put an accent on the use of probabilityin forensic science. In order to understand the necessity for the integration of probabilistic approaches and the adoption of a probabilistic evidence evaluation model, a look at the history and criticism aimed at forensic science, and forensic handwriting examination in particular, is proposed in this chapter. A way forward, through the Bayesian (or logical) approach, providing clarity and transparency, is presented. The Bayesian approach is one of the main theoretical pillars this thesis is built upon, to provide the statistical framework necessary for coherent inference.

1.1 A Change in Forensic Science

In the past, forensic science (initially also often called 'criminalistics') had been defined as the application of scientific principles to help support courts confronted with technical evidence. Forensic science has thereafter acquired the image of a loose collection of 'sciences' applied for judicial purposes. Its definition, scope and nature have long been vague and deemed unimportant by most. As a result, many unproven or even unrelated disciplines have appropriated a 'forensic' tag and tried to legitimize in this fashion, without actually applying the essential principles of forensic science. Forensic science has been in dire need of a solid foundation, to define itself, its scope and objectives clearly. Several authors have studied forensic science and its paradigms [87, 102, 158, 305–308, 310, 357, 478] to unite the splintered collection of disciplines into a single field. Forensic science has gained solid bases through theoretical foundations and universal principles, with inference (see figure 1.1), intelligence, public security and justice, as well as communication being its guiding principles. Forensic science historically had only two purposes: personal (or object) identification and investigative support. The primary goal of forensic science was to determine

[1] Concerned with the evaluation of evidence, often in hope of determining the source or a common source. As opposed to investigative forensic science, which focuses on developing new hypothesis and leads from evidence.

the source of trace evidence, such as a fingerprint, and link it to its originator, the finger of the suspect. The act of linking a piece of evidence to a potential source, activity or crime [98] and quantifying the value of this link, has been called the 'evaluative' role of forensic science [153, 357, 478]. The second goal concerned the use of evidence to locate and determine suspects. Forensic scientists have referred to this application as the 'investigative' role of the forensic scientist. The definition of Forensic science has since been extended to incorporate forensic intelligence, policing and strategy [1, 476–478, 491]. Albeit the adoption of an extended conception of the discipline, the main concern of most of its disciplines remain 'identification'[2] and 'authentication'[3]. Many of the well-known forensic fields are 'evaluative' disciplines and aim at providing information on a source (person or object) or activity. Signature examination is no exception to this, as its objective is to support the verification of the source's identity and authenticity of a signature.

The 'evaluative' fields have undergone deep-rooted changes over the past decades in three major aspects : the inference[4], the decision of concluding[5] or the qualification of results, as well as reporting (how and what is transmitted). Philosophically-minded scholars have grouped particular ideologies in forensic science into paradigms. These paradigms show major differences in their internal workings, but they also affect forensic science's public image. The recent paradigm shift in forensic science is a proof of evolution and improvement in its practice. The following paragraphs briefly described the old and new paradigms of evaluative forensic science, in order to understand why changes in methodology and reporting are essential for signature examination.

Signature examination shares many parallels to other classic forensic disciplines such as fingerprint examination, genetics or firearm and tool mark examination. In the past, all of these disciplines were considered *identification sciences*. At first, identification was to define the purpose of the field, the task to identify (or verify, for signatures) a potential source. Identification has, however, gradually become a mindset and way of reasoning and reporting. The word identification has been popularized by criminal fiction (both in literature and audio-visual form) and transformed into a symbol of establishing identity and individuality. Hence, identification is connotated with certainty and confidence in the presented conclusion. These considerations represent well the first paradigm of forensic science, called the *identification paradigm*.

Originally, *identification* is a synonym of *classification*, a process in which we associate[6] an object with a class of objects with similar properties[7]. When discussing identification, scholars actually often tried referring to *Individualization*. [77, 78, 276, 298, 299, 318]. In the forensic individualization process, a piece of evidence is associated to a single entity or activity as its source, eliminating all other possibilities. In order to individualize an entity, all other entities (in general, all other people on the planet) need to be excluded [78]. Forensic scientists need to base this exclusion on the physical characteristics of the entity. The underlying assumption is that entities sharing the same properties (qualitative identity) must originate from the same source, with no other source sharing these properties. Legal scholars [502] have referred to this as the 'discrete uniqueness' paradigm. This assumption was historically made by forensic scientists for simplicity, but forensic scientists often lacked empirical evidence to sustain it. It inherently implies that as soon as there is an analytical correspondence between features of two objects, the method

[2]Identification, or more correctly 'individualization' consists in determining the identity of a person or object from a population/set. Identification deals with associating a person or object to a class. Individualization is a type of Identification where the classes only contain one entity. Identification is often colloquially used when individualization is actually meant.

[3]Authentication is concerned with verifying an identity claim.

[4]How we reason and obtain conclusions

[5]The examiner may choose not to report any conclusion, if the context or data does not permit robust evaluation. In addition, the examiner must decide how to conclude and how to deal with 'inconclusive' cases.

[6]Association is a decision to categorize a sample.

[7]Qualitative identity, as proposed by Kwan [318].

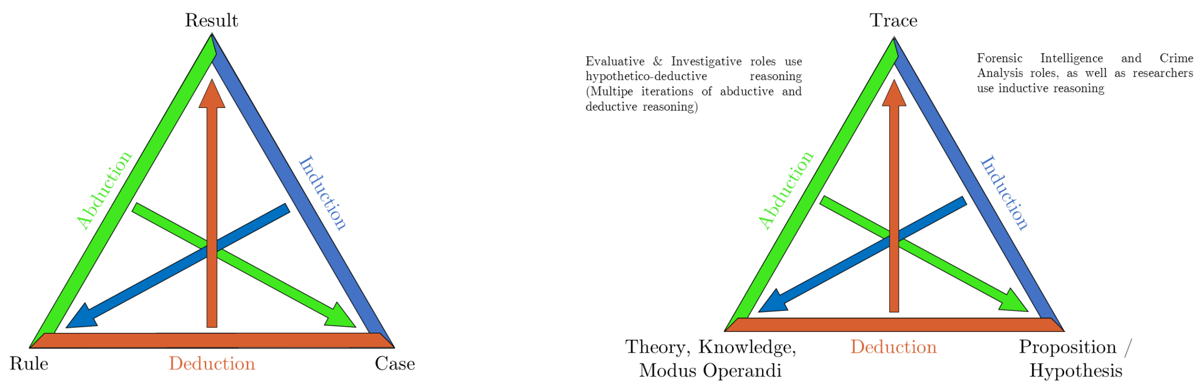


Figure 1.1: C.S. Peirce's description of inference types
(a) General Peirce Triangle (b) Forensic Inference Triangle

permits individualization to the exclusion of all other sources. No further steps, like a decision for instance, are required, as per assumption matching properties (features) imply 'identity'. The empirical observation 'dominates' all other concerns, given that the evidence type is perfect and 'discretely unique'. These deterministic conclusions, without any room for doubt are called 'identifications'. The forensic examiner concluded either to definite identification or, slightly tempered, to 'a reasonable degree of scientific certainty'. Identification conclusions were (and still are to some degree) frequent and highly desired by courts and juries. They inspire confidence and are easy to understand, necessitating no knowledge of the process, but faith in the examiner [500]. Signature examination is used to verify identity claims, which is a slightly different process, but builds upon these very same precepts. Much research has gone into proving the unicity and discriminative power of handwriting and signatures [296, 297, 525, 528]. The identification paradigm foregoes more practical concerns, for example that the methods and features used to distinguish, as well as the quality of the evidence, may cause limitations. The latter render complete individualization impossible, even if the 'discrete uniqueness' assumption is true. Methods used in science are imperfect and incapable of complete separation. Their capacity to differentiate can be measured by the discriminatory power, a 'quality measure' of the methods used [516, 521]. In order to perform well, a method needs to sufficiently and adequately describe the evidence and it needs to be capable of separating based on these descriptors (i.e. features). On another note, progressive thinkers have recognized that unlike deductive reasoning, abductive reasoning[8] [464, 533, 623] does not permit certain inference[9].

The assumptions defining the identification paradigm had been key elements for forensic scientists, because it provided justification and credibility for their conclusions. However, it also permitted the obfuscation of information, such as error rates, the lack of empirical data, absence of statistics or even hide bias. The identification paradigm makes an expert assume *decisions*, which are beyond scientifically defensible conclusions. It therefore allows for semitransparent and overconfident reporting of results. The spectacular miscarriages of justice that resulted from these claims have been its downfall and a crucible for forensic science. The events in a forensic case are disputed, happened in the past and are unknown, therefore they are uncertain by nature. Any inference as to the events in the case must necessarily be probabilistic. Even when embracing this fact, Stoney argued that even the integration of statistics and probability, as well as the use of extensive empirical data could never justify individualization [539]. Deterministic conclusions[10] are unfit for any 'scientific' process. It has since been recognized that concluding and reporting in a deterministic fashion requires a leap of faith [539] or a decision [36, 38, 41, 553]. Prominent examples that shook the confidence in forensic science are the Innocence Project, the Mayfield case or the McKie Inquiry. Following these cases, safeguards have been put into the legal systems, in order to ward off subpar scientific testimony. Evidence admissibility hearings and quality control of evidence, experts and testimony have firmly rooted themselves in the different justice systems. The Federal Rules of Evidence, as well as the Daubert criteria adopted by American courts show rising scepticism and opposition to unchecked testimony. Courts of justice have increasingly rejected the infallibility and unjustified claims made by forensic scientists [32, 161, 162, 193, 410, 444, 482, 483, 501]. This criticism has especially marked handwriting and signature examination [385, 387, 431, 432, 434, 481, 485, 499]. In addition to the inference and logical assumptions, language has created its own problems. The continual use and popularization of 'identifications'[11] have created an expectation and connotation for communicating evidence [89, 90, 159, 217, 510]. Whenever nuances in the testimony are necessary,

[8] Using a general rule and an observed result to determine a hypothesis or possible explanation.

[9] Peirce further asserted that there is no such thing as certain knowledge, later in his career.

[10] Especially if they are based solely on evidence and ignore other information such as base rates, statistics, general knowledge, context, etc

[11] Identification is not the only forensic vocabulary that is ambiguous. Much the same has happened to the 'match' terminology in forensic genetics for example. Because of their continuous (mis-)use, these words have become nearly unusable for forensic scientists.

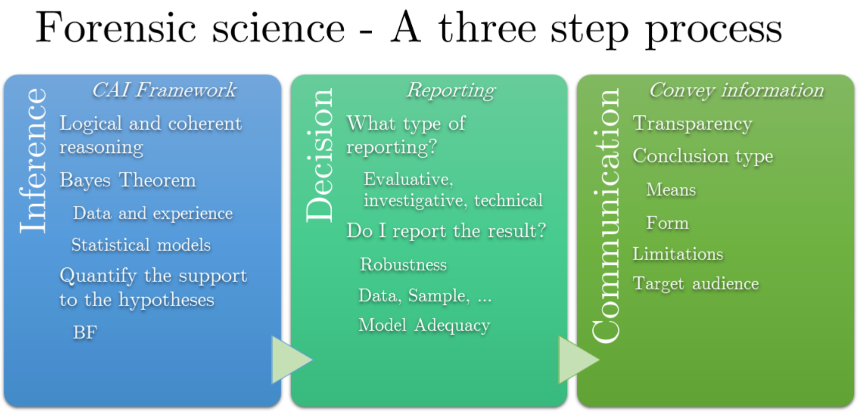


Figure 1.2: A Model of the Forensic Science Process

people fail to grasp uncertainty and limitations inherent to the approaches. This is a consequence of a binary vision (either it's him or not) and failure to consider the uncertainty linked to the assessment, due to language connotation and expression. Clearly, the identification paradigm had outlived its usefulness and a new way for forensic science was necessary.

The forensic science landscape has undergone fundamental changes to find solutions and a new way forward. Finding this path required a revision of the old paradigm. Scholars increasingly embrace the 'evaluative' role of the scientist, a focus on evidence and the clear separation of the scientist's and judge's roles[12]. Forensic scientists are increasingly encouraged to back up their assessments and claims using empirical data and statistical analysis, as well as be transparent in their approach [141, 144, 145, 410, 444]. Another disputed aspect is the prevalence of 'subjectivity' and 'opinion' in pattern matching evidence. In forensic science (as in other sciences), subjectivity is often used pejoratively, although it is omnipresent, no matter the method or scientific process used. Subjectivity is essential in the view of several scholars, and given a logical and coherent framework, deeply subjective elements such as experience, are actually beneficial in the analysis of evidence. One of the key changes is the acceptance of uncertainty in the scientific process and transparent reporting. Forensic reporting should be transparent [141] on methods, data, limitations and errors. Some legal scholars have advocated that FHE's essentially provide 'opinion testimony', to which FHE's object [244]. Forensic scientists have, nevertheless, had to acknowledge the subjective methodological components, human factors [28, 140, 160, 200, 316, 412] and decisions [38, 221] ever present in pattern matching evidence. Subjectivity does not exclude the use of data for assessment [554] and does not make reporting and inference non-scientific. The strong criticism, as well as insights from many scholars in the field have steadily pushed the forensic community towards a paradigm change. Instead of identification, forensic scientists strive to evaluate and assess evidential value. They quantify its support through probability and express their opinions within the limitations. The Bayesian (sometimes called 'logical') approach addresses many of the concerns and continues to garner support from forensic scientists [13, 145, 163, 176, 182, 430, 489]. The road towards a transparent, justified and reproducible forensic science is far from easy, but many forensic scientists have embarked on the journey.

Forensic science is steadily moving towards a probabilistic and information-focused (or perhaps more aptly named *'corroborative information'* [78, 79]) paradigm [10, 13, 31, 61, 78, 79, 81, 87, 88, 145, 281,

[12] According to the ENFSI Guideline on Evaluative reporting, the scientist provides support for one of the considered propositions, while refraining to answer any questions outside his domain of expertise and not requiring his specialist knowledge [145] pp. 5-6, 16.

379, 488, 489, 551, 566, 567]. The novel analysis has recognized the three distinct phases from inference to reporting, as shown in figure 1.2. The ongoing changes concern all of them: inference, decision and reporting. The evaluative paradigm shows an important difference to the identification paradigm on a philosophical and scientific level. Philosophically, the paradigm separates the competence of examiner and legal practitioner by separating hypothesis from evidence. The hypothesis, being the version of 'truth' being evaluated, lies within the confines of the law, while the evidence, the data, lies within the domain of the scientist. In this way, each of the participants can focus on the elements relevant to their distinct tasks, deciding on the 'correct' hypothesis for the court and providing support to make this decision by analysis of the evidence. The separation of expert and judge, isolates the role of investigator and decision maker to create the necessary boundary and distance for each to fulfil their purpose without bias. Scientifically, the paradigm change has created an environment focused on transparency, coherent reasoning, reproducibility and scientific reporting. For evaluative disciplines, such as signature examination, the new paradigm provides a clear way to assess evidence and convey its strength through probabilities. Further, the absence of assumptions, such as the discrete uniqueness, allow more nuanced expression of conclusions. As a result, new fields of forensic evidence, not allowing for statements as strong as forensic genetics, have opened up as a result. Clearly, statistics and probability occupy a key role in the novel paradigm and in legal proceedings [175, 177, 305–311, 522]. Both the inference and reporting steps rely on samples, parameters, variables and populations, using key concepts such as probability and uncertainty [8, 13, 558]. Recognizing the limitations in inference makes the paradigm acceptable and credible, logical frameworks, such as the Bayesian one, guarantee coherent inference. The integration of statistics and empirical data also provide a strong (and verifiable) foundation for examiners' conclusions.

First and foremost, the changes in inference are fundamental to the forensic method. The previous approach, using a single piece of evidence to individualize a source without considering an alternative, is prone to confirmation bias. In order to guarantee coherent and unbiased reasoning, a basic framework for forensic evaluation and interpretation was established [156]. Forensic scientists have started weighing the evidence under several competing propositions, thus clarifying the scope and relevance of their results. They can therefore quantify the probability of the evidence given each scenario and reduce bias. Examiners are also encouraged to do so using statistical frameworks and empirical data in these assessments [9, 13, 99, 152, 153, 158, 339, 340, 558]. Data-driven approaches allow the use of statistics and increase reproducibility, as well as transparency of the assessments. The alternative to the identification process is one of a reduction [78], eliminating sources from a suspect pool and narrowing it down to a smaller population. Individualization forced examiners to consider the world population as suspect pool and narrow the population down to an individual. The current, open paradigm allows for different start and end points in the process. As such, the information provided by the examiner is a reduction factor, which has inferential value by eliminating potential sources. Even if no individualization is (or can be) achieved, the information corroborates one of the considered scenarios and therefore has value for the court. Probabilistic reasoning allows for the use of 'imperfect' evidence and is equipped to deal with method limitations [13, 558]. When accepting the limitations of conclusions and inferential value, the process becomes more scientific and transparent. It also opens up the way for the use of imperfect evidence, nuanced conclusions, as well as multiple hypotheses. In return, evidence is less definitive and straightforward, which complicates the decision processes of the court of law.

Second, is the impact on examiner-specific internal decisions before and for reporting. There are a multitude of smaller, examination related decisions that need to be taken before even considering reporting. Decisions, such as whether or not to pursue the case (pre-assessment), abandon the process due to lack of data, launch a novel study on variation, etc, are necessary when considering imperfect evidence. In addition, this also has consequences on reporting: not all conclusions may be robust enough to report or be helpful. The examiner may then take the decision to report (or not) based on the

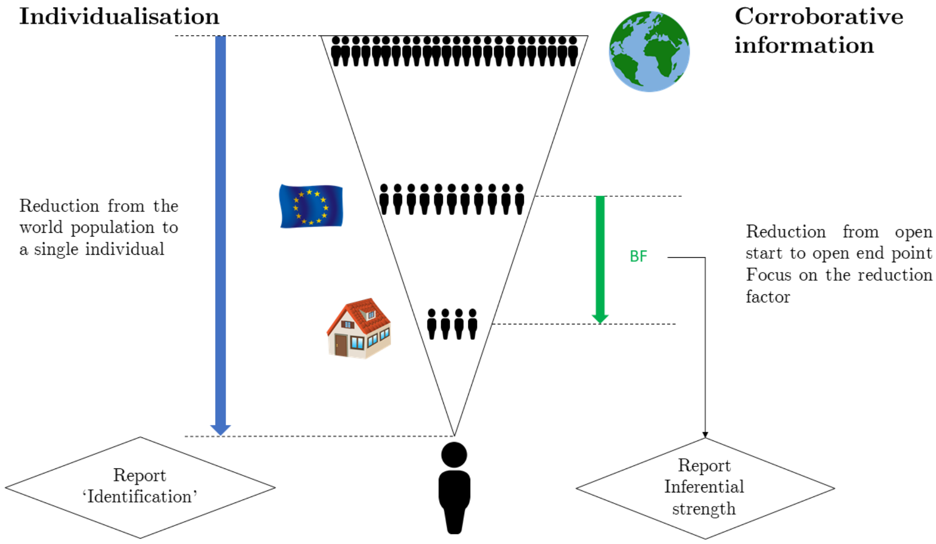


Figure 1.3: Summary of the paradigms in forensic science. Inspired by [77].

information obtained by evaluating and through his process. The imperfect nature of the evidence, the need for the hypothesis to adequately reflect the given problem, and other factors play key roles in this decision. Further, the examiner must decide on what type of information to report. In some cases, reporting evaluative information, through hypotheses and a Bayes' Factor may be unwarranted, or too complex. The examiner must then choose whether to report investigative information, factual reporting or evaluative reporting [145]. A further decision, when reporting in an evaluative fashion, is concerned with how to report the results in a scientific, yet comprehensive fashion. Based on the reduction factor (corroborative information; Bayes' Factor) the examiner may choose a numerical and verbal scale, or decide to report a decision. The difference between decision type conclusions as compared to a deterministic conclusion is the path toward attaining it, as well as the transparency of the process. Introducing a decision component to the reporting process acknowledges uncertainty, while producing a scientifically founded *expert opinion*. The currently predominant recommendation in Europe is to report the reduction factor (Likelihood Ratio, Bayes' Factor, cf 1.3) and a verbal equivalent [145]. However, this approach has encountered some difficulties and produced new challenges, as explained in the next paragraph.

The final step is communication, which is the cornerstone of a good collaboration between forensic scientists and legal practitioners. In this step, the form and content of the conclusion to be transmitted must be chosen carefully. Efficient reporting of scientific results is just as important as the inference step in forensic science. The verdict of the court is taken by judges and juries who need to fully understand the value, limitations and significance of the evidence to make optimal decisions. Communication of probabilistic evidence has proven a non-trivial and open problem [284, 368–371, 564, 565, 568]. No form of communication appears to be expressing only the value of the evidence, as well as adequately reflecting the scope of conclusions and producing the expected decisions. Different types of probabilities, likelihoods, reporting scales and wordings have been proposed, but none has been endorsed internationally. Reporting and presenting results remain a complex and much debated subject, however, methods within this paradigm are more precise and transparent than the previous deterministic conclusions.

1.2 Statistics and the Evaluative Role of the Forensic Scientist

For the interaction between expert and court to bear fruit, the results and process underlying the conclusions need to be understood by the decision makers. The evaluative forensic paradigm features a detailed model of inference, but it also provides a model of the interaction between expert and the court of law, as well as a framework for decision making. As a matter of fact, many elements in forensic examination are subject to choice: the collection of evidence on the crime scene, the sampling of large batch of drugs, the selection of experimental parameters, the determination of an adequate database, the qualification of a conclusion, etc. These decisions are an integral part of the forensic science process and are vital to understanding and communicating evidence. Forensic scientists have used decision theory [41, 221, 553, 558] to describe important phases from examination, sampling, conclusions and communication. Decision analysis is a complex subject, which is outside of the scope of this thesis. At the core of the forensic scientist's task, defined by the law, is the examination and communication of scientific evidence. Inference and statistics are at the heart of evaluative forensic science. Therefore the following paragraphs will focus on the inference step and probabilistic reasoning for signature examination.

A major element in the recent changes to forensic science is the integration of probability and statistics into examination methodologies. A few high profile reports have criticized their absence in most forensic fields and called for their integration [410, 444]. The general framework for forensic scientists has been implemented through the case assessment and interpretation model [99, 153]. Statistics and probability have become essential under the new paradigm, as they are the only way to coherently deal with uncertainty [342]. Statistics are used to describe samples, populations or variables, to create models, to make predictions, to test and select samples and much more. They provide a common vocabulary and the necessary concepts for the forensic scientist to assess the evidence. Forensic examiners use statistical concepts, such as the Likelihood Ratio (LR), to quantify the support provided by the evidence. *Statistics* have contributed reproducibility, transparency and clarity to forensic procedures. They have also instilled a culture of modelling, testing and validation to forensic science, in addition to establishing a greater link to data and developing a more methodical way to use it. In modern times, statistics and validation have almost become *sine qua non* for evidence to be admissible into court. *Probabilities* are used within the discipline to quantify and describe uncertainty. The concept of probability has proven essential within the new paradigm of forensic science, where transparency, uncertainty and error are key elements. To implement a probabilistic framework, the notion of probability must first be understood. Currently, there are two main ideological currents in probability theory, both of which have found their applications in forensic science. The first current, called *Frequentism*, considers probability to be intimately related to the frequency of an event, while the second current, called *Bayesianism*, considers probability to be a subjective belief [219, 229, 236, 238, 493, 554]. Both interpretations have their uses, but also their limitations. Historically, Frequentism garnered more support from scientists and has previously been the standard in most disciplines. Its application, however, requires the studied events to be repetitive and perfectly reproducible. These strict limitations are not respected in most forensic disciplines. Bayesians define probability as a personal belief[13], so even a non-repetitive event may be quantified by a probability. Bayesian methodologies have resurfaced in many fields, such as various social sciences, astrophysics, artificial intelligence and machine learning, as well as forensic science. The Bayesian interpretation of probability provides a logical, structured and coherent framework for reasoning and inference, be it based on data, experience or a combination of both. The core theorem of Bayesian statistics, the law of conditional probability or Bayes' Theorem, is universally accepted and forms a solid scientific basis for inference. It allows for reasoning, communication and decision within a single framework and is used by forensic scientists all over the world [25, 145, 282, 430, 613]. The best way to achieve reliable (forensic)

[13]Personal or subjective as in *linked to the individual*, not to signify *arbitrary*

science, regardless of statistical paradigms, is to integrate statistics, empirical data, as well as validation procedures into forensic science.

A 'Bayesian' reasoner has initial beliefs on hypotheses or parameters of interest, observes some information and updates his beliefs accordingly. He may then use these updated beliefs, called posterior probabilities, to make informed decisions. Forensic Science presents an additional difficulty in the process, as the decision maker and the informer are not one and the same person. Therefore experts do not control the reasoning and decision, which are done by the court. Experts and judges have clearly separated roles and competencies, but also vastly different knowledge [4]. The forensic scientist's duty is to convey the information correctly and help the judge reach a justified conclusion. The decision maker determines his initial beliefs and then receives the information about the evidence from the forensic expert. The decision maker then obtains his updated beliefs, which he uses together with his internal compass (A utility[14] or loss[15] function). The decision maker decides on guilt or innocence, source, activity and crime, while the forensic scientist only expresses his conclusions on the probability of the evidence, rather than the probability of the propositions. In this fashion, he does not implicitly assume the role of the decision maker. In Bayesian terms, the information is conveyed by a ratio called the Bayes Factor (BF)[16] or sometimes a Likelihood Ratio (LR)[17]. The LR is a well documented and established element of forensic methodology and reporting. Through the use of these concepts, the forensic scientist can apply statistics and discuss uncertainty, while respecting the distinctions and limits of the roles of expert and decision maker. The Bayesian framework hence addresses many of the points criticized in past practices. Nonetheless, no framework is perfect and the Bayesian framework is no exception. It has a strong subjective component, meaning that its result depends on the decision maker's belief. Two people, because of different prior beliefs and utility functions, may arrive at diverging conclusions (or decisions) based on the same information. Both may have applied the process correctly, and arrived at their 'correct' belief, but their initial 'conditions' are different. This issue, called the reference class problem, is the main criticism aimed at the Bayesian approach, which is contested by Bayesian scholars [54, 558]. Many scholars expect 'objectivity' and correct prediction based on the outcomes, which require validation and calibration, preferably with reproducible examination and conclusions. While some view this as a problem, it can, philosophically, be an advantage to some others. Each reasoner and decision maker is free to believe whatever he wants, based on his subjective experience and his understanding of circumstances. Nevertheless, the framework also expects the reasoner to attribute his subjective values logically and coherently. Further, this does not fundamentally affect the reasoning processes, such as the assessment of the value of the evidence, which is the key component of the Bayesian method for forensic science. This process is ideally determined by empirical data and 'physical' models, and can therefore claim additional objectivity and reproducibility.

Much research has been devoted to finding ways of assigning and determining BFs, a complicated and domain specific endeavor. Every forensic field is subject to different assumptions concerning populations, feature dependencies, distribution models, constancy and impactful parameters. Using multivariate data to assess the value of the evidence is not a simple task and requires reconsideration for every field. In signature examination, little research on probabilistic inference has been conducted. While some authors are aware of the Bayesian framework, most appear not be using it [244, 392]. Although the case assessment and interpretation framework (CAI) has changed forensic science fundamentally, acceptance and implementation for signature examination have been slow [144, 430]. The recommendation of probabilistic inference in handwriting and signature examination appear to be mostly limited to the European

[14] Decision theory concept; Utility functions define both the action space and consequences (gains) of possible outcomes; the final decision depends on both the posterior probability, but also the desirability of the outcome

[15] Decision Theory concept; A loss function is conceptually the inverse of a utility function. Instead of maximizing gain (utility) it minimizes loss

[16] Shortened to *BF* for the remainder of the document.

[17] Shortened to *LR* for the remainder of the document.

continent [40, 108, 144, 201, 282, 359, 361]. Most authors use the logical foundation and principles of these frameworks, to guarantee coherent inference. Their approaches are however qualitative [108, 361], with few authors actively using collected data [194, 201, 317, 347, 351, 354] for inference. Signature examination as a field, in particular dynamic signatures, can benefit from the integration of data-driven statistical frameworks. Their integration will strengthen the evaluative methodology currently employed, as well as provide an empirical basis and justification for conclusions. A more detailed description of the subject can be found in chapter 4.

1.3 Reporting in Forensic Science

Communication is the final step of the evaluative forensic process. It is a key component for the court to make an informed decision of guilt or innocence, by providing relevant information for their decision. According to decision theory [221], coherent and justifiable decisions maximize utility by considering both the consequences of the decision and the probability of the event[18]. The 'best' decision, which depends on a subjective utility function, is called the Bayes action. In forensic science, only posterior probabilities are adequate for inference and decision making, as they relate to the events. Therefore, information such as the LR needs to be transformed before being used by a decision maker. The role of the forensic scientist is also to accompany the decision maker in the process to understand the information and its scope as fully as possible. The efficiency of forensic science depends on the ability of the expert and decision maker to transmit, assimilate and use the information extracted from the evidence. Forensic principles enforce strict separation of roles in the process, meaning that the forensic examiner should generally not provide his posterior belief and influence the decision maker. Probabilistic assessments of evidence have been a great asset for transparency, however they have complicated the communication of results. Deterministic statements may be simplistic representations, however, understanding them is not an issue. This is not true with statistical inference and uncertainty. Subtle changes in wording may cause misunderstandings, due to the complexity of the probabilities and elements involved. A failure in transmitting, comprehending or updating due to logical fallacies may cause faulty inference and 'bad' decisions. Further, the human component also plays a role in the framework. Several authors [284, 368–371, 564, 565, 568] have been able to show that people may misunderstand or misuse the information provided to them. Kahneman [295] has shown that people do not always obey the 'rules' of rationality when taking decisions. It is insufficient to dispose of a 'perfectly rational' system if the users lack education and training to use it correctly.

The methodological and conceptual changes to forensic science have had an impact on communication between the forensic scientist and the court. First, the type of reporting plays an important part in the communication of evidential value, uncertainty and limitations. Jackson et al. [284] classified the types of possible reporting into several categories: explanatory statements, deterministic conclusions, features and simple match conclusions, numerical conclusions, as well as verbal equivalents (to numerical conclusions). All of these categories can be further divided into several subcategories. The authors mention a fundamental difference between these types of conclusions. Historically, forensic scientists used deterministic conclusions. Explanatory statements and deterministic conclusions leave no place for uncertainty and error. They convey an impression of absolute certainty, a position which is impossible to defend in court. Due to strong criticism, experts have had to change their approach to reporting, to express uncertainty and limitations in addition to evidential value. As a result, reporting turned towards numerical solutions to quantify the uncertainty. Numerical conclusions, such as probabilities or matching numbers, provide the decision maker with a number, most often the value of the evidence or the 'random

[18] In our previous jargon, these events would be called hypotheses or parameters of interest.

match probability'[19]. They quantify the uncertainty inherent to the assessment, but may confuse the decision maker, as their meaning and form may be very diverse. They may cause various logical fallacies during their interpretations [13]. Verbal equivalents translate those numbers to a qualified verbal scale, often spanning several categories based on strength of the support provided. According to Jackson et al., no conclusion type is universally well received and understood. While some numerical conclusions appear more precise than the other types, their use requires some statistical knowledge for the receiver to handle them correctly. Personal preference and affinity, as well as education and background play a large role in the comprehension of conclusions [28, 266, 267, 283, 284].

Different numerical reporting solutions have been proposed, such as the random match probability, the probability of inclusion, the probability of exclusion, the likelihood ratio, etc. Several fields proposed posterior probabilities concerning the events, such as the paternity index (PI) in genetics, rather than the evidence-related probabilities such as the LR as reporting solutions. The numerical ratios and probabilities do not appeal to all audiences. This led to the adoption of probabilistic and qualified reporting scales. These are a conventionally agreed upon classification of conclusions, which distinguishes a variety of levels of support. The first of these reporting scales was based on posterior probability, with thresholds on specific probability values, warranting various degrees of 'certainty'. Depending on the interval the numerical values fall in, the evidential value is translated with a specific wording. These probabilities and verbal scales were in conflict with the separation of the roles of expert and decision maker. The changes in forensic science and probabilistic inference, such as the case assessment and interpretation model have led to a different type of probability for expressing conclusions. These probabilities are focused on the evidence and measure its likelihood given one of the propositions. To this effect, forensic scientists have redefined their verbal scales based on the LR, rather than on the posterior probabilities. Currently, the predominant position appears to be the use of both numerical values and verbal scales [25, 145, 284, 359] for reporting. While these methods are the preferred ways at present, the Bayesian framework is capable of supporting numerical, verbal or decision-based conclusions. Communicating LRs, or other numerical statements for that matter, is a demanding exercise, requiring rigour and some level of statistical literacy from both parties. Even using a combination of the diverse communication methods, the exchange between a legal and a technical expert is far from easy. The Bayesian Framework appears to be a worthwhile solution. However, efficient communication must be built upon more than only a statistical framework. The way to efficient communication is paved by the development of clear terminology, as well as statistical and linguistic education for all parties involved.

[19] Also called 'conditional (profile) probabilities.

2

Handwritten Electronic Signatures

Handwritten electronic signatures and dynamic signatures in particular are the main subject of the present research project. These signatures are a composite construct mixing electronics, sensors, movement and neuroscience. They contain many complexities, ranging from terminology to creation process. Additionally, electronic signatures are rapidly evolving on the technical, legal and standardization level. A general overvire of the subject of handwritten electronic signatures is provided. The field of dynamic signatures and its interest for forensic handwriting examiners is presented in greater detail. Finally, the structure and analysis of these signatures is briefly summarized.

2.1 General Information

2.1.1 Context, definitions & terminology

Handwritten signatures hold an important place in our legal system. They are used as signs of authenticity and intent in almost all of our official transactions, ranging from small purchases to important leases, work or commerce deals. They enjoy such popularity because they are universally accepted, difficult to forge and free of negative connotation[1]. Along the same lines, electronic signatures are first and foremost a legal concept, meant to accompany and facilitate the increasing digital[2] and online[3] transactions, commerce and government. Electronic signatures[4] are a consequence of the age of big data, as well as the increase of formality[5] and complexity of modern society. An ever increasing number of contracts, leases and other official documents are being signed. These transactions are increasingly made digitally and occasionally remotely over the Internet. The Swiss federal law on electronic signatures provides a short definition, based on the European **e**lectronic **Id**entification, **A**uthentication and Trust **S**ervices (*eIDAS*) framework, which reads as follows *"data in electronic form, which is logically associated with other data in electronic form and which is used by the signatory to sign."*. In order to facilitate international commerce, legal texts all over the world were designed for compatibility and feature similar definitions. Handwritten electronic signatures are a combination of both handwriting and the electronic signature concept. A handwritten signature is the product of the execution of a complex and highly automatic movement. It is

[1]Unlike fingerprints or DNA, which are heavily associated with criminal activities.
[2]Digital here means dematerialized, as in computer-based.
[3]Conducted over the World Wide Web (Internet).
[4]Sometimes shortened to eSignature.
[5]Formality induces the increase in administrative documents and controls.

a well accepted behavioral biometric, characterized through both the graphic image of the signature, but also through the complex movement sequence composing it. Creating a handwritten electronic signature requires transforming either the movement sequence or the image of the signature into digital data and securely linking it to a document. With the high resolution digitalizing technology deployed on a large scale (e.g. in tablets, smartphones and computers), a handwritten signature can be recorded virtually anywhere in a matter of seconds. Given the demand for easy, familiar ways to obtain electronic signatures, an increase in the popularity of handwritten electronic signature solutions is to be expected. The wide distribution, confidence and familiarity in the signature make it an attractive solution for businesses and very relevant for forensic scientists.

The word 'electronic Signature' is a collective noun, regrouping many different types of signatures under a common legal umbrella. Many types of electronic signatures exist, including vocal, textual, cryptographic or handwritten signatures (see chapter 10). The function of (handwritten) electronic signatures is to authenticate a document. Signed documents are often important, as they create legal or financial obligations. Transactions of this importance carry an increased risk of fraud. One type of fraud is directly related to the handwritten signature, namely signature forgery and disguise. Handwritten electronic signatures are just as likely to be disputed and challenged as their inked equivalents. Furthermore, in cases of transaction conducted remotely, risk of dispute may be even greater, as there is no meeting with a physical intermediary and thus no witnesses to the signature. The complex interplay of the electronic support, digital data, connectivity and human aspect pose a great challenge to forensic science. The field is a complex interplay of both digital and physical evidence. Handwritten electronic signature cases require a skill set beyond that of pattern matching, as well as an extended methodology for evidence evaluation. Digital evidence and pattern matching specialists need to be involved together to determine originality through digital evidence and metadata, as well as source through the biometric data. The forensic handwriting examiner's[6] main task, helping in verifying the authenticity of the handwritten signature by providing scientific support to the court, has not changed, however the path toward it has. The field is plagued with ambiguous and complicated terminology, which is confusing both for lay people and experts. Even within the handwritten electronic signatures field, terminological specificities and subcategories exist. For example, handwritten signatures can either be present under image form, or as list of measured points. They can also be digitized by an image scanner or camera, or be captured on an electronic device during the signature. Distinctions and technicalities are numerous and may cause confusion. In this thesis, all types of handwritten electronic signatures, no matter if scanned or recorded during signature, are referred to as handwritten electronic signatures. Within this main class, there are two broad categories, based on the type of data representing the signature. One may distinguish static (offline; image-based) and dynamic (online; list-based) signatures, which are briefly described in the following paragraphs.

Static signatures are essentially digital images, pixel-based representations of signatures. They are currently the most widespread form of handwritten electronic signatures. They are called 'static', as they ignore (or suppress) the kinematic and temporal data during their creation. They have been called offline signatures in biometrics, although the terminology is confusing in regards to either connectivity[7] or biometric verification[8]. They are generally captured on point-of-sale devices, using old resistive touchscreen technology. Other solutions are based on non-sophisticated programs using capacitive touchscreens on computers and smartphones. Common applications include delivery receipts and post or banking transactions. Some authors distinguish digitally captured static signatures from digitalized 'physical' signatures. The latter were transformed into digital images via an image scanner or camera. In addition, one may

[6]Referenced as FHE for the remainder of the document.

[7]As in *connected to the internet*.

[8]The matching and verification computations can either occur simultaneously with capture (online) or separately (offline).

add that static signatures can be generated from dynamic signatures by saving them as images. Both scanned and transformed dynamic signatures contain at least some information on signature dynamics, while digitally captured static signatures do not.

Dynamic (or online) signatures are digitally captured signatures. They are recorded using a digitizing device capable of recording both static and dynamic information. In accordance to biometric standards [253, 279, 503], they need to contain specific information and obey a predefined structure. Dynamic signatures feature temporal sampling of the signature signal. They are chronological lists of data points from the signature in a given coordinate space. The data collected during the capture may include kinematic measurements, such as pen force, speed, or inclination, in addition to the coordinates and the timing information. Many different devices can produce dynamic signatures, although by far the most common means are dedicated signature pads. These signature pads are equipped with touchscreen technologies, displays (and digital ink) and (active or passive) pens to recreate regular signing conditions. Other devices capable of creating dynamic signatures, such as gloves, virtual reality device controllers, or camera systems, exist. The recorded data will depend on the recording device. This thesis will focus solely on signature pads and pen-based capturing solutions.

The recently updated ENFSI Best Practice Manual on the examination of handwriting now includes a section on digitally captured signatures. It introduces new terminology [144]. Handwritten electronic signatures are therein referred to as digitally captured signatures (DCS)[9]. Their definition reads :

> A DCS is a handwritten signature which is digitized during its production. Even though both DCSs and conventional handwriting and signatures are products of writing behavior, a DCS usually contains more information, such as spatial coordinates, time and pressure values.

This terminology is confusing as it contains the term 'digital', linked to digital signature, as well as being a close acronym to the Digital Signature Certificate (DSC). By the proposed definition, it appears clear that scanned signatures do not fall within the DCS category. Further, the definition points to all digitally captured signatures (both image and list-based). However, the terminological section declares them to be only of the list-based, dynamic type. Instead of the newly proposed terminology, this thesis will use the clearer dynamic and static signature distinction.

2.1.2 A History of Handwritten Electronic Signatures

Handwritten electronic signatures have a rich history, which is still rapidly evolving. They share the history of signatures, handwriting and digital signatures, and have some specific events that relate to them. They link handwriting, electronics and cryptography into a single complex object. A graphical overview of some of the key events can be found in 2.1. Their history begins with handwriting. The first records of handwritten signatures come from the Jewish and Muslim communities in the third and sixth century AD. The first officially documented signature was made in 1098 by Rodrigo Diaz (El Cid), a Spanish military leader. In the past, handwritten signatures were the reserve of nobles and educated people. The general society started to use signatures only much later, when literacy and education spread. In 1677, England passed the *Statute of Frauds*, which made signatures mandatory on important contracts. Ever since, handwritten signatures have become a standard way of authentication when conducting business. Handwritten signatures have kept their firm monopoly on legal agreements since then.

The dawn of technology progressively changed this long status quo. In 1869, The United States

[9] The acronym DCS will be used for the remainder of the document.

Supreme Court first accepted telegraphed Signatures, the first precursor of electronic signatures, as equivalent to handwritten signatures. In 1965, Automated Signature Recognition Research is launched by a feasibility study of the North American Aviation Co. in Anaheim, marking the start of handwritten electronic signatures. Research progresses through the 1970s, focusing on static signatures. Diffie and Hellman [122] first describe the concept of digital signature (cryptographic signature) in 1976. In 1977, the patent for the acquisition of dynamic signature data is awarded to Veripen, Inc. and research on dynamic signatures begins. The same year, the RSA[10] algorithm for asymmetric cryptography is invented and produces the first cryptographic signature. The technological evolution, as well as the marriage of both concepts, has not stopped since.

In legislation, the way for the dynamic signature was paved in the 90s. In 1996, the United Nations Commission on International Trade Law publish the UNCITRAL Model Law on Electronic Commerce, which influences electronic signature legislation all over the world. The year 1999 is marked by the proposal and adoption of electronic signature 'laws' in Europe and the United States of America. The European Union proposes a first directive for members, the *DIRECTIVE 1999/93/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 December 1999 on a Community framework for electronic signatures* [435], while in the US the National Conference of Commissioners adopt the *Uniform Electronic Transactions Act (UETA)* [411] and Australia adopted its own *Electronic Transactions Act (ETA)* [26]. These legislations provide a legal framework and basis for the use of electronic signatures for government and business transactions. The US further adopted the Electronic Signatures in Global and National Commerce Act in 2000 into federal law [19], ensuring legal validity of electronic signatures. Subsequently, European countries adopted their national legislations on electronic signatures.

Once the legal situations allowed for the use of eSignatures, there was a strong need for regulation, standardization and regular updates. In 2005, the American National Standards Institute (ANSI) standard defines the format for interchange of biometric sign and signature data in the *INCITS 395-2005* standard [20]. A similar international standard has been published in 2007 by the International Organisation for Standardisation (ISO), as *ISO/IEC 19794-7*[11] [279, 280]. The demand for electronic signatures on mobile devices increased with the first releases of smartphones in 2007. In 2008, the ISO officially included digital signatures in the open standard of the portable document file (pdf) standard. It also added a specific dynamic signature standard, *ISO 19794-11*, in 2013 [278]. The widespread touchscreen and acquisition capabilities also increase possibilities of use of dynamic signatures. The European Union has since replaced their initial directive with the *REGULATION (EU) No 910/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market.* More recently, the ISO committees on biometrics have published the *ISO/IEC 30107-1*, *ISO/IEC 30107-2* and *ISO/IEC 30107-3* standards on presentation attack detection between 2016 and 2017 [277]. These standards specifically define physical impersonation attacks on biometric systems, such as forgery in dynamic signatures, and are therefore relevant for 'forensic applications' of biometrics.

[10] From the authors initials: Ron **R**ivest, Adi **S**hamir, and Leonard **A**dleman.
[11] The standard was revised in 2014.

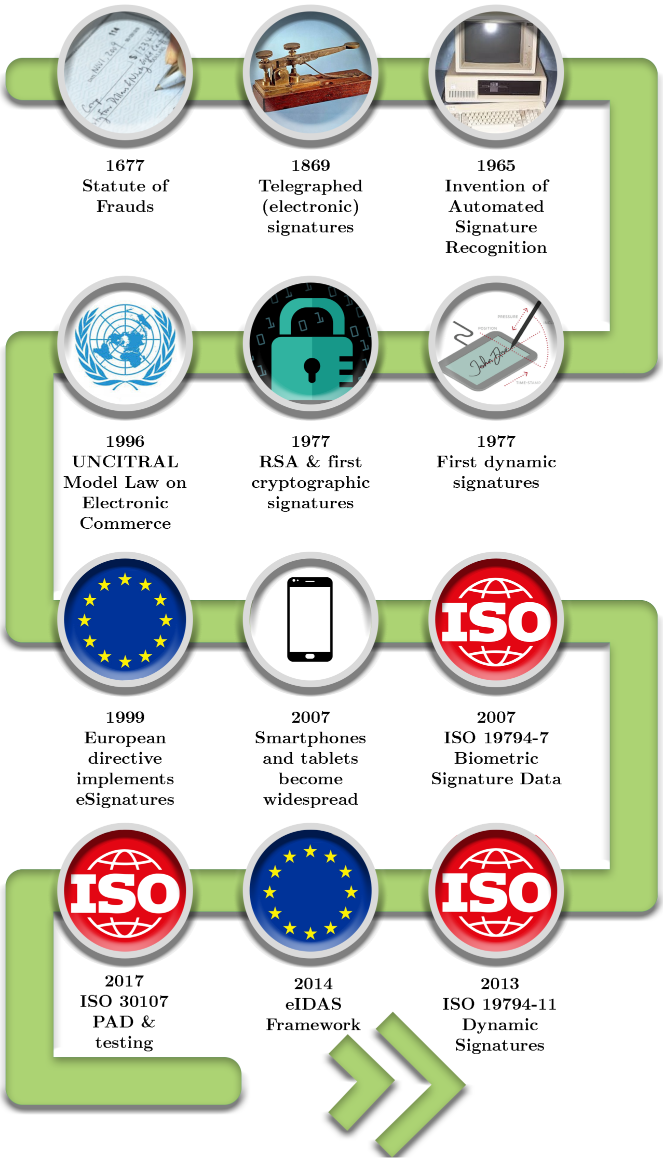


Figure 2.1: Infographic summarizing the history of dynamic signatures.

2.2 Dynamic Signatures

2.2.1 Applications and Cases

Dynamic signatures are solutions that are cost-effective, fast, eco-friendly, traceable and easy to integrate [330, 331, 333]. They are additionally perceived as very secure [332, 333], more so than 'wet' signatures. Further, people are comfortable using handwritten signatures and consider them nonintrusive, as well as free of connotation[12] [59, 286, 383]. They simplify and speed up workflow by replacing paperwork and planning with logical data, as well as automated treatment and archiving. Advantages for businesses are numerous, which has led to an increasing interest in these dynamic signature solutions throughout the world. Applications already include banking and financial sectors, insurance (e.g. life insurance), medical sector (e.g. responsibility and consent sheets), tourism, deliveries and sales. Development and deployment of technology is still ongoing and it is expected to result in further increase of dynamic signatures.

While the current use of eSignatures resides mostly in deliveries, the documents signed with dynamic signatures are not limited to 'small' cases. Documents such as life insurances or medical consent sheets are highly relevant to forensic scientists, because their of their importance and thus their potential of appearing in court cases. One prominent case in the United States of America, concerning a life insurance in 2007 has been documented[13]. Several other minor cases, mostly involving credit card fraud have been documented [100, 587, 588]. In Switzerland, there have been several cases involving deliveries, none of which have gone to court because of the small disputed sum. Additionally, more important cases are starting to surface. In Eastern Europe, a dozen cases have been reported. Germany's BKA has shown an interest in electronic signatures, especially in counter-based transactions and has started a research project on dynamic signature and posture. The European Network of Forensic Science Institutions (ENFSI) and its expert group on handwriting examination (ENFHEX) have recently updated their best practice manual on handwriting examination to include the examination of DCS. Interest in the subject is on the rise both nationally and internationally.

2.2.2 Hardware, Software & Data

Dynamic signatures can be made in a variety of ways, using many different devices. All of the hardware capable of recording dynamic signatures will be referred to as digitizers or digitizing devices. Digitizers sample the analogue writing signal[14], record several features, such as coordinates and kinematics, and finally output a signature file. They are of many different types (such as tablets, cameras, gloves, input devices or pens) and use various sensors (inductive, capacitive, accelerometers, gyrometers, cameras, ...). Most of the dynamic signatures produced today are made on dedicated signature pads or mobile devices, such as tablets and smartphones. Signature pads are, contrary to mobile devices, computer peripherals, requiring a personal computer. They are operated via a device driver, often written by either the Operating System provider, or the hardware manufacturer. Their operation is ensured via a graphical user interface[15] in the form of a software package. Software and hardware packages are often not provided by the same manufacturer. Many software packages are compatible with multiple hardware solutions. These hardware peripherals communicate the signature data to the computer, which creates,

[12] Contrary to fingerprints, mugshots or DNA, which are strongly associated with the criminal and carceral worlds.

[13] AFLAC v Biles; [586].

[14] the handwriting movements

[15] GUI for the remainder of the document

encrypts and embeds the signature into the digital document.

Signature pads need two elements to record signatures, a writing surface (often a display) and a writing instrument (often a passive pen; stylus). The writing surface's crucial component is a sensor, which (most often) uses touchscreen technology. The most prevalent of these are inductive (Electro Magnetic Resonance; EMR), capacitive (operating on conductivity of electrical current; e.g. tablets and smartphones) and resistive sensors. High quality systems have almost all adopted inductive sensors, while resistive sensors are disappearing. The writing instrument is variable and ranges from pens, to fingers. It may either be active[16] or passive[17]. All digitizers have their own specific characteristics, most importantly their spatial resolution[18] (and precision) as well as their temporal resolution[19]. Other factors, such as the presence of digital ink, surface rugosity and the pen nib are equally important in forensic examinations. These provide the writer with proprioceptive information, which is essential for the writing process to feel natural to the writer and for the writing product be representative of their writing habits.

The digitizer records pen data in regular or event-based[20] time intervals and saves every sampled point in a chronological list. The pen data is pretreated by the driver and sent to the computer. Pretreatment may include interpolation [197] to smooth the data, as well as normalization of the signals (amplification or reduction, especially for pen pressure). The signature data is composed by several input-related columns, such as an indexing of data points, button-related columns or pen state (in air vs on surface), and the measured data. The measured data must include spatial coordinates of the data point, as well as sampling time. Additional data and measurement depend on the hardware, but may include pressure, tilt, inclination. Signatures are encapsulated in signature container formats, including pen data, but potentially also metadata, and a timestamp, which are encrypted. After decryption and extraction, the pen data is available as a simple table[21]. An example of the structure of this signature data is available in table 2.1.

The manufacturers of the signature pads often use proprietary data formats as signature container files, which complicates standardized data analysis. Most manufacturers propose their own comparison and verification software suite for their clients, as well as more specialized software for forensic experts[22]. Forensic analysis software is currently provided by the manufacturers of the dynamic signature solutions and may differ significantly in measurements, calculus and methodology. In most cases the software is a tool for the visualization and qualitative comparison of signatures, rather than a statistical or forensic analysis suite. Some of these solutions permit the extraction of features, such as diverse measurements and calculated features (i.e. number of strokes, number of velocity inversions, average speed, ...). NeuroScript's MovAlyzeR software [419], designed for research purposes can adapt to many signature pads and may be a solution for the current situation. These software packages feature very variable depth of analysis and visualization capabilities. Most of them permit only qualitative (as opposed to statistical or automatic) comparison between two signatures. No open-source analysis software, for use on all dynamic signatures or with a statistical evaluation framework, is available. Visualization and qualitative comparison of dynamic signatures are the first steps to integrate them into forensic signature examination

[16] As opposed to passive; an active pen uses electrical components and measures data and input directly. Active pens are often battery dependent, e.g. Apple Pencil.

[17] As opposed to active; passive pens do not contain electronic components. They need only be adequate to write on the surfaces; Passive writing instruments can range from regular pens, plastic nibbed pens, to the human finger.

[18] Often measured in lines per inch [lpi].

[19] Number of points sampled per second in [Hz].

[20] The temporal sampling depends heavily on the device used for digitizing. Some devices, such as smartphones and tablets show irregular (event-based) sampling periods.

[21] Often a comma separated values (.csv) file.

[22] Prominent examples for this are Wacom's SignatureScope, Topaz's SigAnalyze and Namirial's FirmaCerta software.

Input Data			Signature Data			
Index	**Button**	**Pen State**	**X**	**Y**	**P [Level]**	**T [ms]**
1	1	1	15005	7600	300	0
2	1	1	15015	7605	334	5
3	1	1	15035	7615	370	10
4	1	1	15040	7613	400	15
5	1	1	15045	7612	476	20
6	0	2 (UP)	15063	7625	511	25
7	0	2 (UP)	15078	7627	480	30
8	0	2 (UP)	15120	7633	430	35
9	1	3	15150	7638	400	40
10	1	3	15167	7639	412	45
...	..	...	...	...	...	...

Table 2.1: Example of the data structure of a dynamic signature. The input data describes the state of the implement, with a column for the chronological order of the points, the button pressed while recording, and the stroke number. The following column group contains the actual signature data, with 3D coordinates (X,Y,P) and the timing component.

methods. These methodologies assimilate and reduce the dynamic signatures to 'wet' signatures and do arguably waste much of their potential. A methodology and tool capable of using and statistically comparing the novel dynamic and measured features would greatly complement current forensic examination methodology.

2.2.3 Technical Analysis of Dynamic Signature Data

Dynamic signature data is complex; it is formed by multiple measurements and descriptors of the handwriting movements. Biometricians and forensic scientists select and extract specific characteristics of the signature for analysis and comparison. They are commonly called *features* in biometrics. Features have been classified into several types, according to their detail. Richiardi et al. [302, 479] propose a three level classification for signatures (see table 2.2). In addition to being complex data, the spatio-temporal signals that are signatures exhibit strong correlation in their features. Diverse parameters, such as shape, velocity and pen pressure have been shown to be correlated [116, 598], adding complexity due to the interplay of variables. The additional complexity of signature data warrants care in the choice of the statistical analyses, models and interpretation of their results.

Forensic scientists use diverse types of features in their subjective assessments, combining them intuitively. The scheme takes on much more importance when considering signatures in the light of statistical analysis, seeing how the feature types require different statistical analyses and models. In forensic science, the main method of signature examination is the qualitative comparison of shape, dynamism and fluidity, as well as specific artefacts. They operate similarly to pattern matching methods in biometrics, using a variety of local, regional and global features. They combine the information intuitively, rather than through a statistical model. Independently of the feature-type used, signature data is multivariate[23]. Each of the three feature types has different advantages and issues. *Global features* convey general information about the signatures that complements shape-based information, and are easy to use within statistical models. They have however been shown to produce worse verification accuracy than systems using more detailed local features. *Local features* convey more extensive, in-depth information about the

[23]It contains multiple variables and is therefore multidimensional.

Feature Type	Level of Detail	Variable Type	Examples
Local (Functional)	'Point per point descriptor'	Timeseries, Pseudo-Timeseries	Instantaneous Speed, Instantaneous Pressure, ...
Regional (Segment, Partition)	'Segmented descriptors'	Scalar, Boolean, ...	Length of Strokes, Average Speed during Stroke A, ...
Global (Parametric)	'Unique descriptor'	Scalar, Boolean, ...	Average Speed, Number of Strokes, ...

Table 2.2: A taxonomy of signature feature types with the level of details, the variable type and an example representing the category.

variation, however they present various difficulties in matching and treatment. They may have diverging lengths and require complex matching or modeling algorithms (Dynamic Programming, Hidden Markov Models, ...). *Regional features* are easier to use, in the vein of global features and are slightly more detailed. They lack the detail of local features, as well as 'universality' as they are signature specific and require reproducible segmentation of the signature. Local and global features have been shown to be complementary and improve performance of verification systems when used in conjunction. In the current state of the art, little if any information on the treatment and combination of different feature types within the scope of forensic signature examination is available. As a starting point, this thesis focuses mainly on global features, which present less of a challenge in terms of statistical modeling and additional opportunities for cross signature comparison.

As mentioned above, the diverse feature types require data treatment, analysis and modeling adequate to their type. Traditionally, histograms, as well as parametric or non-parametric statistical models are used to represent and compare global features. Local features are generally represented as a (pseudo-) timeseries, a continuous signal through time (cf 2.2). Some researchers use time series averaging or bounds to represent multiple local features. Local features are often compared by template matching techniques, such as Dynamic Time Warping (DTW)[24], Longest Common Subsequence (LCSS), Fréchet Distance, ... Alternatively, there are also modeling approaches, using Hidden Markov Models (HMM) or Gaussian Mixture Models (GMM) to deal with local feature matching. In biometrics, results are often rendered as classification and prediction. To this effect, a variety of algorithms have been used, including Support Vector Machines (SVM), Vector Quantization (VQ), various Neural Networks (NN, ANN, CNN, ...) among others are used [27, 30, 121, 271, 273, 460, 548]. An extensive description of the subject can be found in several reviews [121, 271, 273]. In this thesis, only global feature data is used for evidence evaluation. Methodology-wise, a parametric model for within-writer and between-writer variation is used to evaluate the evidence and produce the Bayes' Factor. In stark contrast to previous approaches, this methodology assesses the probability of the evidence under competing hypotheses and adheres to the ENFSI Guideline on evaluative reporting [145].

As for inference, two approaches currently exist within the evaluative framework: feature-based and score-based inference. Feature-based approaches are easily applicable to global features, while score-based approaches are predominantly used for local features. Feature-based approaches use features themselves.

[24] A type of dynamic programming algorithm that enables elastic matching of two signals.

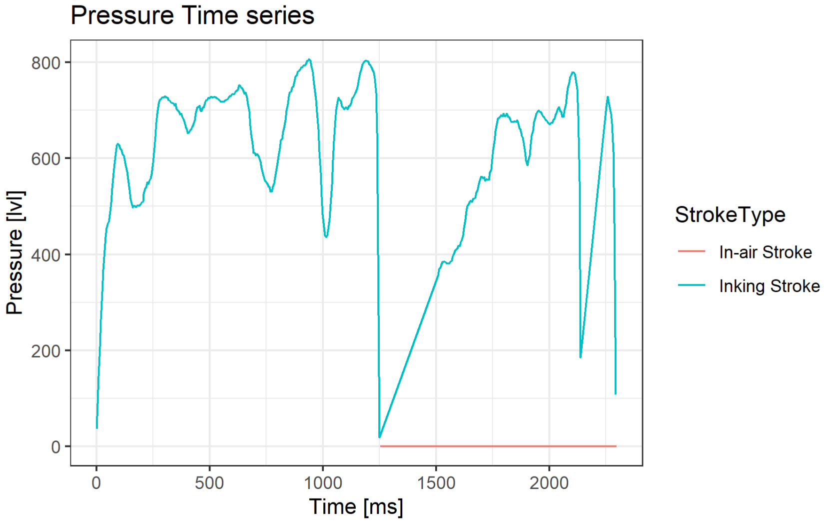


Figure 2.2: An example of a signature pressure timeseries.

They are limited by 'curse of dimensionality' [301] when used for a high number of features. Score-based approaches match templates one-to-one and produce (dis-)similarity scores between the templates. The statistical models are not applied directly on the feature values, but on comparison scores. As a result, the scoring causes a loss of information as to the original feature value and rarity, but reduces the dimensionality in the process. Therefore, score-based approaches may use a high number of features and still use a simple statistical model, which is an advantage in terms of data volume necessary to obtain robust results. Score-based approaches are however often computationally costly and slow because of the matching step. The mathematical equivalence of Score-based Likelihood Ratios (SLR)[25] [15, 50, 254, 504, 505] to the BF, as well as their theoretical adequacy for application within the Bayesian Framework have also recently been questioned [414]. The application and validity of SLRs remains controversial in the forensic community.

Both approaches have previously been used in signature examination. Marquis et al. propose a qualitative feature-based approach [361] to signature examination and Bayes' Factor assessment. While not data-driven, their approach uses qualitative and subjective assessments of feature-values. Gaborini et al. [201] use a feature-based model for 1:1 writer comparison. A signature verification competition [354] used a 'calibration'[26] procedure to transform match scores from automated systems to Likelihood Ratios. In this thesis, we will focus on feature-based inference, rather than on score-based approaches. The mathematical and statistical adequacy and 'purity' of SLRs being in question, feature-based approaches are currently of more use and interest to the FHE community. Further, feature-based approaches may help with interpreting numerical data, such as lengths, sizes and writing angles, which are often neglected during comparison. Bayesian statistical models are key components in this type of reasoning and will be explained in the next chapter.

[25] Sometimes also Similarity-based Likelihood Ratios.

[26] Calibration has multiple meanings. According to Jacquet et al. [285], Calibration is used to either transpose matching scores to Likelihood Ratios, or to scale a predicted probability to an observed probability [58, 62, 63, 110, 111, 470]

3 The Bayesian Framework and Bayesian Statistics

The application of statistics and probabilistic reasoning in handwriting examination is a core subject in this thesis. In order to guarantee transparency in the methodology, the statistic approach used in this thesis needs to be defined. For the reasons explained in chapter 1, the chosen framework is the Bayesian one. The latter has garnered strong support among forensic scientists in Europe and presents a rational and transparent framework. This chapter lays down the statistical basis of the computations and experiments described further into this manuscript. The following paragraphs explain the basis and the processes of the Bayesian statistical framework, including Bayes' Theorem, statistical models, belief updating and model selection. Bayes' Factors, the main method of assessing and communicating evidential strength in forensic science [145], are a result of statistical model comparison. The paragraphs also include an example of the application of these processes to a forensic science problem. The comprehension of the model contributes to transparency and scientific accuracy in the conclusions reported by a forensic scientist. Understanding the models and the process is also essential to understand the methods and computations during the experimental part of the thesis.

3.1 Bayes' Theorem

The Bayesian framework is a subjectivist framework for reasoning and decision making in which probabilities are defined as the subjective belief of the reasoner. As a consequence of this definition, the application of the rules of probability to singular events is possible [109].[1]It has alternatively been called the 'logical' approach [430, 489], as it provides a coherent way for inductive reasoning. It provides a model encompassing both inference and decision processes, as introduced in section 1.2. Forensic science involves and invokes both parts of the Bayesian Framework [221]. The focus of this thesis lies however with inference, rather than with decisions. The manuscript will hence solely focus on the elements relevant to probabilistic inference and statistics.

The foundation of the Bayesian framework is a mathematical formula known as Bayes' theorem, named in honour of clergyman and mathematician Reverend Thomas Bayes. The formula is derived from universally accepted rules of probability and is uncontroversial [522]. It models a belief updating process. Succinctly, Bayes' Theorem shows how the occurrence of one event affects our belief in another event. The events are *hypotheses*, which are often called *propositions* in forensic science. The observed data is often

[1]By this, de Finetti meant that Frequentist definitions of probability require perfectly reproducible experiments, which does not apply to situations such as elections or crime.

also called the evidence E in statistics and forensic science. All information is conditional on background knowledge, which is often noted as I. Bayes' theorem links these elements through probabilities, which express the relationship between two or more events. The subjective probability about the occurrence of an event is updated by the observation of some data, leading to a new, updated belief. This general frame holds for forensic science, with the only differences that the hypotheses are of judicial interest, and the data is the combination of features of recovered (traces) and reference (comparison) objects.

The prior probability on H_i, where i = 1,...,n describe the state of knowledge on an event before observing the data. The Bayes Factor describes the support offered by the data to the hypotheses of interest H_i. The posterior probability on H_i describes the new state of knowledge after having observed the data. Bayes' Theorem (Eqn. 3.1) describes a process of 'updating' probabilities on H_i after obtaining the new information E.

$$\overbrace{P(H_i|E,I)}^{\text{Posterior Prob.}} = \frac{\overbrace{P(E|H_i,I)}^{\text{Evidence}} \times \overbrace{P(H_i|I)}^{\text{Prior Prob.}}}{\underbrace{P(E|I)}_{\text{Normalization Constant}}} \tag{3.1}$$

$$P(H_i|E,I) = \frac{P(E|H_i,I) \times P(H_i,I)}{\sum_{j=1}^{n} P(E|H_j,I) \times P(H_j|I)} \tag{3.2}$$

$$P(H_i|E,I) = \frac{f_{E|H_i,I} \times P(H_i,I)}{f_{E|I}} \tag{3.3}$$

$$\overbrace{\frac{P(H_1|I)}{P(H_2|I)}}^{\text{Prior Odds}} \times \overbrace{\frac{P(E|H_1,I)}{P(E|H_2,I)}}^{\text{Bayes Factor}} = \overbrace{\frac{P(H_1|E,I)}{P(H_2|E,I)}}^{\text{Posterior Odds}} \tag{3.4}$$

$$\overbrace{\frac{P(H_1|I)}{P(H_2|I)}}^{\text{Decision Maker}} \times \underbrace{\frac{P(E|H_1,I)}{P(E|H_2,I)}}_{\text{Forensic Scientist}} = \overbrace{\frac{P(H_1|E,I)}{P(H_2|E,I)}}^{\text{Decision Maker}} \tag{3.5}$$

Bayes' theorem (Eqn. 3.1) is a universally accepted mathematical formula involving conditional probabilities. Divergences in use between statistical paradigms come from the definition of a probability and from the usage of prior probabilities. For simplicity, the necessary concepts will be explained for the case of discrete data. In Bayesian statistics and forensic science, it is common to see the odds form of the Bayes' theorem with two events H_1 and H_2 (Eqn. 3.4). This form is obtained by dividing the equations for $P(H_1|E,I)$ and $P(H_2|E,I)$. Ultimately, Bayes' theorem is a model of a learning process, a guide on how to integrate new information with previous knowledge. Although forensic scientists use the Bayes Factor as tool to quantify and communicate their results, only posterior probabilities have actual meaning for inference. Therefore, combining the Bayes Factor with the prior probabilities is essential for the decision maker. As this is generally not the role of forensic scientists, this aspect is often not communicated or sufficiently stressed in research. It proves confusing and frustrating for the courts and may lead to misinterpretation of evidence.

3.2 Bayesian Models and Model Comparison

Bayesian statistics diverge from Frequentist statistics in several important aspects. For an in-depth review, the interested reader may referr to [11]. Frequentist scholars rely on the definition: "probability is the limit of the relative frequency of a target event that has occurred in a large number of trials [...] under identical conditions" ([11], p58), while Bayesians define it as a subjective belief. Frequentism is strongly focused on the analysis of samples, whereas Bayesians focus on parameters. Therefore, their inference can inform them about the parameter value, such as the proportion of voters in a population. Bayesian statistics are more intuitive in their interpretation, even if their computation is more arduous. Bayesian inference employs probability distributions to model these parameters and use the expectations of the distribution for inference. Bayesian Hierarchical Models, layered (nested) probability distributions, can be used to describe complex processes with interdependencies. Bayesian estimation is computationally intensive, because it requires integral computation. As a result, various approximation techniques, such as *Markov Chain Monte Carlo* (MCMC) along with importance sampling, or *Nested Sampling*, are required to determine posterior and marginal distributions for complex models. These distributions are essential to calculate the BF when using continuous data and thus vital to forensic scientists.

In evaluative forensic science, experts express the likelihood of the evidence under several competing models. Thereby, they perform a Bayesian model selection process, focusing on the Bayes' Factor, which expresses the probability of the evidence given the alternative scenarios. The ultimate goal is to compare two different models, to determine whether a questioned signature is better explained by one or the other model.

In the following paragraphs and equations, the architecture of a simple statistical model for signature length is described. Such a model is then fitted to the two alternative scenarios, namely the signatures being either genuine or forged. The two resulting models are used for the evaluation of the questioned signature. When there are several competing models, they can be compared using the Bayes Factor. The BF indicates which model explains the observed data better. This model selection process is exactly what Bayes Factors (or Likelihood Ratios) are used for in forensic science. The BF measures support for each of the propositions and expresses under which the evidence is more likely, therefore providing the information needed for model selection. Forensic scientists use this process to assess evidence and report their conclusions. In the case of signature examination, the models that are being compared are those of two distinct populations, the first being genuine signatures and the second being forged signatures (freehand forgeries).

$$\theta \sim \mathcal{N}(\mu,\, \tau^2)\,. \tag{3.6}$$

$$E|\theta \sim \mathcal{N}(\theta,\, \sigma^2)\,. \tag{3.7}$$

$$\theta|E \sim \mathcal{N}(\mu^*,\, \tau^{*2})\,. \tag{3.8}$$

A simple example of a Bayesian model, based on a parameter θ representing the length of a signature, is provided in this paragraph. The parameter of interest is assumed to be distributed according to a Normal[2] distribution (prior distribution; Eqn. 3.6). The likelihood of the evidence E is also assumed to be normally distributed (Eqn. 3.7). In this configuration, the posterior distribution for signature length θ will also be normally distributed (Eqn. 3.8, Figure 3.1). Using the newly calculated posterior distribution, we can now provide an estimate of the most likely length of the signature (the expectation

[2]Gaussian distribution; $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

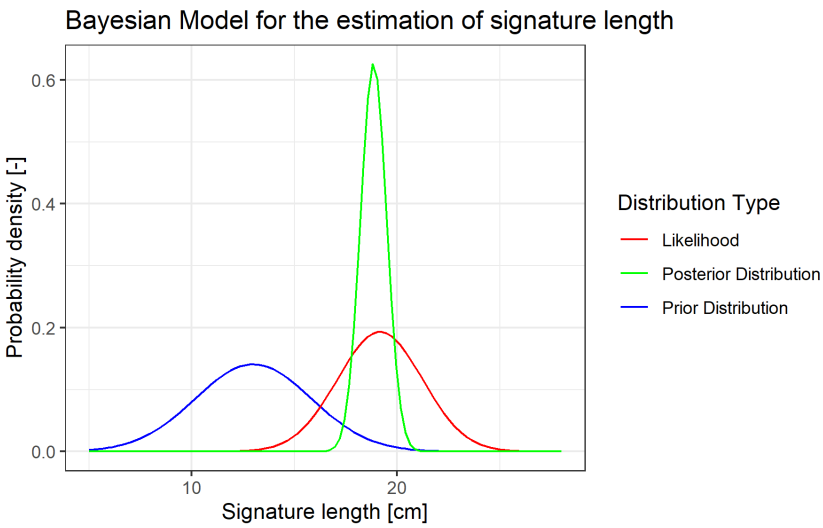


Figure 3.1: Simple Bayesian Model for the signature length example

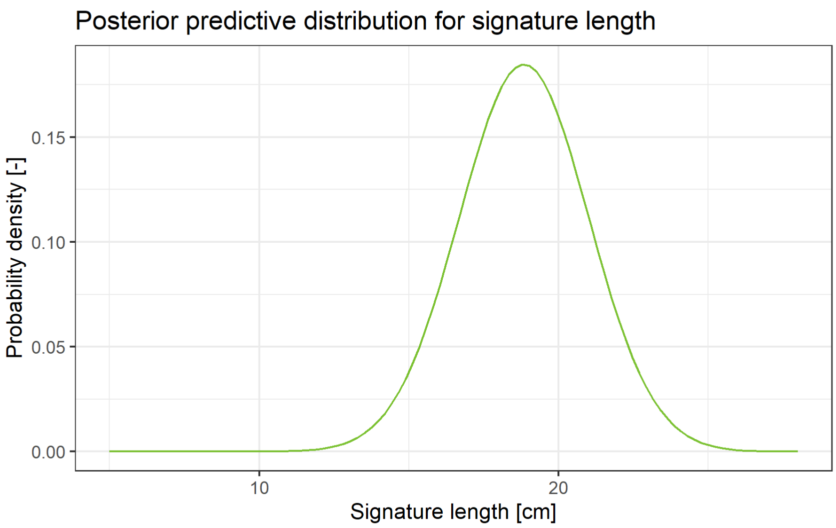


Figure 3.2: Posterior Predictive Distribution for the signature length model. The posterior predictive distribution shows how future datasets may look like and how likely the different signature lengths are to occur in such a dataset.

of the distribution), but also a credible interval for the parameter . Furthermore, we can also look at the posterior predictive distribution of the parameter (cf Fig. 3.2) in order to predict the length of signatures from the same writer based on our newly updated knowledge on both the mean value and his natural variation.

A univariate example of a univariate model comparison can be found in Fig. 3.3. The green and red distributions represent the model of the genuine and forged signature lengths respectively. In the example, the questioned signature shows a length of 18.8 cm. The probabilities of such a value under either model is shown by the vertical bars. In this case, the BF is determined by the ratio between these probabilities. The BF is 13.6, meaning that the questioned signature's value is approximately 13 times more likely to occur in a genuine signature rather than in a forgery. If a decision as to the authenticity of the signature had to be made, the posterior probability of the signature being a genuine signature would need to be calculated by multiplying the BF with the prior probabilities (odds) of the signature being genuine. Forensic scientists are generally encouraged to report the likelihood of the evidence (as BF or

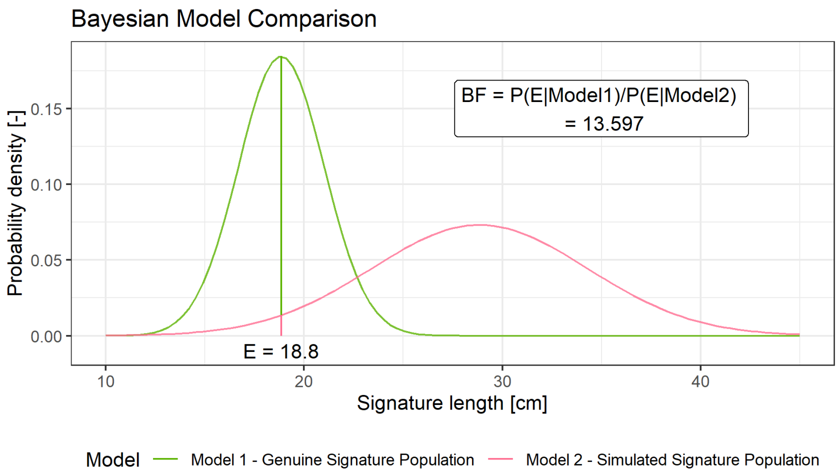


Figure 3.3: An example of a Bayesian model comparison and BF calculation for questioned signatures

LR) rather than the posterior probability, because they do not have the necessary information to access the prior probabilities.

3.3 Bayes' Factor

The Bayes' Factor is an essential element of Bayesian statistics, even if the inference within the framework is based on the posterior probability. The BF is a weighted measure of support given by the evidence and is used for model comparison. The propositions define the formula and context of the Bayes Factor. The observations and background knowledge define the inferential strength. Inferential strength is a concept related to how strongly the information impacts the probabilities involved. For example a BF of 10 means that however strong your prior beliefs were, your posterior beliefs should be 10 times stronger after hearing the evidence. When visualizing the belief in hypotheses as a scale, the BF shifts the scale, without necessarily tipping it. The actual state of the scale (tipped toward left or right) also depends on the prior information. In forensic science, the Bayes Factor, rather than posterior probability, occupies the most important role. It is the main method for the quantification of evidential value and for the communication of the results. Forensic scientists have however taken to using a slightly different concept called the 'Likelihood Ratio'[3] as surrogate for the Bayes Factor. The BF can be seen as parent concept to the LR in the Bayesian framework. For a Bayesian, all LRs are in fact a special case of Bayes' Factors[4]. The interested readers are referred to the Handbook of Forensic Statistics [29] (pp. 125). In the remainder of this document, we will prefer the terminology of Bayes Factors, even when they coincide with a LR.

$$\text{Weight of Evidence (WoE)} = log(BF) \tag{3.9}$$

The Bayes' Factor has provided forensic science with a solution to provide the court with information while respecting their 'sovereignty'. Instead of reporting unsustainable claims of individualization, it permits the examiner to express his probabilities *on the evidence given the proposition.* It further works within a clearly defined framework and with unambiguous propositions and assumptions. It therefore contributes to clarity and transparency in the scientific process and its reporting. Its use promotes a logical and transparent way of reasoning and communication. Bayes Factors, in much the same way, are

[3]Shortened to LR.
[4]BFs reduce to LRs when using simple hypotheses. A more detailed explanation can be found in [555]

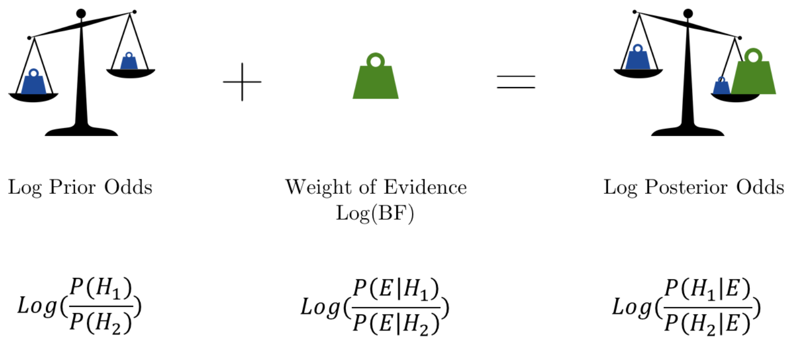


Figure 3.4: A representation of the belief updating process using the weight of evidence. The I term from the formula was omitted for brevity.

used in forensic science [25, 145, 282–284, 359, 488] to communicate the value of findings. As shown before, Bayes Factors are combined with the prior odds by multiplying the two elements. These ratios may take on very large numbers, considered intimidating and confusing for people without a scientific background. Some authors have proposed to use the *Weight of Evidence*; Eqn. 3.9) [225, 489] rather than the BF. The WoE maps the evidential value to the logarithmic scale, which shrinks the numerical values. The combination of the WoE and the log Prior Odds also only requires an addition, rather than multiplication. Some authors consider this form simpler and more intuitive, as it works analogously to a scale. The log Prior Odds represent the initial state of the scale (tipped or in equilibrium) while the *weight of evidence* represents the weight to be added onto the scale. The positivity of the WoE indicates the side the weight is added to, the numeric value indicates the mass of the weight. The log Posterior Odds represent the state of the scale after it re-equilibrates, showing the final support to either hypothesis (cf Fig. 3.4). Both procedures yield the same result, except for the calculus and scale of evidential value. While combination of the different elements is more straightforward in the logarithmic scale[5], probabilities and odds are easier to grasp in the regular scale. Both representations of inferential strength are useful. Ideally, both forms should be used for calculations and explanations respectively, whenever the case permits it.

3.4 Bayesian Statistics in Forensic Science

Bayesian statistics have played an important role in forensic interpretation and case assessment. They attribute meaning to the analytical results obtained by defining a logical frame. Bayesian inference enabled forensic scientists to abandon two-stage approaches to evidence evaluation [152, 339] , to be more coherent and diminish bias by considering alternative explanations. Bayesian methodology is transparent, as it requires clearly stating hypothesis and assumptions [99, 153, 155, 488]. Hypotheses are the cornerstone of valid evaluation and interpretation in forensic casework. Researchers have accordingly proposed a hierarchy for propositions with several levels, such as source, activity and crime-related propositions [98]. This hierarchical classification was further researched, in particular the source-level (e.g. impression, method, part, geographical) and the logical linkage that can be established [276, 318]. Further, Forensic scientists have shown the benefits of applying a solid methodological and statistical framework [158]. The Bayesian Framework has enabled forensic scientists to develop data driven approaches [13, 56, 67, 415,

[5] At least when the odds form of Bayes theorem exists.

558]. They provided the framework for combining the examiner's experience and theoretical knowledge with empirical data from collected case data and databases. The framework's transparency also helps during the communication of the results to the court. The framework also allowed for the integration of rational decision analysis and support for forensic scientists. These tools help practitioners with sampling strategies, examination strategy, prioritization and trace triage. These methodological elements have helped to add scientific and statistical rigour to many forensic disciplines, as well as respond to criticism concerning the methodological and fundamental issues of subjectivity, logical incoherence and unfounded assumptions previously associated with forensic science by several authors [386, 482, 483, 485, 499, 501, 502]. Their integration into dynamic signature evaluation conforms to the ENFSI evaluative reporting guidelines [145] and is a step toward a uniform reporting solution.

The Likelihood Ratio / Bayes' Factor occupies a particularly important position in forensic science. Therefore, many associations and forensic institutions now recommend using these concepts for expressing evidential value and communicating with the courts [25, 145, 359, 488]. The concepts and theory of Bayesian statistics have found their application in many forensic disciplines, including Genetics [66, 67], Fingermark examination [76, 472], Shoemark examination [519], Handwriting examination [56, 362, 556], Glass (and other microscopic elements) examination [6, 7], Firearm and toolmark examination [51, 82], Speaker identification [83, 405, 469], Facial Recognition [285, 569], as well as Chemiometrics and Investigative approaches, such as Profiling for drug and fire accelerant analysis [50, 473, 624, 625]. In forensic handwriting and signature examination several applications and models relying on Bayesian statistics have been documented. Day [108] explains the Bayesian Framework as applied to handwriting examination. Recent publications [244, 392, 428] in signature examination show the rising awareness of Bayesian methodology and statistics in the field, even if not yet fully embraced by examiners. European scholars have been more open to the Bayesian framework within forensic science. Applications involve both qualitative [361] and quantitative [56, 194, 201, 362, 363, 562] approaches. Various authors have created score-based models for signature evaluation [317, 345, 351, 352, 354, 355]. A more detailed account of forensic applications and methodology for dynamic signature examination is proposed in the next chapter.

4 Dynamic Signatures in Forensic Science : A State of the Art

Dynamic signatures have only recently moved into the focus of forensic scientists. In comparison to 'wet signatures', many of the relevant circumstances, including writing conditions and resolution, but also fundamental information, such as recorded data and data structure have evolved. Nevertheless, the underlying neurological and motor processes for signing are the same no matter the writing surface, implying that the subject is worthy of interest but unlikely to revolutionize signature examination. The shift to the digital world, as well as the novel kinematic data, call for a revision and extension of the existing signature examination methods. The novelty of dynamic signatures is technological, mechanical and metrological. These changes have caused gaps in the technical knowledge, such as human-sensor interaction, recording quality and writing conditions and variation, but also in methodological questions, such as the examination of dynamic features, their comparability to 'wet' signatures and the probabilistic evaluation of features.

Exploratory analysis of technical aspects of dynamic signatures is a crucial step to approach the new subject and lay down groundwork for coherent inference. A second, albeit just as necessary, step is the validation of the probabilistic procedures used to assess the value of features. Risinger and Saks [486, 500, 501] have criticized a lack of validation and testing in forensic science, notably in questioned documents. Since, many performance and validation studies have been conducted [42, 43, 189, 297, 355, 517]. Several of these studies were in turn criticized [498] for not reflecting and validating the procedure that forensic examiners actually use. The validation of methods forensic examiners can actually use in practice and integrate into their examination procedures are therefore vital to the credibility of signature examination in legal proceedings. Knowledge of the signature data and its underlying distribution is essential for the forensic examiners if they aim to increasingly use empirical data in their assessments. It is especially relevant when fitting statistical models to signature data, as one must determine valid assumptions and limitations, as well as a suitable probability density. As part of this thesis, the methodological advances in dynamic signatures have been reviewed in chapter 10. Since the publication of this paper, new studies have nevertheless been published. It appears worthwhile to review some general points, as well as novel additions in the following paragraphs.

4.1 Methodological Advances

Electronic and dynamic signatures have only appeared recently in the forensic landscape. These novel signatures have first been mentioned and categorized by LaVelle [327] and later by Flynn [185]. Harralson

[242, 243] adopted further clarifications for terminology, which were further clarified by Linden et al. [335]. Di Toma [115] calls for the establishment of clear terminology in order to guarantee reproducibility and validity in handwriting and graphonomics research. Galbally et al. [202] have reviewed the traditional classification dichotomy applied in biometrics. They insisted on the theoretical aspects of 'presentation attacks'[1]. They qualify signature forgery as 'mimicry', signifying a clear dependence on a target signature. The difference in 'attack' and type should be reflected through comparison data and in computations. Recent work on presentation attack detection (PAD) [571, 577] has proposed to classify forgeries according to the forgers knowledge of the target signature. This levelled classification enables researchers to clearly describe their samples and methodology in a unified framework. Adopting their terminology explains both the collected data and experimental conditions, which are vital to understand dynamic signature studies. A common vocabulary and clear terminology create a solid basis for further research and communication.

Forensic examiners have been examining 'wet' signatures for more than a century. The principles for their analysis and comparison have been laid out in great detail and serve as a solid basis for pattern-matching based signature examination. Guiding principles, such as detailed documentation, progressing from general to detailed analysis, as well as the ACE-V[2] procedure [23, 268, 321] are universally applicable and remain valid. Many textbooks on signature examination lay down the fundamentals for qualitative analysis [72, 244, 247, 259, 312, 334, 392, 396, 427]. Further, Australian researchers have shared a complete method for the analysis of questioned handwriting [187]. The subjective and qualitative examination of shape and signature features are disciplines forensic examiners have proven proficient and accurate in [42, 43, 189, 297, 355, 517]. Recently forensic handwriting examination has been further validated and its reliability assessed [376]. Further studies strengthened the basis of forensic handwriting (and signature) examination and addresses many of the points made by critics, finding subjective judgment to be acceptable [130] and procedures to be reliable. Several authors have tried to qualify [251, 336, 391] whether dynamic signatures are of sufficient quality to allow forensic examination and testimony. Especially the book chapter by Flynn et al. [392], in a previously unpublished study, support that previous examination methodologies remain applicable to dynamic signature examination. These studies also include validation of both the hardware and the 'traditional' signature examination method on electronic handwritten signatures. They conclude that dynamic signatures may have lower resolution than inked signatures, but make up for this disadvantage by recording previously inaccessible temporal and kinematic data on the signatures. This data and the resulting features are useful to the forensic examiner. The existing method is structured and proven and should prove adequate for the visual features of dynamic signatures, with a few exceptions due to differences in ink deposition and digitalization [14, 197, 218, 235]. Harralson et al. [244] propose some adaptations due to the examination of dynamic signatures in the recently revised version of their textbook. The amendment to existing methodology mostly concerns the detection of electronic handwritten signatures, as well as standard questions to ask whenever they appear in cases. Mohammed [392] lays out the new challenges for forensic scientists posed by handwritten electronic signatures. He also summarizes several technical studies and insists on the methodology remaining applicable.

The basis for statistical analysis and evaluation of signatures is not as well documented as the subjective analysis method. The framework for case assessment and interpretation [99, 153, 489] is laid out in general for evaluative forensic disciplines, but a specific adaptation for signature examination is yet to be found. The ENFSI standard on the examination of handwriting evidence **??** vaguely alludes to propositions, while the ENFSI guideline for evaluative reporting [145] recommends using probabilistic approaches (Bayesian or not), as well as respecting the principles of evidence interpretation [156]. Ostrum

[1] An attack on the biometric sensor where the impostor is physically present and produces the signature.

[2] Abbreviation for the four-phase process used across many forensic disciplines. The phases are, in order, Analysis, Comparison, Evaluation and Verification.

[430] has recently released a position statement of the Canadian Society of Forensic Science's (CSFS) Document Section endorsing the logical approach. Mohammed [392] very briefly touches upon Bayesian reporting and confirms the limited awareness and support of said approach in FHE circles. There have been discussions on formulating propositions, especially regarding vocal comparison [40, 80, 155, 156, 257, 258, 398]. The discussion shows that the determination of propositions is far from a simple task and further shapes the data collection and the evaluation process. The arguments and considerations made in these vocal comparison cases are equally relevant to signature examination. Several comprehensive publications on assessing probabilities and LRs in signature examination [108, 359, 361] through a qualitative assignation of subjective probabilities lays groundwork for the Bayesian framework in signature examination. They demonstrate how probabilistic methodology and a logical, structured approach can be integrated into a field, even without using extensive databases.

Applications and methodology incorporating empirical data and statistical models in handwriting examination can be found in several scientific articles [13, 56, 194, 201, 360, 362, 363, 552, 555, 562, 563]. Specific applications to signature examination are rarer. Franco-Pedroso et al. [194] have proposed a generative model for signature evaluation using Gaussian Mixture Models (GMMs). Gaborini et al. [201] have presented a generative probabilistic model for writer to writer comparison for signatures. Their model is feature-based and relies on Gaussian distributions for the BF calculation. Both of these studies are specific to signature examination and show that probabilistic methodologies can be developed and used to support examiners. Montani et al. [394, 395] discuss the integration and usage of statistical and data-driven approaches with subjective inference. Their research shows the way toward successful integration of the two approaches and further collaboration between forensic and biometric researchers, as attempted by several other authors [194, 241, 287, 351, 374, 381, 443]. Several authors have participated in studies and competitions integrating both disguised signatures and using score-based likelihood ratios [317, 345–347, 350–353]. In these applications, a calibration method (isotonic regression or logistic regression) was used to 'translate' matching scores to Score-based LRs [285].

4.2 Technical Studies

A review of studies on dynamic signatures or dynamic features measured on 'wet' signatures can be found in chapter 10. This section aims to complement the proposed review with a few technical, technological, matching and verification related developments.

Dynamic signatures are complex blends of digital signatures, cryptography, hashing algorithms, time-stamping, as well as metadata. In order to provide proof of originality, date signatures, and technical consideration, such as determining illicit copying and assessing the security of the signature generation process, IT expertise is indispensable. Harralson [242, 244] stresses the complex interplay of technology, evidence and sensors (IT, Biometrics & Forensic Science) and recommends collaborating with digital evidence experts and IT experts [242, 244]. The delicate and novel elements especially concern timestamps, metadata and digital files, all of which are subjects outside of the 'traditional' FHE's area of expertise. Electronic handwritten signatures are first and foremost electronic evidence [372], which requires computer forensic analysis and extensive logging. Reporting and presenting this type of evidence in court should therefore rely on principles formulated by these scholars and profit from their know-how.

Diaz et al. [121] have determined that a (technology-oriented) state-of-the-art for dynamic signatures is necessary approximately every 10 years. They proceed in reviewing new relevant developments in signature verification, such as models operating with few reference signatures, novel verifiers and technical matching. Mazzolini et al. [374] have proposed a computational framework to support forensic handwriting examiners in their daily task. They use a majority-voting approach to combine dynamic feature information and support decision-making. Other authors propose automatic systems and computational frameworks to support examiners [443, 591]. Further, several researchers have used a model of movement dynamics, the sigma-lognormal model [457, 458], for dynamic signature verification [118, 183, 184]. Many other researchers have applied machine learning techniques to signature verification [27, 33, 34, 121, 273] (CNN, DP, DTW, DWT, NN, RF, Deep Learning, ...) [138, 576, 599–602, 617–620]. Several authors [27, 440, 441, 450, 452, 453] propose to analyze signature datasets to determine stable regions (for local features) and weighing these accordingly in verification and examination. Important new advances also include distortion models to generate synthetic dynamic signatures for research datasets [75, 171–174, 203–205, 208, 237, 467], high quality forgeries [174] or verification tasks [203]. Novel approaches also integrate concepts such as signature complexity [294, 496, 535, 537, 580, 592] and template quality [210, 262–264, 315, 406, 494] into the verification task and have achieved encouraging results. Several authors [119, 120, 423] have proposed a system able to deal with signature verification using a single reference through a distortion model. Recently, a system using no references was proposed [117]. These methods, specifically synthetic signatures and distortion models, may resolve issues such as the curse of dimensionality, by creating additional, although synthetic data. With this additional data, cases with little available materials [120, 578] may still be treated reliably with statistical models.

As far as exploratory studies go, several researchers have confirmed the correlations between signature trajectory and dynamics [116, 319, 509, 560]. Helm et al. [252] have shown that disguised movements, such as feint throws in ball games, use slightly different kinematic patterns compared to natural movements, that can be recognized. This may also extend to signature kinematics and to forensic examiners investigating the signature data. Mohammed [392] documented previously unpublished trials on electronic and dynamic signatures by Flynn and Nicolaides. They concluded that a temporal resolution (sampling rate) of 100 Hz produces signatures of sufficient quality for forensic examination. Additionally, if these conditions are fulfilled, the examination methodology for paper-based signatures applies neatly to electronic handwritten signatures. Further, the same authors have explored dynamic features such as writing rhythm, by measuring it through velocity, acceleration and jerk plots. Their results support that writing rhythm is consistent within writers and carries discriminative information for forgery detection. Some additional studies focusing on determining robust and coherent features for forensic examination exist [261, 337, 384, 433, 436–439, 442, 443, 550, 591, 620]. Multiple authors have investigated sensor interoperability issues [18, 46, 455, 572, 573, 581]. Tolosana et al. [572] in particular have conducted a study on digitizer interoperability and the resulting verification performance. Their tests include five different signature pads and mobile devices, as well as pen and finger as writing implements. Finger-drawn signatures appear to be less precise than pen-drawn signatures and also present lower verification accuracy, especially when mixing finger- and pen-drawn materials. Overall, comparing signatures from different digitizers appears to be a strongly limiting factor for conclusions. Signatures on mobile devices are also very popular research subjects [22, 45–47, 129, 181, 231, 265, 272, 314, 349, 364, 495, 546, 626]. Their verification produces encouraging results, although finger-drawn signatures and the small display sizes may cause lower accuracy than traditional systems.

4.3 Summary

Research into dynamic signatures has picked up over the last decade. Most of the studies are however focused on technical elements, such as better matching or classification, with a strong machine learning theme, rather than a focus on fundamental aspects of signature examination and forensic science. The few fundamental studies focus on applying traditional signature examination methodology on dynamic signatures, instead of re-examining and developing novel ones. The developments made on the technical aspects concerning analysis and comparison of dynamic signatures are impressive and valuable, however they lack a theory-based framework to work within. The forensic science specific evidence evaluation model has still not been fully integrated into signature examination. Research should further focus on score- and feature-based approaches and their robustness. These elements complete methodology by providing coherent ways to deal with data and quantify the value of the evidence reported by a forensic scientist. Developing this framework will provide the necessary basis for examination of relevant features and reporting their value, but will also transparently communicate and share knowledge among researchers. As such, they would greatly contribute to general acceptance of knowledge between researchers, validation of methods as well as overall transparency in signature examination.

Part II

Methodology & Experiments

Data Acquisition, Sampling and Databases 5

The thesis has the ambition to produce usable information to practitioners out of large-scale empirical trials. The practical part of the thesis is accordingly focused on reconstructing casework and studying a variety of factors, such as the statistical model behavior, the validity and generality of the model, the feature selection and the dynamic signature variation. The evidence evaluation model in particular requires the precise specification of the relevant population and of databases representative of said populations. In order to produce results that are relevant and usable by examiners in their work, detailed descriptions of writing and hardware conditions are necessary. Alas, conventions regarding signature form (and content) are intimately linked to the geographical region, the local education system, as well as the culture and the alphabet. Furthermore, the signature may also vary depending on the type of document, whether a frame or dotted line is provided. Ideally, data acquisition strategies need to take these factors into account. Further, the choices for data acquisition made on a case-specific basis should always be well documented. While already critical in casework, data acquisition is fundamental for research as well. In particular, methodologies around probabilistic inference require extensive testing and therefore a large dataset of known source data in strongly controlled conditions. In order to deliver useful results, data used for a study needs to conform to a chosen set of criteria. The selected conditions, databases and hardware used in this thesis are summarized in the following sections.

5.1 Databases and Hardware

One of the goals of this research is to study personal variation of dynamic signatures in detail, while another is the development, validation and testing of a probabilistic model. These tasks require massive amounts of data for exploration. The dataset also has to serve both as parameter estimation data[1] and validation data[2]. In order to study the variation of dynamic signatures, a high number of acquisition sessions, spread out through time, is required. Additionally, the dataset needs to feature a high number of 'skilled' forgers and forgeries for every signature, preferably freehand forgeries, to extensively test the model. Our first criterion is the extent and applicability of the data. The dataset should cover long time periods, as well as present a high number of signatures per studied signer. For the results of the study to bear any relevance for practitioners, a solution already in use by businesses is necessary. Hence, the dataset for the study should be acquired on currently available hardware, which can be found in point-of-sales and offices. It should preferably be a signature pad with high resolution, as might be used for important contracts. The second criterion is therefore deployment of the solution and practical relevance

[1] Often referred to as *training data* in biometrics.
[2] Often referred to as *test data* in biometrics.

of the results. Our third and final criterion was applicability to casework and 'realism'. The dataset used in the study serves for exploratory data analysis, but also for testing and model validation. The latter milestone involved creating pseudo-casework cases and evaluating these using the probabilistic model. The 'realism' of the generated cases highly depends on the quality of the data and conditions in which the forgers were allowed to work. In order to produce meaningful results, both the original signature quality and the forgery quality needed to be high, to be close to real world scenarios. Ideally, the forgery data set should feature pre-session training, assessment of forgeries and selection of high quality forgeries by the forger.

In an effort to determine the necessity of a data acquisition, a search for adequate publicly available dynamic signature databases was performed. A summary of the proposed and published datasets is provided in Tables 5.1, 5.2, 5.3 and 5.4. Of all the reviewed datasets, none fits all the requirements for the hypotheses proposed in this research. Most recently published datasets are repackaged or extended subsamples of older ones, like the ATVS long-term database. As of late, no acquisitions with multiple acquisition sessions on novel hardware were organized. Further, the questions of interoperability of hardware and writing instruments [18, 21, 22, 455, 572, 581, 590, 593, 595], synthetic signatures [75, 117, 119, 171–173, 205, 208, 237, 467] and mobile devices [22, 265, 272, 314, 349, 367, 495, 546, 572, 590] are currently of special interest. It appears clear that template age, variability and signature tablets are moving out of focus of the current research. As shown in Tables 5.1, 5.2, 5.3 and 5.4, no single dataset fulfils all requirements completely.

As none of the publicly available datasets are adequate, an original data collection was organized. The acquisition campaign, while time consuming, yielded an adequate, large longitudinal and population-specific dataset, which is perfectly adequate for the use with the proposed probabilistic model. Further, a state-of-the art digitizing device and optimal conditions can be chosen, which is essential for interpreting differences observed in the results of the probabilistic evaluation. First, the technical parameters of the data acquisition needed to be selected. These considerations required hardware, methodological and organizational choices. As far as hardware was concerned, the important parameters included high accuracy, easy usability and fast setup. A suitable hardware and software package, permitting easy recording, decryption and exportation of signature data needed to be selected. In order to find a suitable solution for the study, a market study was conducted with the previously defined criteria in mind. These factors included spatial and temporal resolution, technology and reliability, kinematic measurements, as well as user-friendliness, usability, deployment in businesses and practical relevance. A listing of the top models at the time of the research is provided below in table 5.5. All dedicated business solutions had associated forensic or biometric software for analysis and comparison purposes available. These tools were however proprietary and often linked to the manufacturer. These tools were not used beyond data extraction.

After careful consideration of the previously enumerated criteria, the Wacom DTU-1141 Signature tablet was chosen (see Figure 5.1). This pad is a signature pad with a large display, causing no visual or mechanical constraints while signing. The tablet uses the (formerly) patented electromagnetic resonance [373, 422, 621] technology, a passive pen and axial pressure measurements. The tablet features high precision measurements (0,1 mm precision on coordinates) and is also used in the smaller STU-530 signature pad. The hardware is usually bought with a software suite to create dynamic signatures. Unlike smartphones and mobile devices, signature pads are focused on high quality signature capture and mainly use more reliable static sampling. Technology evolves ever faster in the current age, bringing new developments and devices to the forensic examiner. During the present thesis, several new and improved devices have appeared (see Table 5.6). After the expiry of the patent on the EMR technology, almost all developers have opted for the EMR technology. Vast improvements in data quality have been observed as a result. This further validates the choice of hardware for the study, as well as proves the relevance of the

Dataset	Year	Alphabe	Signers	Genuine Signatures per signer	Forgerie per genuine signature	Sampling Rate [Hz]	Resolutior	Sessions	Real Signatures	Digitizer	Ref.	Public
Thesis	2020	Latin	3	700 (approx.)	200 (approx.)	200	5080 lpi	46	Yes	Wacom DTU-1141	[338]	No
MOBISIG	2018	Latin	83	45	20	approx. 60 (event-based)	NA	3	Yes	Nexus 9 (Tablet)	[22]	Yes
SCUT-MMSIG Mobile, tablet and inair	2017	Chinese	50	20	20	120 (mobile), 60 (tablet), 120 (camera)	4000 lpi (UGEE)	2	Yes	LG-G3 (Phone), UGEE EX05 (tablet), iPad Air 2 (High Speed Camera)	[349]	Yes
e-BioSign DS1	2016	Latin	65	8	6	Various	Various	2	Yes	Wacom STU-500, Wacom STU-530, Wacom DTU-1031, Samsung ATIV 7, Samsung Galaxy Note 10.1	[572]	Yes
e-BioSign DS2	2016	Latin	81	8	6	Various	Various	2	Yes	Wacom STU-500, Wacom STU-530, Wacom DTU-1031, Samsung ATIV 7, Samsung Galaxy Note 10.1	[572]	No
Sae-Bae and Memon	2015	Latin	180	30	0	NA	NA	6	No	iOS Devices	[495]	No
BiosecurID-SONOF DB Online Subcorpus	2015	Latin	132	16	12	100	5080 lpi	4	Yes	Wacom Intuos3 A4	[178, 205, 208]	Yes
xLongSignDI (Ext. ATVS-SLT DB)	2015	Latin	29	46	20	100	5080 dpi	6	Yes	Wacom Intuos3 A4	[207]	Yes

Table 5.1: Summary of dynamic signature datasets;
lpi - lines per inch; dpi - dots per inch; ppi - pixels per inch; NA - Not Available, Information not found.

Dataset	Year	Alphabe	Signers	Genuine Signatures per signer	Forgerie per genuine signature	Sampling Rate [Hz]	Resolutior	Sessions	Real Signatures	Digitizer	Ref.	Public
Thesis	2020	Latin	3	700 (approx.)	200 (approx.)	200	5080 lpi	46	Yes	Wacom DTU-1141	[338]	No
ATVS-SGNOTE	2014	Latin	25	20	0	NA	NA	2	Yes	Samsung Galaxy Note (Phone)	[364]	Yes
SigWi Comp'13 Online	2013	Japanese	31	42	36	200	50 px/cm (approx. 127 ppi)	NA	NA	HP EliteBook 2730p (PC)	[354]	Yes
ATVS-DooDB Subcorpus 2, Pseudo-signatures	2013	Latin	100	30	20	100	NA	2	No	HTC Touch HD (Phone)	[365]	Yes
Blanco-Gonzalo	2013	Latin	43	60	0	Various	Various	3	NA	Wacom Intuos4, Wacom STU-500, Asus Eee PC Touch, Samsung Gal. Note, BlackBerry Playbook, Apple iPad 2, Samsung Galaxy Tab	[46]	No
Krish	2013	Latin	25	20	0	NA	NA	2	NA	Samsung Galaxy Note (Phone)	[314]	No
GyroSig Db2012	2013	Latin	21	10	NA	100	-	1	Yes	Pen equipped with three-axis accelerometer and gyrometer	[227, 228]	No
ATVS-SSig	2012	Latin	0	17500	0	-	-	-	Synthetic	None	[205, 208]	Yes
PDA-64	2012	Latin	64	30	20	NA	NA	NA	NA	Qtek 2020 ARM (PDA)	[264]	No
Bissig	2011	Latin	NA	20	NA	NA	NA	NA	NA	HTC Desire 3.7' (Phone)	[44]	No
SigComp'11 Dutch	2011	Latin	64	25 (approx.)	11 (approx.)	200	5080 lpi	1	Yes	Wacom Intuos3	[347]	Yes

Table 5.2: Summary of dynamic signature datasets, continued;
lpi - lines per inch; dpi - dots per inch; ppi - pixels per inch; NA - Not Available, Information not found.

Dataset	Year	Alphabe	Signers	Genuine Signatures per signer	Forgerie per genuine signature	Sampling Rate [Hz]	Resolutior	Sessions	Real Signatures	Digitizer	Ref.	Public
Thesis	2020	Latin	3	700 (approx.)	200 (approx.)	200	5080 lpi	46	Yes	Wacom DTU-1141	[338]	No
SigComp'11 Chinese	2011	Chinese	20	19 (approx.)	45 (approx.)	200	5080 lpi	1	Yes	Wacom Intuos3	[347]	Yes
AccSigDb1	2011	Latin	40	10	5	1000	-	1	Yes	Pen equipped with three-axis accelerometer	[65, 228]	No
AccSigDb2 (Extended)	2011	Latin	40	20	10	1000	-	1	Yes	Pen equipped with three-axis accelerometer	[65, 228]	No
SUSIG Blind Subcorpus	2009	Turkish	100	20	10	100	1000 lpi	1	Yes	Wacom Graphire 2	[304]	Yes
SUSIG Visual Subcorpus	2009	Turkish	100	20	10	100	300 dpi	2	Yes	Interlink Electronics ePad-ink	[304]	Yes
BiosecurID	2009	Latin	400	16	12	100	5080 lpi	4	Yes	Wacom Intuos 3		Yes
SigComp'09 NFI-online (NLDCC-online)	2009	Latin	100	12	6	NA	2540 lpi	1	Yes	Wacom Intuos2	[48]	Yes
SIGMA	2008	Malaysiar	213	30	10	200	5080 lpi	2	Yes	Wacom Intuos3	[3]	Yes
BioSecure DS2	2008	Latin	650	30	20	100	5080 lpi	2	NA	Wacom Intuos3	[425]	Yes
BioSecure DS3	2008	Latin, Turkish	713	15	10	100	NA	2	NA	HP iPAQ hx2790	[425]	Yes

Table 5.3: Summary of dynamic signature datasets, continued;
lpi - lines per inch; dpi - dots per inch; ppi - pixels per inch; NA - Not Available, Information not found.

Dataset	Year	Alphabet	Signers	Genuine Signatures per signer	Forgeries per genuine signature	Sampling Rate [Hz]	Resolution	Sessions	Real Signatures	Digitizer	Ref.	Public
Thesis	2020	Latin	3	700 (approx.)	200 (approx.)	200	5080 lpi	46	Yes	Wacom DTU-1141	[338]	No
MBioID	2007	Latin	120 (approx.)	20	0	NA	2540 lpi	2	Yes	Wacom Intuos2	[113]	No
Guest	2006	Latin	274	10-74	0	100 interpolated to 300	500 lpi	Variable	Yes	NA	[233]	No
MyIDea	2005	Latin	104 (approx.)	18	18	100	2540 lpi	3	Yes	Wacom Intuos2	[136]	No
SVC2004	2004	Latin, Chinese	40	20	20	100	2540 lpi	1	No	Wacom Intuos A6	[622]	Yes
MCYT-330 Online Subcorpus	2003	Latin	330	25	25	100	2540 lpi	1	Yes	Wacom Intuos A6	[426]	Yes
MCYT-100 (MCYT-330 Subset)	2003	Latin	100	25	25	100	2540 lpi	1	Yes	Wacom Intuos A6	[426]	Yes
BioMET	2003	Latin	130, 106, 91	15	17	200	2540 lpi	1	Yes	Wacom Intuos2	[212]	Yes
Phillips	1998	Latin	51	30	Variable	200	NA	NA	NA	Philips Advanced Interactive Display (PAID)	[126]	No
Caltech	1998	Latin	56	25	10	NA	NA	2	NA	Camera	[407]	No

Table 5.4: Summary of dynamic signature datasets, continued;
lpi - lines per inch; dpi - dots per inch; ppi - pixels per inch; NA - Not Available, Information not found.

Device	Display	Sensor Type	Resolution	Sampling Rate [Hz]	Pressure	Pen Tilt	Comments
Wacom DTU 1141	LCD 10,6"	EMR	2540 lpi	200	Axial 1024 Levels	No	Business solution Large surface High-end technology
Wacom STU 530	LCD 4,5"	EMR	2540 lpi	200	Axial 1024 Levels	No	Business solution High-end technology
Wacom STU 430	LCD 4,5"	EMR	2540 lpi	200	Axial 1024 Levels	No	Business solution High-end technology
Wacom Intuos Pro	No	EMR	5080 lpi	200	Axial 2048 Levels	60 Levels	Design solution Additional Features
Signotec Sigma	LCD 4"	RTP	5080 lpi	500	Orthogonal 1024 Levels	No	Business solution Low Reproducibility (Pressure) Sampling Interpolation
Signotec Omega	LCD 5"	RTP	5080 lpi	500	Orthogonal 1024 Levels	No	Business solution Low Reproducibility (Pressure) Sampling Interpolation
StepOver naturaSign Pad Color	LCD	NA	1000 dpi	500	NA 512 Levels	No	Business solution Most likely RTP Low Reproducibility (Pressure)
Topaz SignatureGem LCD 1x5	LCD 1x5"	NA	NA	NA	NA	No	Business solution Most likely RTP Low Reproducibility (Pressure)
Topaz SignatureGem Color 5.7	LCD 5.7"	NA	NA	NA	NA	No	Business solution Most likely RTP Low Reproducibility (Pressure)
Apple iPad	LCD 9.7"	CTP	NA	NA	No	No	Large Availability Easy to use Low Sampling Rate No Pressure Information
Apple iPad and Apple Pencil	LCD 9.7"	CTP and Active Pen	NA	60	Axial NA	Yes NA	Design solution Additional Features Expensive Limited battery life

Table 5.5: Summary of Hardware available at the beginning of data acquisition;
RTP - Resistive Touch Panel; CTP - Capacitive Touch Panel; LCD - Liquid Crystal Display; lpi - lines per inch; dpi - dots per inch; ppi - pixels per inch; NA - Not Available, Information not found. The first tablet, marked in green was the one used during this thesis.

Figure 5.1: Photograph of the Wacom DTU-1141 digitizer used for this study

Device	Display	Sensor Type	Resolution	Sampling Rate [Hz]	Pressure	Pen Tilt
Wacom STU 540	LCD 5"	EMR	2540 lpi	200	Axial 1024 Levels	No
Wacom DTH 2452	LCD 24"	EMR	2540 lpi	187(pen) 100 (finger)	Axial 2048 Levels	No
Wacom DTK 1660E	LCD 15,6"	EMR	5080 lpi	187	Axial 8192 Levels	No
Wacom DTH 1152	LCD 10,1"	EMR	2540 lpi	200	Axial 1024 Levels	No
Signotec Sigma	LCD 5"	EMR	2.400 x 2.910 ppi	500	Axial 1024 Levels	No
Signotec Delta	LCD 10,1"	EMR	1.264 x 1.292 dpi	500	Axial 2048 Levels	No
StepOver duraSign Pad US 10.0	LCD	NA	2560 x 2560 dpi	330	Axial 2048 Levels	No
Topaz GemView 16	LCD 15,6"	EMR	5080 lpi	200	Axial 2048 Levels	No
Apple iPad & Apple Pencil 2 (2018)	LCD 9.7"	CTP & Active Pen	NA	60	Axial NA	Yes NA
Microsoft Surface & Surface Slim Pen (2019)	Various	CTP & Active Pen	Various	NA	Axial 4096 Levels	1024 Levels

Table 5.6: Summary of new state-of-the-art hardware
RTP - Resistive Touch Panel; CTP - Capacitive Touch Panel; LCD - Liquid Crystal Display; lpi - lines per inch; dpi - dots per inch; ppi - pixels per inch; NA - Not Available, Information not found

results for forensic practitioners. Wacom developed the EMR solution and is currently the market leader in signature pads. The Wacom tablets are already in use in several businesses [603], and their quality guarantees some longevity of these particular sensors. These dynamic signatures are already in use and may be encountered by forensic scientists in the near future. They are relevant to forensic practitioners and are equally well suited for research. Wacom provides a Forensic Analysis Toolkit (Wacom Signature Scope [605]) free of charge to forensic examiners. This toolkit can be used for data collection and pen data extraction without resorting to external software suites [418].

5.2 Writing Conditions and Data Collection

Writing conditions were standardized for all writers during the data collection for this thesis. The volunteers had to use the Wacom stylus provided with the DTU-1141 tablet. All signatures were made on the tablet surface, solely in digital form. It was purposefully chosen to refrain from adding a sheet of paper on the digitizer display while recording specimens. Many researchers have adopted this method to produce inked and dynamic signatures simultaneously, however the pen interacts differently with the paper surface than with the display. As a result writing conditions reflect paper-based signature behavior, rather than 'pure' tablet surface behavior. By omitting the sheet of paper from the acquisition process, the pen-display interaction of pure dynamic signatures are preserved. Participants all signed (and forged) in a sitting position, while seated in an office chair in front of a medium height desk[3]. They could adjust the chair position and height for comfort. They could turn the signature pad for writing comfort, the tablet

[3] About 1 meter from the floor.

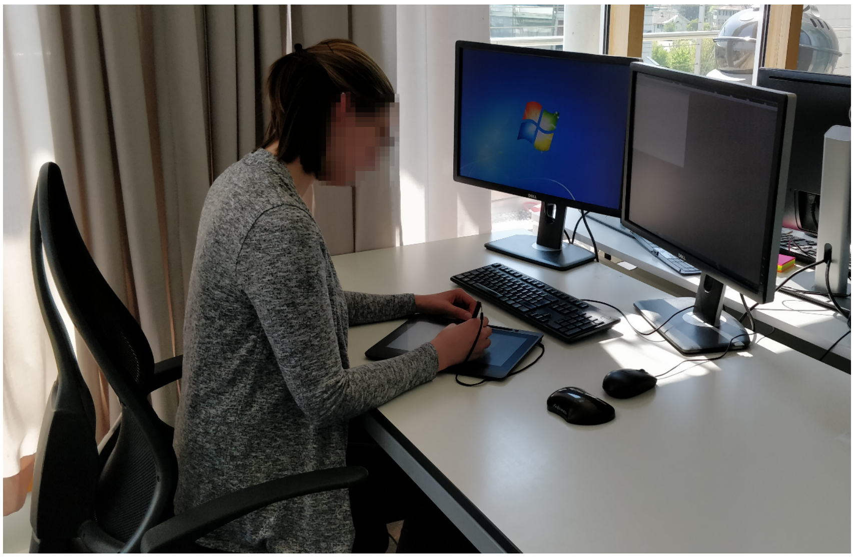


Figure 5.2: Writing conditions and setup for the data acquisition

could however not be inclined[4] (see Figure 5.2). Participants could train using the provided materials before signing.

The subjects proposed in this thesis require three different 'types' of data, coming from 3 different populations. These three types are (1) case-unrelated genuine signatures (or genuine population data), (2) case-related genuine signatures (control or reference data; presumed source data), as well as (3) case-related simulated signatures (forgery population data). The two populations and the data are described in the following paragraphs. A more detailed description can be found in chapter 11.

5.2.1 Case-Related Genuine Signatures / Reference Data

The case-related genuine signatures are produced by the presumed source of a forensic case. They were used to validate the probabilistic methodology and explore subjects relating to variation of dynamic features, feature selection and contemporaneity. The case-related genuine signatures were produced by three participants, which were chosen after a screening of volunteer signatures. They were selected based on signature style[5] [389, 392] and complexity [16, 57, 188]. One signature of each style, which presented the lowest complexity[6], was chosen. 'Low' complexity within the signature type and complexity category were criteria to facilitate forgery[7], which is difficult even for simple signatures [68]. Every one of the participants was asked to provide signatures in irregularly spaced intervals, with decreasing frequency. The aim of this irregular sampling was to study the variation of the signature within and between sessions in several time intervals, such as days, weeks and months. A total of 44 distinct acquisition sessions per participant, over the approximate duration of 18 months, were conducted. The participants had to provide 10 samples per session initially. During the acquisition process, high variance was observed. In order to sufficiently capture variability, the number of acquired signatures per sessions was raised to 20 samples. Every one of the signers produced at least 750 signatures during the acquisition period. The differences in the number of signatures per signer are due to the timing of sampling appointments and small adjustments, which differed slightly for the participants. The detailed description of the number of acquired signatures per signer is provided in Table 5.7.

[4] The tablet had to remain parallel to the desk surface (0° inclination).
[5] Signature style may be classified as either stylized, text-based or mixed.
[6] Complexity was assessed qualitatively by two experienced forensic handwriting examiners.
[7] Freehand forgery and tracing specifically.

	Signature 1	Signature 2	Signature 3
# Signatures	770	750	780
# Sessions with 10 Signatures	15	13	12
# Sessions with 20 Signatures	31	31	33
# Sessions	46	44	45

Table 5.7: Number of available genuine signature per signer

5.2.2 Case-Unrelated Genuine Signatures / Genuine Population

The case-unrelated genuine signatures were acquired to estimate the probabilistic model's parameters. The case-unrelated signatures represent a genuine signature population, as presented for the model comparison process in chapter 3. They are unrelated to the presumed source's signature. They were collected from 23 participants, each providing 20 samples of their genuine signature. Signers produced all of their samples in a single session[8]. The 20 samples made by the writers were used to estimate the variability of their signatures within a session. No attempts at capturing the variation of their signature over time were made. A brief summary of the population can be found in table 5.8.

	Signers	Signatures	Total
#	23	20	460

Table 5.8: Number of available signers and signatures per signer

5.2.3 Case-Related Simulated Signatures / Forgery Population

The case-related simulated signatures were acquired to both estimate model parameters and validate the methodology. The forgeries were case-related, meaning they were specific to one signature 'target'. In the experiments, no general simulated signature population, made by mixing different simulations, was used. Simulated signatures were only acquired for the three participants' case-related genuine signatures. Forgeries were collected as a prize contest, to create competition and a reward for the participants. Forgers were ranked based on shape- and velocity-based similarity to the target genuine signature. The three best forgers won a book from a forensic science collection. The signatures produced were directly inspired by the presumed source's signature, as they were based on a subsample of the case-related signatures. Forgers were shown 6 genuine signatures made on paper. They could then practice forging the signatures for 20 to 30 minutes, either on the signature pad or on paper, before providing the simulated signatures. All forgers had to provide 10 forgeries per signature model they chose to simulate. They could however stop during the process, as well as reject 'botched' forgeries. The simulated signatures were produced by 57 distinct forgers. Not all forgers provided simulated signatures for every case-related genuine signature, several chose to only simulate one case-related genuine signature. All forgers except one chose freehand simulation as their preferred method of forgery. The number of forgeries per signature is detailed in table 5.9.

	Signature 1	Signature 2	Signature 3
# Forgeries	260	400	160
# Forgers	26	40	16

Table 5.9: Number of available forgeries per signer

For further details on the data and its usage inside of the probabilistic model, the interested reader can refer to chapter 11.

5.3 Features and Feature Sets

There are many categories of features, but all can be classified as either continuous (pressure, speed, time to sign, ...) or discrete features (e.g. the number of strokes, number of velocity inversions, ...). In

[8] Within an approximately 20-minute-long session.

this study, it was decided to exclusively use continuous global features. Discrete features may be equally valuable, but they require specific discrete probability distribution. Combining these with continuous variables, alas, harbors many complex statistical challenges lying outside of the scope of this thesis. Global features are an opportune starting point for statistical signature models. A set of 60 features was used in the study. Tables 5.10 and 5.11 show a complete description and classification of the features into different types (see Chapter 2.2.3), according to what kind of characteristic is being described. These characteristics were picked because of their simple extraction, intelligibility and complementarity to shape-based and graphical features. Features are used in several ways, either for univariate statistics[9] as features or for multivariate statistics[10] as feature sets. By the terminology *feature set*, a simultaneous and multivariate use of the features is implied. In general, we chose feature sets with three features (of dimension $p = 3$), however in some parts of this thesis we varied the number of variables in the sets. With the 60 features used in this thesis, there are a total of 34'220 distinct trivariate[11] feature sets. There are still many other features to be used within the model, several authors [118, 273, 451, 460] report hundreds of possible features. Possibilities for extensions, recombinations and novel measurements are nearly infinite. This thesis focuses on comprehensible features with a physical interpretation, meaning features that will make sense to practitioners and whose equivalents can be seen in the signature image.

[9] Statistical techniques using only one variable.
[10] Statistical techniques using multiple variables.
[11] Dimensionality of the feature vector $p = 3$.

Index	**Feature Name**	**Feature Description**	**Category**
1	Totaltime	Duration of the Signature, including pen lifts and inking strokes	Time-related
2	Uptime	Cumulated duration of pen lifts	Time-related
3	Downtime	Cumulated duration of inking strokes	Time-related
4	UpTot	Ratio of pen lift duration to signature duration	Time-related
5	DownTot	Ratio of inking stroke duration to signature duration	Time-related
6	DownUp	Ratio of pen lift duration to inking stroke duration	Time-related
7	TotLength	Cumulated length of the signature trajectory, including inking strokes and pen lifts	Spatial
8	DownLength	Cumulated length of the inking strokes	Spatial
9	UpLength	Cumulated length of the pen lift strokes	Spatial
10	Width	Horizontal distance from leftmost to rightmost pixel	Spatial
11	Height	Vertical distance from highest to lowest pixel	Spatial
12	WHRatio	Width to Height Ratio	Spatial
13	XY_bar	Average distance to centroid (center of gravity of the XY coordinates)	Spatial
14	XY_var	Variance of the distance to centroid	Spatial
15	XY_max	Maximum distance to centroid	Spatial
16	P_bar	Average (axial) pen pressure	Pressure-related
17	P_var	Variance of (axial) pen pressure	Pressure-related
18	P_max	Maximum (axial) pen pressure	Pressure-related
19	dp1_bar	Average of the first differential of pressure (time)	Pressure-related
20	dp1_var	Variance of the first differential of pressure (time)	Pressure-related
21	dp1_max	Maximum of the first differential of pressure (time)	Pressure-related
22	dp2_bar	Average of the second differential of pressure (time)	Pressure-related
23	dp2_var	Variance of the second differential of pressure (time)	Pressure-related
24	dp2_max	Maximum of the second differential of pressure (time)	Pressure-related
25	dp3_bar	Average of the third differential of pressure (time)	Pressure-related
26	dp3_var	Variance of the third differential of pressure (time)	Pressure-related
27	dp3_max	Maximum of the third differential of pressure (time)	Pressure-related
28	dt1_bar	Average of the tangential velocity	Velocity
29	dt1_var	Variance of the tangential velocity	Velocity
30	dt1_max	Maximum of the tangential velocity	Velocity

Table 5.10: Description of the continuous global features used in this thesis.

Index	Feature Name	Feature Description	Category
31	dx1_bar	Average of the horizontal velocity	Velocity
32	dx1_var	Variance of the horizontal velocity	Velocity
33	dx1_max	Maximum of the horizontal velocity	Velocity
34	dy1_bar	Average of the vertical velocity	Velocity
35	dy1_var	Variance of the vertical velocity	Velocity
36	dy1_max	Maximum of the vertical velocity	Velocity
37	dt2_bar	Average of the acceleration	Acceleration
38	dt2_var	Variance of the acceleration	Acceleration
39	dt2_max	Maximum of the acceleration	Acceleration
40	dx2_bar	Average of the horizontal acceleration	Acceleration
41	dx2_var	Variance of the horizontal acceleration	Acceleration
42	dx2_max	Maximum of the horizontal acceleration	Acceleration
43	dy2_bar	Average of the vertical acceleration	Acceleration
44	dy2_var	Variance of the vertical acceleration	Acceleration
45	dy2_max	Maximum of the vertical acceleration	Acceleration
46	dt3_bar	Average of the jerk (third differential of position)	Jerk
47	dt3_var	Variance of the jerk (third differential of position)	Jerk
48	dt3_max	Maximum of the jerk (third differential of position)	Jerk
49	dx3_bar	Average of the horizontal jerk	Jerk
50	dx3_var	Variance of the horizontal jerk	Jerk
51	dx3_max	Maximum of the horizontal jerk	Jerk
52	dy3_bar	Average of the vertical jerk	Jerk
53	dy3_var	Variance of the vertical jerk	Jerk
54	dy3_max	Maximum of the vertical jerk	Jerk
55	TVD_bar	Average trajectory vector angle to a horizontal line	Writing Angle
56	TVD_var	Variance of trajectory vector angle to a horizontal line	Writing Angle
57	TVD_max	Maximum of trajectory vector angle to a horizontal line	Writing Angle
58	TAD_bar	Average of the acceleration direction to a horizontal line	Writing Angle
59	TAD_var	Variance of the acceleration direction to a horizontal line	Writing Angle
60	TAD_max	Maximum of the acceleration direction to a horizontal line	Writing Angle

Table 5.11: Description of the continuous global features used in this thesis, continued.

6 Evaluation of Signature Evidence

In this chapter, the statistical model applied to the signature forgery problem is described. A detailed account along with justification of the divergence from more classic models is given in chapters 8.2 and 11. This modified model is used for the description of both the genuine and forged signature populations in the thesis. Albeit the model being the same, the population parameters are specific to each population type, either genuine or forged. This statistical model, in conjunction with the population specific parameters, is used for the Bayes Factor calculation, as described in chapter 3.

6.1 Scope of the Evaluative Framework and Hypotheses

A statistical model is a simplified representation of reality. As the statistician George E. P. Box put it *"All models are wrong, but some are useful"* [55]. Models can be used to evaluate evidence in forensic science with respect to the considered populations. Evidence evaluation works within a frame defined by some hypotheses of interest. Hypotheses[1] are statements that can either be true or false [155]. Forensic evidence evaluation only works within the confines of the considered hypotheses, implicitly ignoring any that are not considered. Careful consideration is necessary to correctly and exhaustively frame a problem. The considered hypotheses define the *'meaning'* of the BF and are essential to its comprehension. A basic principle of case assessment and interpretation [156] is to consider at least two mutually exclusive alternative hypotheses. Before moving forward, the hypotheses of interest need to be defined. In dynamic signature casework, the court is interested in whether or not a presumed source made a signature *and* if it is an original. The subtle difficulties in the definition of these hypotheses are the originality checking as well as the specificity to a source, which makes the task a verification of identity rather than an individualization. Assessment of originality is generally taken care of before probabilistically assessing the source of the signature. A signature can be non-authentic even if it was produced by the real presumed signer if it is not an original. A tracing or copy of signature for example may have been made by Person A, but has only been reproduced onto the document. A simple pair of hypotheses focusing on the source might look as follows:

H_p 'Person A made the signature';

H_d 'Someone other than Person A made the signature'.

This pair of hypotheses lacks a sufficient level of detail and might misinform the decision maker. Therefore, forensic examiners most often use a combination of three hypotheses: the signature may be

[1] Sometimes called propositions.

genuine, may be disguised or may be a simulated signature. Simulated signatures include all types and methods of forgery, including tracings, stamps, freehand simulations, ... Sincerity[2] and specificity to a presumed source complicate the FHE's task. In an effort to limit the complexity of the model, only forged signatures are included while signature disguise is excluded from the assessments. This choice was made for simplicity and development of a working model for this research. Before applying such simplifications to casework, assumptions and allegations, such as if the person needed to provide further credentials and an ID check or the interest of disguising a signature, need to be checked for compatibility. Therefore, the simplification made here in this thesis is not always warranted in casework. In the majority of cases, signature forgery is however the most likely and most important alternative to consider. Future research may extend the current model to include signature disguise, or extend the model to different forgery behaviors. Further discussion of the formulation of relevant hypotheses can be found in chapter 11. In the reduced form used in this document, the following hypotheses for signature examination can be formulated:

H_1 'The questioned signature is (sincere) genuine and was made by Person A';

H_2 'The questioned signature is (insincere) a simulated signature y of signature x_A made by someone other than Person A'.

A particular point to be noted in the H_2 hypothesis is that the questioned signature is a simulation *of Signature A* rather than a random genuine or simulated signature made by someone else. The difference lies in the specificity of the statement. It signifies that not only is the signature of a different type (simulated rather than genuine), the simulation is inspired by *Signature A*. This specificity differentiates signature examination from most other evaluative forensic disciplines. Rather than the alternative being just part of genuine (sincere) non-specific population, a forgery is an intentionally similar product, through some copying or reproducing mechanism [202]. The alternative population is thus specific and depends on the presumed source, in our case Person A. To be even more precise, the fact finder may be interested in a particular type of forgery (such as tracing, freehand simulation, ...) which could be added to the alternative proposition H_2 (or used to create a nested[3] hypothesis considering different types of forgery). The data used in the model has to reflect the hypotheses properly in order for the model to produce rational and relevant results.

These hypotheses define the scope and application of the model, as well as the context of the BF. The context is important, as the value of the evidence can only be correctly interpreted within the defined hypotheses. It answers one specific question given by the hypotheses. The model is used to determine under which of these hypotheses the observations made on the questioned signature are more likely. This quantitative assessment - a Bayes Factor - is the 'inferential strength' of the support: the value of the evidence. The Bayes Factor expresses how much more likely the updated belief should be, given the considered hypotheses.

[2] Genuine signatures are considered 'sincere' while disguised and forged signatures are examples of 'insincere' signatures.

[3] A nested hypothesis is an ensemble of a parent hypothesis that branches into several sub-hypotheses.

6.2 The Signature Model

6.2.1 Statistical modelling

The signature evaluation model will use three distinct populations, presented in chapters 5 and 11. The population of signers (or forgers[4]) will be shortened to Z_{ij}, with j representing a signature specimen and i the signer. Additionally, it is assumed that the signer from the populations (G and S, for Genuine and Simulated signatures respectively) produce signatures following a Normal distribution, with unknown parameters θ and W. The feature vector ψ_i can be described by the mean vector θ_i for every signer i and the covariance matrix W, which is assumed to follow a Wishart distribution. x is the presumed source's signature (case-related genuine signature, control sample) and y the questioned signature (unknown source signature, recovered specimen). Each population follows the same model in this framework, but is differentiated by its prior parameters, based on the collected population data. This different starting point makes the model selection process possible.

An essential assumption in signature examination relates to the difference in natural variation between genuine and forged signatures. Genuine signatures and forgeries are supposed to show different features (their value and statistical 'mean') but also have different variation (variance) due to the motor processes engaged during their production. The difference may stem from the use of different processes while signing and copying a signature [209, 615]. Therefore, one may expect forged signatures to lack 'natural' variation, while being pictorially almost identical to the other signatures. Lack of variation is also an indication of a 'false' signature, as some forgery methods generate carbon copies of the original. As a result of these assumptions, signature evaluation frameworks cannot consider genuine signatures and forgeries to be part of a same, homogenous population. It is essential for logical coherence to separate both populations and use model comparison procedures to evaluate the evidence. Hence, the genuine signature population G_{ij} and the simulated signature population S_{kl} will be used separately in the Bayes' Factor formula (c.f. Eqn. 6.7). Each will have their own parameters (Eqns. 6.1 and 6.4), as well as separate prior (and hyper-prior[5]) distributions.

$$G_{i,j} \sim \mathcal{N}(\theta_{gi},\, W_{gi}) \tag{6.1}$$
$$\theta_{g,i} \sim \mathcal{N}(\mu_g,\, B_g) \tag{6.2}$$
$$W_{g,i} \sim \mathcal{W}i(U_g, \nu_g)\,. \tag{6.3}$$

$$S_{k,l} \sim \mathcal{N}(\theta_{sk},\, W_{sk}) \tag{6.4}$$
$$\theta_{s,k} \sim \mathcal{N}(\mu_s,\, B_s) \tag{6.5}$$
$$W_{s,k} \sim \mathcal{W}i(U_s, \nu_s)\,. \tag{6.6}$$

The Bayes Factor is obtained as explained in chapter 3. It determines under which population model the evidence is more likely to have occurred and guides the decision maker by defining the inferential strength and 'direction' to tip toward. The BF formula for the signature case can be found below in Eqn. 6.7. This BF conveys the 'strength' of the evidence and is an integral part of the conclusions to be reported in casework.

[4] In the following paragraphs, 'forgers' will be omitted for legibility.
[5] Hyper-priors are prior distributions for distribution parameters situated higher in the model's hierarchy.

		Model Conclusion	
		Support H_1	Support H_2
Ground Truth	H_1 True	Correct	Misleading toward H_2
	H_2 True	Misleading toward H_1	Correct

Table 6.1: Summary of possible outcomes for BF accuracy and RME

$$BF = \frac{f(y, x|G_{ij}, H_1, I)}{f(y, x|S_{kl}, H_2, I)} \tag{6.7}$$

6.2.2 Evaluation & performance

The model and the feature sets are evaluated through computer simulations mirroring Bayesian assessment in casework. Every simulation is constructed as a case, with some presumed source reference materials, a single questioned signature and the available background information. Multiple pseudo-cases are generated by repeated random sampling. These simulations are grouped into random trials. Within one random trial, presumed source materials and background data remain the same, while the questioned data changes. Each random trial contains multiple simulations, varying between 100 to 200 each. Between random trials, the presumed source data is randomized. Multiple random trials are performed and the results are summarized and studied. Averages and variances of the random trials are measured and stored. To fully explore one experimental condition, multiple random trials are necessary. The repeated random sampling plays the same role as a bootstrap procedure [107], allowing to virtually upscale the size of the dataset. An experiment is constructed by associating multiple hundreds or thousands of random trials per experimental condition. This computational statistics technique allows one to study model characteristics even with limited data. The results are then reported with the concepts presented in the following subsection.

6.2.2.1 Accuracy, Rates of Misleading Evidence, Reproducibility and Inferential Strength

Accuracy, rates of misleading evidence (RME), reproducibility and inferential strength are defined by Aitken et al. [11] and Taroni et al. [558]. A short formulaic description can be found in the equations below:

$$Accuracy = \frac{\#\text{Correct BFs}}{\#\text{Cases}}, \tag{6.8}$$

$$RME = \frac{\#\text{Misleading BFs}}{\#\text{Cases}}, \tag{6.9}$$

$$RME_{H_1} = \frac{\#\text{BFs misleading towards } H_1}{\#\text{Cases where } H_2 \text{ is true}}, \tag{6.10}$$

$$RME_{H_2} = \frac{\#\text{BFs misleading towards } H_2}{\#\text{Cases where } H_1 \text{ is true}}. \tag{6.11}$$

The performance of the method can be measured using several criteria. In this thesis, three major types of criteria are going to be used. First is the *accuracy* (Eqn. 6.8), which reflects how reliable the BFs are in supporting the correct proposition. In the accuracy category, there are also the rates of misleading evidence (RME) towards either hypothesis, which show if the model is biased toward a hypothesis. Accuracy measurements ignore inferential strength and are limited to controlling the final 'verdict' of the model, in this case which hypothesis is supported. They are conceptually related to the much used confusion matrix in biometrics. The second category is the *inferential strength*, measured through the numerical BF value. In this way, the model's production of strong or weak inferential strength for correct and misleading statements can be measured. When the same case data is used for different conditions, a direct comparison between the BF values can be used to determine differences in BF scaling. There are measures, like calibration (C_{llr}, the calibrated log-loss) that are being used to measure misleading evidence in a more precise fashion. Finally, we look at the *reproducibility*[6] of the results through the change of questioned and reference samples. Repeated random sampling (RRS) is used to vary the materials and produce many random trials, akin to the k-fold Cross Validation (CV) and bootstrap procedure. With this procedure, every random trial produces an accuracy and variance value. The variance of these values over all the trials represents reproducibility. While repeating the method on different samples, the reproducibility of conclusions and the importance of the sample quality, contemporaneity and size for the BF can be assessed. Reproducibility is especially important for generalization of the model and justifying that the same method can be applied to different cases and data sets.

A number of additional criteria for validation have been established [241, 380]. However accuracy, reproducibility and inferential strength measurements can be used in a first step to support that the method is reliable and valid. When comparing methods and feature sets in this document, the ideal characteristics are high accuracy, low variance, as well as low misleading WoE values. A method fulfilling all three of these criteria is accurate, reliable and reproducible. Such a method would consequently be justifiable and credible for court use. Currently, there are no 'threshold values' or strict numerical criteria for the admissibility of a method in court. The reliance on a model is based on hearings and personal judgment. However, professional deontology, as well as legal rules require experts to be transparent on error rates and limitations. It is my firm belief that these criteria, along with the presentation of the data and the model, reflect these values and fulfil the criteria for court admissibility.

6.2.2.2 Calibration

There are alternative ways to evaluate model performance. Research in automated forensic systems[7] has brought forward the calibration procedure [62, 110, 240, 380, 470], the log-likelihood ratio cost C_{llr}, as well as criteria such as the Empirical Cross Entropy (ECE). A precise definition of the mathematical concepts involved have been described in Aitken et al. [11]. A more intuitive and simplified description is given here: The main idea of calibration is to assign a cost function that measures the models' performance with regard to a known-source test-set. A posterior probability is calculated and compared to the ground truth, to penalize strongly misleading results. The results of the calibration procedure describe the methods' performance through two separate values: calibration and discrimination loss. Calibration loss measures the robustness of the BF, whereas discrimination loss benchmarks the discriminative power of the method. Both of these factors are very informative in theory, but can be difficult to interpret in practice.

Calibration procedures have several disadvantages for forensic science. A first point is that the calibration measures and output depend on the 'test' dataset used to calibrate on. Merely measuring

[6]Reproducibility is used as an extension of repeatability. An experiment is repeatable if the same operator in the same conditions obtains the same results. The experiment is reproducible if a different operator in the same conditions obtains the same results. As the statistical procedure is mathematically defined, given the same input data, the results are perfectly reproducible.

[7]E.g. Fingerprinting, Facial Recognition, Vocal Recognition, Gait Recognition.

calibration for the dataset does not change the BFs, and can be an informative and important step in the validation of a model. However, some authors advocate for the correction of calibration loss, therefore adapting the value of the evidence given on the results of the calibration procedure [404]. This *calibration*[8] of BF values alters them and either 'shrinks' or 'grows' them [404] to minimize log-likelihood ratio cost. Calibration essentially does not affect the accuracy of a system, but it affects the inferential strength produced. It alters the original BF values and binds them to the calibration dataset. Essentially, calibrating a system rescales and limits the BF values produced, by transposing them from an absolute to a relative scale.

As an analogy to the transposition from absolute to relative scale, one may compare the BF assessment to a teacher grading exams. The teacher will assign an 'absolute' grade to exams, which reflects the performance of the student. If the teacher wishes to calibrate, he will start by assigning his absolute grade, and decide passes and failures of the exam. The actual calibration happens when the quota of passes is unsatisfactory. The teacher will then regrade (rescale) his absolute grades to a new (relative) scale, reevaluating the absolute grades in comparison to another class of students. Some of the graded students may end up with a 'relative' grade that is either higher or lower than their previous 'absolute' grade, based on the new grading (calibration) scheme applied by the teacher. The calibrated result does not reflect the true[9] worth of the exam, but it helps respecting a condition (here the comparability of two different classes, or years).

Calibration of probabilities essentially changes the reference system, shifting it from an absolute to a relative scale. This poses several methodological issues, especially when comparing evaluations. As their values are no longer in a common scale, one item may be disproportionately weighed. Every evaluation would be relative to the calibration dataset. It also shifts a lot of importance onto the selection of a 'good' calibration dataset. This poses a plethora of new questions on how to choose good calibration data and the under- and overfitting debate in statistics. On a more philosophical note, comparing the worth of a prediction using the final outcome does not truly reflect the *quality of the prediction.* Take for example a decision making process: Decisions are made using the available information at the moment of the decision, not in hindsight. When gambling on a roulette board, an experienced player knows that the expected gain is highest when betting on color, rather than on numbers or a combination of both. Winning on a streak of luck while betting on a number does not change the fact that the bet was not a rational decision to make at that time, before being aware of the outcome. Kahnemann [295] describes such problems with decision making and the perception of 'optimal decisions' against truly rational ones. Additionally, there is a fundamental issue with the methodology of the calibration procedure. Calibration can be seen as an optimization of the inferential strength over all possible prior probabilities. This means that posterior probabilities need to be generated before calibrating. This in turn means that a scientist chooses a 'rational' prior distribution, most often a uniform[10]. Mathematically, the choice of a flat prior distribution makes the posterior reflect the Bayes Factor. Conceptually, this choice appears problematic, in my opinion. The calibration procedure shifts the value of the evidence to minimize the impact of the prior probability. This means that the system is designed to minimize the effect of the priors of the court and making the evidence evaluation the predominant element. The resulting BF may indeed be mathematically 'optimal', but *cheats* the 'libre appréciation des preuves' principle, anchored in the Swiss legal system, by minimizing the decision maker's impact in the process.

The use of calibration procedures to measure method performance may be useful for many forensic fields. Measuring calibration is a great asset for studying and validating methods [11, 380]. Calibration,

[8] As opposed to the usage of the term made by biometricians, who use calibration to transpose match scores to BFs.

[9] On an absolute scale.

[10] Also known as 'weak' or 'flat' prior in Bayesian Inference.

as in shifting probabilities to a different scale, itself seems disputable on various aspects. Some of these issues include overfitting[11] and loss of generality, choosing an 'optimal' function for the proper scoring rule and *warping* the BF. For signature evidence in particular, choosing a calibration dataset is a difficult task. The calibration data would have to be formed by source-specific forgeries, else it would fail because of being an unrepresentative overgeneralization. In light of these considerations and difficulties, no attempt at measuring calibration was made during the thesis.

6.3 Limitations, perspectives and further use-cases

6.3.1 Limitations

The signature model adopted in this research project represents one possibility to model signature data. Any modelling choice can be discussed as for theoretical relevance and performance on a relevant dataset. Many aspects of the model have underlying assumptions and are based on the current theoretical and statistical grasp of the problem. The important aspect is to be transparent on such assumptions. Research in the future may deepen our knowledge on the generative mechanisms of signatures, variation and ageing, helping in the development of novel movement based models. Examples of such models can already be found [462]. The model proposed in this thesis has several limitations, which are briefly stated in the following paragraphs.

First, the model assumes the Normality of the data. As a general rule, the data may, however, not be distributed Normally. In some features (see chapter 13) multimodality was observed, even if the simplification generally held. The Normality assumption however holds best for large sample sizes. For small sample sizes, as are available in forensic science [11], distributions are often heavy-tailed and may require a different distribution (Student-t, lognormal, ...). In signature examination, reference materials rarely exceed 30 signatures. Further research into Bayesian approaches with the use of artificially generated or diversified data may offer a perspective to apply more sophisticated models. Possible solutions, such as distortion models and synthetically generated data [119, 205, 208, 423] have already been proposed in biometrics, but lack an adaptation for forensic science [579].

Second, the used population (prior) model is very simple. It is based solely on within-writer variation, ignoring between-writer variation. Other forensic models [6, 56] account for both of these, which may be an adequate extension to the model. Along the same lines, the population model constraint on constant within-writer variation can be relaxed, as proposed by Bozza et al. [56] in their model for handwriting evidence. These changes may make the model more complex, however also closer to reality. Being closer to the ground truth may significantly improve model accuracy, but also has a large impact on data requirements and computational performance. The disadvantages in this approach include the significant time loss and increase in computational load due to Bayesian approximation techniques (MCMC and computation of the marginal likelihood by Importance Sampling or Chib's Method, ...).

6.3.2 Perspectives and use-cases

The presented methodology is first and foremost extensible. The basic construct respects the theoretical framework proposed in handwriting and signature examination. The model and the probability distribution (priors, hyper-priors[12], likelihoods) can be changed to better fit the data in the different applications.

[11] Through the linkage of results to a dataset, entailing loss of generality and applicability of the model.

[12] A hyper-prior is a probability distribution of a parameter in a hierarchical model's prior. The model is considered hierarchical, because it has multiple layers of parameters and hyper-parameters.

Further extensions may add either parameters (such as adding a model for between writer variation), or hierarchical layers (allowing for variable within writer variation). These changes may further improve and extend the methodology. Albeit limited, this form of the model has clear advantages in terms of speed, computational load and approximation. This form of model does not require any approximate Bayesian inference techniques, as all elements have analytical solutions. Therefore no computationally difficult (and slow) approximation techniques such as Markov Chain Monte Carlo (MCMC) simulations along with bridge sampling (or other equivalent methods) need to be used. These techniques often lack good reproducibility and require tuning of parameters for the approximation to work well. Using conjugate distributions, as proposed here, allows for perfectly reproducible results. The proposed signature model is fast, adaptable and reproducible. As such, it provides a perfect first step for exploration, but also for applications requiring quick results.

The presented model was specifically developed for use in questioned signature cases, where freehand forgery is suspected. This type of forgery assumes some knowledge about the presumed source's signature, as the point is to produce a signature as close as possible to the original. Galbally et al. [202] referred to this type of behavior as mimicry, which is different from what most other forensic disciplines deal with. Nevertheless, the signature model, with its dependency assumptions between the questioned and reference materials (see chapter 11) can also be applied to similar problems. In forensic science, such dependencies may also appear in cases with fake evidence or disguises, in particular when a suspect claims to have been framed. In facial comparison tasks, this may be the case if a mask, a virtual 3D model or an image of the person is used. Other examples include the use of synthetic fingerprints, using the same clothing or wearing disguises, voice changers and imitation, etc. The potential forensic applications are very specific, but may be useful in these cases.

Another aspect is the use of two different populations to assess parameters and Bayes Factors. The model permits interchange and combination of populations in assessments. This is essential in the case where reference and questioned materials are not independent. Additionally, the statistical elements used to generate the Bayes Factor can be changed (comparing for example variance instead of means, or both) and adapting both hypotheses and data. The model is hence a good way to study influences of time, specificity and data on the signature examination problem. The differing populations assumptions may be valid and important for other fields, where significant differences in populations exist. A possible application may be different manufacturing procedures in narcotics. Narcotics created from two different chemical processes may be identified through the use of two separate populations and be tested either based on their variation or mean values. When there is an independence assumption involved, only a single population would be used instead. As a result, there may be two or more possible alternative populations, which all in turn lead to a different inferential strength.

The model captures and differentiates 'common root' type scenarios with its probabilistic structure and inbuilt dependencies. Therefore, it is especially useful for differentiating subclasses that share some common properties, or identity theft or obfuscation scenarios. Given all these considerations, the probabilistic model could find more applications in the signature examination field, specifically with large databases. Because of its simplicity, it is fast and easily implementable. It could also be used for quick pre-case assessments, and cases with very little available data. Other possibilities include biometric applications, or even for investigative or research purposes. Further, the model can be applied to specific situations in other forensic fields that also feature mimicry or disguise behavior, including facial, voice or fingerprint comparison.

7 Research Propositions

7.1 Proposition 1 - The Novelty of Dynamic Signature Examination

Dynamic signature examination is novel for forensic scientists in both technological and data terms. Dynamic signatures are complex constructs involving multiple research fields (e.g. computer science, engineering, biometrics, cryptography). The technology used to record them is capable of measuring dynamic and kinematic signature data, which was previously inaccessible to forensic examiners. These developments have transformed dynamic signatures into complex multivariate data constructs. The new developments and fundamental changes to dynamic signatures warrant a revision and extension of our knowledge on signature dynamics and data-driven evaluation of signature evidence. It is important to raise awareness, provide new insights and encourage research in the field.

As a matter of fact, dynamic signatures have made large leaps due to technological evolution recently. The technology involved in their capture and analysis has been evolving ever more quickly, leading to a recent gain in popularity and application in e-commerce and e-government. Dynamic signatures have come to the forensic scientist's attention recently and continue to gain exposure. Devices such as smartphones, remote signing via the internet and the diverse means of signing have changed the context of signature examination. This proposition concerns the many novel challenges and the gaps in our knowledge related to dynamic signature examination. The sheer speed in the evolution of standards and technology, as well as materials and applications calls for continuous reviewing and updating of existing methods and procedures for data collection and their evaluation. Important elements and standards on dynamic signatures are continuously appearing and evolving. Recent examples are the standardization of biometric signature data, signature formats, encryption or system evaluation, which will impact both the data recording as well as examination [277, 278]. The rapid evolution of dynamic signatures has made it difficult for the forensic community to follow suit. The quantity of data produced has greatly increased, so novel methods also need to deal with time constraints, computational loads and large volumes of data. The collaboration and exchange between forensic science and other connected research fields will be necessary to deal with upcoming issues [287, 379, 381].

Dynamic signatures have already revealed a plethora of new challenges. The most important challenges for forensic cases fall into several rough categories, namely:

- Mechanical and human interaction issues (e.g. writing instruments, depreciation[1] of hardware,

[1] Wear and tear, material fatigue.

surface physics[2], human-sensor interaction, digital ink and visualization);

- Technological, software and hardware issues (e.g. new technologies, diversity of hardware, measurement, data pre-treatment, life-cycle of soft- and hardware);
- Interoperability and comparability issues (e.g. comparability between different signature digitizers, different means of creating signatures);
- Knowledge gaps (e.g. dynamic feature variation, feature selection, engineering, sensor technology, movement science and neuroscience, statistics);
- Methodological issues (e.g. collection of control materials, choice of probabilistic model for evaluative reporting, digital signatures and PKI[3], metadata management, collaboration with digital forensics).

Examiners need to be equipped to deal with the data under its new form and use the full extent of data available to them. Forensic examiners are used to ink-paper interactions and shape-based comparison, while signature dynamics are mostly used to assess fluency. Some features (such as in-air strokes or timing) are completely unknown, while others have not been used for probabilistic conclusions. Examination methodology needs to be revised to coherently measure features, use them in probabilistic models and interpret the results. To this effect, they need to grasp the differences in features, variation and production of the signatures, as well as acquire a methodology to integrate the newly available measurements effectively. The methodology and knowledge categories are therefore more imminently needed in casework, with the other issues being less pressing. Methodology and knowledge issues may pose issues in presenting and justifying signature evidence testimony in general. Interoperability and wear issues will be an obstacle to examiners in a few years, when there are many additional devices on the market and the current ones have had time to age. The novelty of dynamic signatures and the challenges accompanying them should be thoroughly examined and researched.

To address the proposition adequately, a four-step process is proposed:

- Review available knowledge on dynamic signatures, methodology and applications of statistics in signature examination.
- Identify and prioritize areas in need of further research.
- Explore the relevant areas.
- Distribute and share knowledge through academic publications.

In order to determine what forensic scientists should focus on, it should first be established what is known. A thorough review of knowledge relevant to the examination of dynamic signatures is proposed as a first step to identify strengths in existing methodologies and expand knowledge, as well as pinpoint knowledge gaps. This is a gargantuan task that will only be possible to complete over a lengthy period. First, all available knowledge on dynamic data, should be reviewed and summarized to be easily accessible to the forensic community. Second, the two sectors identified as fundamental, methodology and knowledge (especially feature variation), should be explored. Third, the forensic community should be encouraged to produce knowledge on these subjects by researching and collaborating with other research fields such as computer science and biometrics. Finally, it is of the essence to share and distribute the knowledge to both the end-users and forensic examiners. A first step toward this goal was taken in this thesis. The results are presented in chapters 8.1 and 10.

[2]Pen-pad interaction.

[3]Public Key Infrastructure. A cryptographic infrastructure relying on a public and private key pair.

7.2 Proposition 2 - The Methodological Changes Warranted by Dynamic Signature Examination

The analysis of the best evidence available is essential for reliability and admissibility of evidence. Dynamic signatures are files containing large amounts of complex multivariate data. Traditional pattern matching techniques largely ignore the dynamic and kinematic measurements made during the recording of the signature. Previous methodology does not permit the use of the dynamic signature to its full potential and therefore arguably does not constitute the 'best available evidence'. A probabilistic evaluation methodology for dynamic signature assessment provides a framework for the analysis of this additional information and may support examiners in quantifying evidential value, as well as in communicating reliable and justifiable conclusions. Its adaptation to dynamic signatures is an essential step to guarantee coherence and admissibility.

Dynamic signatures are, contrary to most signatures, not natively in image form. They are complex data lists, containing complementary information to the static signature image. Dynamic signature evidence has faced admissibility issues[4] on the grounds of the use of low quality images instead of the signature data. In much the same way, reproducibility and features may become criteria to exclude or favor some piece of evidence. Forensic examiners need to be equipped to go beyond their usual qualitative pattern matching techniques in order to produce admissible testimony. Dynamic signatures have created the need for a methodological way of data treatment for novel features. They are described by patterns, measurements and signals, which are all forms of numerical data. This information is well suited for statistical approaches. The legal community has often challenged forensic scientists on the lack of rigorous statistical procedures and use of empirical data. The integration of a rigorous statistical framework, complementing existing pattern matching methods can help forensic examiners face criticism to their field, as well as justify their evaluations through empirical data and a logical approach. Forensic examiners already have working methodologies and standard operating procedures minimizing bias and regulating the examination process. They are currently in need of an adequate probabilistic framework to assess measurements and numerical features from signatures, complementing shape-based and pattern matching techniques. The Bayesian approach provides both a means to quantify the probabilities involved, as well as a logical framework to combine them with the qualitative pattern-matching information. Nevertheless, this aspect is very complex and out of scope of the present thesis.

The second research proposition is a complex construct of theoretical and practical aspects. Mainly, it aims to show that dynamic signatures contain information beyond 'wet' signatures [244, 392] and should be used to their fullest potential. Additionally, this difference in data treatment and methodology, may also have legal implications, specifically on the admissibility of signature evidence. Finally, the step towards numerical data is an opportunity to integrate probabilistic evidence evaluation, as well as rigorous statistical procedures into forensic signature examination.

To support this research proposition, several aspects need to be shown:

- the difference between 'wet' signatures and dynamic signatures;
- the need to utilize the entirety of the available dynamic signature data;
- the applicability of a probabilistic methodology to a questioned signature case;
- the role of empirical data, statistical model and 'experience' in signature examination;
- the performance of a probabilistic methodology in operational conditions;

[4]See Biles v. AFLAC case, explained in chapter 8.

- the reproducibility of the probabilistic assessments;
- the flexibility of the model to different signatures and feature sets.

Evidence for each of these points is going to be provided with three different elements: a description and illustration of both types of signature data in a literature review, a case study on the admissibility of dynamic signature evidence and finally a validation study of the probabilistic model described in chapter 8. A validation study operates on known source data and compares the obtained results with an expected value, based on the known origin of the data. On the known source data, results can be validated and performance can be quantified. To determine reproducibility, the study uses large scale computer simulations. In these simulations, the many pseudo-cases in varying conditions are simulated. These simulations recreate conditions similar to regular casework (one questioned and n reference signatures) and using the proposed model and four distinct feature sets per signature, to generate Bayes' Factors as measure for the value of the evidence. A procedure called repeated random sampling (RRS), in analogy to k-fold cross validation, was used to vary and multiply the cases and permitted the measurement reproducibility. The validation study needs to respect operational conditions, here we opted for limited quantities of reference materials (5 to 30 signatures). In summary, the steps taken to validate research proposition 2 are:

- describe the difference between 'wet' and dynamic signatures;
- present and analyze a case study involving a rejection of electronic handwritten signature evidence;
- describe and identify the potential cause(s) of rejection of signature examiner testimony;
- propose an examination and probabilistic methodology for dynamic signature evaluation;
- validate on known-source data and provide justifiable estimates on reproducibility and reliability;
- illustrate the versatility and extendability of the methodology.

7.3 Proposition 3 - The Specificity of (Dynamic) Signature Examination

Signature examination requires a different probabilistic approach than other forensic fields, like fingerprint or DNA evidence. The identity verification scheme is different from identification tasks. Further, simulated signatures are a type of mimicry of a 'target' signature. Dealing with these forms of impersonation requires a probabilistic model tailored to the signature problem and its specific assumptions. A novel model for source-level signature evaluation is necessary to avoid misapprehension of the problem and erroneous interpretation of evidence.

Most forensic disciplines are concerned with assigning the probability of finding another entity possessing the same features of interest *by chance*. This is however not the case for signature examination. Signature examination is often confronted with the allegation of forgery, a very particular kind of claim involving an intentionally produced similarity between a target and the unknown source object. The features used in most forensic fields are often fixed and cannot be influenced willingly, for example glass refractive index, short tandem repeats (STR) in genetics, fingerprints, etc. Excluding extreme actions (such as surgery, accidents, ...) these features are mostly immutable over a long period of time. For signatures, as well as for many other behavioral biometrics, these ideal properties are not respected. Signatures, much like voices, are observable results of conscious human behavior and can therefore wilfully be affected. Hence, in these fields, we use the concept of sincerity[5] in our evaluation. Forgery is a type of insincerity, masking one's own features by recreating those of another person. In its essence, forensic signature examination treats a different kind of logical problem than most traditional forensic disciplines. Instead of assessing the occurrence of some set of features in a population, it's necessary to measure how well features can be impersonated or masked[6]. It is unclear whether this conceptual difference warrants a major change in the interpretation of the evidence and the statistical models applied to the problem. The conceptual basis of a probabilistic evaluation model for signature examination and the differences with other forensic evidence types needs to be further explored. This hypothesis stipulates that addressing the simulated signature issue requires a case-specific and sincerity-based model, specific to forgery situations. Therefore, a revision of existing forensic examination methodology is necessary.

In order to validate the research proposition 3, we need to:

- prove that the standard individualisation and signature problems are conceptually different;
- propose statistical models fitting the problem at hand;
- determine the effect and impact of using a theoretically adequate vs an inadequate model.

A thorough investigation would need to focus on the three aspects of signature evaluation. First, the investigation should clearly point out both conceptual differences. Second, it should translate the theoretical and statistical implications of the conceptual differences into an adequate model. Finally, the empirical impact of using inadequate models on forensic casework interpretation needs to be determined. If the empirical impact of the theoretical inadequacies is low, using traditional models may be a sufficient approximation. To address the previous three points with a conceptual description of the types of problems, a model of sincerity-based inference, as well as an empirical comparison of the performance of different models are proposed. These points are summarily:

[5]Sincerity signifies the absence of intentional changes in the biometric. Examples are voice impersonation, deepening, heightening your voice tone, forging a signature, inventing a signature or disguising one's signature.

[6]in case of signature disguise.

- theoretical description and conceptualization of standard scenarios (Signature and Handwriting examination)
- intuitive and logical comparison of rarity and sincerity (forgery) approaches
- comparison of statistical models fitting both approaches through parameters and assumptions
- study of the empirical impact through the variation of accuracy and inferential strength (BF) on known-source data

In order to clearly lay out and study the conceptual differences, three forensic fields were chosen for a brief comparison: forensic genetics, signature examination and handwriting examination. First, a conceptual description of both types of scenarios and then a focus on the differences between the two approaches is offered. The thesis then proposes a statistical model specific to signature examination and compares it with existing models. Finally, testing the different models on the same mock cases permits direct comparison of performance and inferential strength. The empirical impact will help conclude as to the importance of choosing a 'sincerity' model.

7.4 Proposition 4 - Contemporaneity and Time in Dynamic Signature Examination

Contemporaneity is an important factor in forensic handwriting examination. Dynamic features are different from their static counterparts in terms of variation and stability in time. There is a need for long-term studies of dynamic signature variation to define a period of contemporaneity, as well as for validation of a probabilistic model in regard to template age.

Time is a crucial element in security, identity and forensic sciences in particular. Forensic science has a strong connection to both future (prevision, policing and strategy) and past (events under evaluative and investigative modes), whereas biometric systems are centered on the present (access control, efficiency). Signature examination is a *re-examination* of a signature that has been accepted in the past. In casework, cases may date back years or decades and the presumed source may have aged, his physical or mental health declined or he may even have expired. Collection of control signatures from the given time period is often limited and the production of new material can prove problematic because of trustworthiness and age-related changes. Ageing and long-term evolution of signatures may cause significant changes, going as far as to make samples non-comparable. Unsurprisingly, ageing has caused scholars in forensic science to question the numerical identity[7] identity [318] of entities. Signature template ageing, evolution of writing behavior and reference material collection are all directly related to time and the ageing process. While there appears to be general agreement on sampling a sufficiently large timeframe between and after the case, as well as a preference for contemporaneous materials, little is known about the impact of time on signatures and on case assessment.

Many authors recognize the key place time holds in signature examination, albeit almost all recommendations and guidance on handling time and ageing in signature examinations are vague. Many authors recommend using contemporaneous control signatures, but without defining contemporaneity and the width of a contemporaneous time-frame. The real impact of time and contemporaneity on signature evidence and probabilistic methodologies is unknown. Precious little empirical research [157, 563] backs up the few claims made, such as a contemporaneity time-frame of a year [392]. Despite the supposed impact on casework, extensive studies on the subject have proven difficult in both conception and execution. In order to study specific variation, sampling needs to happen with some regularity on multiple occasions. Data collection in these studies suppose extensive sampling from individuals, which has only been insufficiently done in existing datasets (see chapter 5, tables 5.1-5.4). For a conclusive study of ageing, researchers need to consider the high natural variation inherent to signatures (within and between session, as pointed out by Evett & Totty [157], as well as Thiéry et al. [563]) and separate them from time-related effects. The aim is therefore to examine the effects of ageing and sampling on dynamic signature data as a first step. Following this assessment, a second step consists in determining the impact of temporal distance (contemporaneity) on evidence evaluation for dynamic signatures, using a probabilistic and data-driven approach. In summary, the aims of this proposition are both exploratory and evaluative:

- illustrate variation of dynamic features and signatures over time;
- document natural and temporal variation in dynamic signatures;
- study natural correlation and *mutual information*[8] in signature features;

[7]Numerical identity is meant as the concept of being a unique and selfsame entity. The concept is opposed by qualitative identity, which is established through common properties.

[8]Mutual information is an information-theory concept that expresses how much information is shared between two pieces of information.

- investigate representativeness of small reference material samples (for participants in the 'stable' age-range);
- define a time period for 'contemporaneity';
- establish good practices for sampling strategies in dynamic signature sample collection;
- study the impact of contemporaneity on a probabilistic evaluation methodology.

In order to explore these points, a two-part study is necessary. The first part of the study focuses on visualizing and analyzing signature variation in general, and highlight time-related changes in dynamic features. A variety of statistical techniques, visualizations and models are employed to determine within and between session variation for several selected dynamic features. The exploration of the features and the individuality of ageing effects are in themselves of interest to forensic scientists. Multivariate visualization techniques can show a drift of signature features due to ageing and the high correlation in its features. Finally, sample representativeness and the influence of time on sample is examined through a variety of sampling methods and non-parametric modelling. The second part of the study aims to determine the impact of time and ageing on a probabilistic case assessment. Different sampling strategies are used to investigate specific effects of sample size and contemporaneity on signature evidence assessment. The following steps are proposed to achieve the previously cited aims:

- visualize dynamic features and signature data for different sessions to explore feature ageing and trends, as well as within and between sessions variation;
- visualize dynamic signature data in multidimensional space to explore overall ageing and feature correlation;
- visualize dynamic signature data using non-parametric models for assumption and model checking;
- compare subsamples and sample collection strategies qualitatively through non-parametric models to empirically support representativeness of data and validity of conclusions;
- validate the use of a probabilistic model on 'aged' signature samples;
- test the impact of sampling strategies and temporal distance on the value of the evidence;
- summarize the results and produce recommendations for representative sample collection.

Results and Discussion 8

8.1 Proposition 1 - The Novelty of Dynamic Signature Examination

Dynamic signature examination is novel for forensic scientists in both technological and data terms. Dynamic signatures are complex constructs involving multiple research fields (e.g. computer science, engineering, biometrics, cryptography). The technology used to record them is capable of measuring dynamic and kinematic signature data, which was previously inaccessible to forensic examiners. These developments have transformed dynamic signatures into complex multivariate data constructs. The new developments and fundamental changes to dynamic signatures warrant a revision and extension of our knowledge on signature dynamics and data-driven evaluation of signature evidence. It is important to raise awareness, provide new insights and encourage research in the field.

Through a thorough review (see chapters 4 and 10) of the available scientific and technical literature on dynamic signatures, it was shown that dynamic signatures have changed forensic signature examination. Not only do they change the way signing works, they also produce different traces and data during their production. Consequently, they require a new skill set for their examination. Additionally, dynamic signatures utilize rapidly evolving methods, requiring frequent and regular revision of technological and methodological knowledge on their subject [121]. They are produced and recorded electronically, which enables the quantification of signature data. As a result, dynamic signatures deserve further attention as a novel field from the forensic community.

The review has revealed that automatic signature verification (ASV) and online signature verification (OSV) in particular have flourished [27, 121, 179, 273, 289, 328, 460] (Figures 8.1 and 8.2) in the last two decades, while forensic applications and studies are rare (Figure 8.3[1]). Many technical advancements in matching algorithms and methodology, signature complexity assessment and classifiers have been made. Some exploratory studies [114, 232, 245, 514, 549] are publicly available. Forensic scientists have begun exploring variability and the selectivity of features [114, 186, 197, 337, 389, 420], as well as the sufficiency[2] of dynamic signatures for expertise [251, 336, 392]. Neuro- and movement science has been concerned with development and degradation of signatures, as well as substance effects on the kinematics of handwriting and signatures. Finally, biometric studies focus on data treatment, matching algorithms, efficiency and error rates [121, 179, 273, 289]. It should however be noted that most of the studies on dynamic signatures do not actually use 'pure' dynamic signatures. Many of the studies on handwriting kinematics do not

[1]Please note the change of the scale of the vertical axis.

[2]Meant as sufficient in quality and resolution to allow for forensic examination.

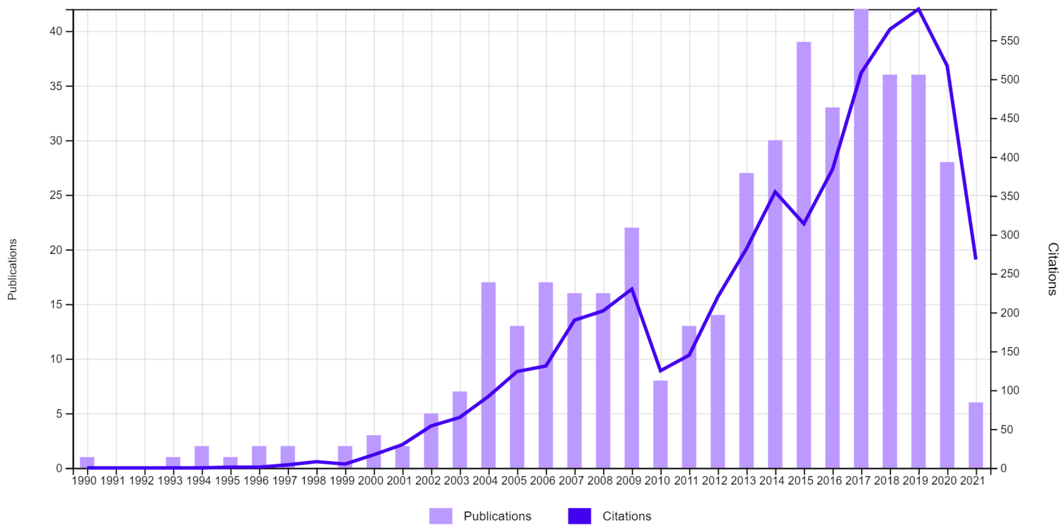


Figure 8.1: A barplot showing the increase in contributions through the number of papers published in scientific journals and citations in the field of dynamic signatures.

use the tablet surface and stylus as writing medium, but instead fix a sheet of paper over the tablet and sign using an inking pen. Therefore, they reflect 'wet' signature kinematics and ignore the influence of the new surface and writing implement. Tables 8.1 and 8.2 summarize our main findings and their significance for this proposition.

Dynamic signature examination needs to adapt methods from traditional signature examination for acquisition and treatment. Researchers should also look into new techniques for dealing with numeric and kinematic data. Methodological elements linking the evidence evaluation with the use of statistical models and data-driven approaches for signature examination are especially rare. The addition of said methodology can help address criticism to signature examination, as well as become a valuable tool for both researchers and practitioners. A dynamic signature examination procedure, integrating the various features with coherently used analysis techniques and evaluated using a probabilistic method is required. A complete methodology should integrate the principles of a rigorous examination process [244, 300, 312, 396], bias-avoiding procedures [131, 132, 140, 191, 269], as well as the empirical knowledge and pattern matching techniques used by forensic examiners [72, 244, 300, 392] and complement them with data analysis, statistical modelling and evidence interpretation. A chain is only as strong as its weakest link, therefore none of the different steps should be neglected. As yet, no research unifying all of the concepts and advancements into a common framework exists. This research tries to address and reunite most of these steps into a comprehensive study.

The complete review can be found in chapter 10. While awareness of dynamic signatures as a novel challenge is increasing, deep knowledge of the subject is still scarce. The FHE community actively researching dynamic signatures have however produced a number of important advances. Impressive achievements in technical aspects, such as feature extraction and matching techniques, as well as novel insights on signature dynamics and kinematics, have been highlighted in the review. Nonetheless, there is currently a lack of methodological and theoretical advances to accompany these technical improvements. Evidence assessment has not been actively developed and merits more attention from researchers. More specifically, few studies are concerned with the use and combination of correlated (dynamic) signature information and even fewer are compatible with probabilistic inference. Further research should equally focus on methodological advances and extend the use of probabilistic models to signature examination.

Finding	Conclusion
Dynamic signatures are economically (and ecologically) advantageous for businesses. They also benefit from high usability and user-acceptance.	Handwritten electronic signatures will become more widespread and relevant for forensic examiners.
Dynamic signatures are different from 'wet' signatures in both their nature and information content. They contain kinematic data, and are characterized by numerical data.	Forensic scientists need to adapt their methods to account for these differences and review their knowledge on biometric sensors, data collection and analysis and finally evaluation procedures. Methods and methodology should be adapted to reflect the new circumstances.
Novelty, interdisciplinarity and technical complexity pose new challenges, some outside of the FHEs' area of expertise. In particular concerns regarding originality, timestamps and metadata lie within IT forensic expertise.	Collaboration with IT Experts, Engineers and Biometricians will become necessary for the complete examination of dynamic signatures.
Dynamic signatures recorded on high quality digitizers have been found to be qualitatively adequate for forensic examination.	Quality of dynamic signatures has progressed rapidly in the last decades. The current quality of digitizers is sufficient for forensic and statistic analysis.
Movement kinematics and their usage for forensic casework [72] have been researched.	Forensic Examiners have a good grasp on execution and dynamic aspects of signatures. Previous research on movement kinematics can be used for dynamic signatures, as the mechanisms involved in the signature production are the same.
Movement and dynamics studies focus on pen-paper, rather than stylus-tablet interaction [59].	The impact of the different surface rugosity and writing instruments in signature examination is still an open research subject. Human-Device interfacing studies are necessary [60, 142, 143, 218, 596].
Basic dynamic features and their discriminative power for signature examination have been studied (speed, velocity, jerk, pressure, inclination, tilt, in-air movement).	Dynamic features contribute valuable information to the signature examination process. Some features (inclination, tilt, pressure) have produced ambiguous results and may strongly depend on hardware. Overall, the novel information and precise measurements require a change in examination procedures, but are reliable and discriminating.

Table 8.1: Summary of major insights gained from reviewing the relevant literature relating to dynamic signatures.

Finding	**Conclusion**
Powerful matching algorithms, stability region detection, as well as biometric verifiers have been researched. Data science provides helpful methods for the exploration and analysis of dynamic signatures. With these methods, novel, detailed characteristics, such as local features may be analyzed.	Computer science and biometric techniques could complement the FHE's repertoire of analysis techniques. They contribute to reproducibility and speed of the examination. A novel methodology should include elements from both forensic science and biometrics for validation and reporting.
Methodological advances and fundamental research concerning dynamic signatures are still necessary, especially in evaluative aspects. Awareness of the existence of dynamic signatures, as well as knowledge about the probabilistic model needs to be spread among the FHE community.	Methodological advances and exploratory studies concerning data treatment, as well as case interpretation are priority subjects for further research in signature examination. A methodological framework, which uses the full extent of available data to assess evidential value, is especially important.
To the best of our knowledge, there is currently no specific probabilistic model for evidence evaluation in dynamic signatures. Several statistic models exist [194, 201], but they are not specific to the dynamic signature problem.	Dynamic signatures present an opportunity for research and proposals into transparent, logical and reproducible frameworks for assessment in handwriting examination. A signature examination specific framework and statistical model would be valuable tools for forensic examiners and provide a strong basis for justifiable forensic conclusions. The development of such a model, with a probabilistic approach and computational techniques appears vital.

Table 8.2: Summary of major insights gained from reviewing the relevant literature relating to dynamic signatures, continued.

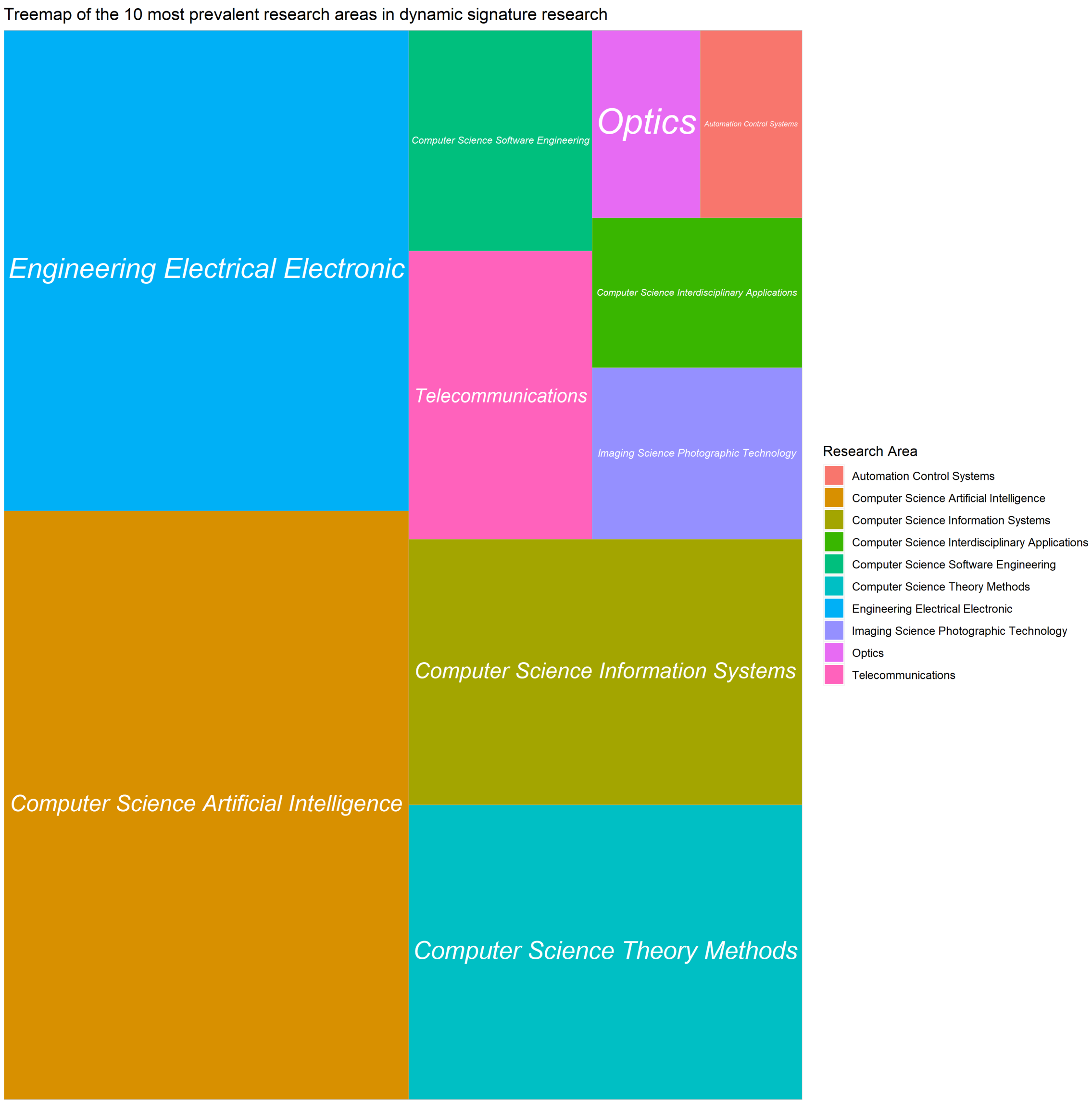


Figure 8.2: Treemap of the 10 research fields with the most content relating to dynamic signatures. The same query as in Fig. 8.1 was used. Research comes mainly from the computer science and electrical engineering communities. In total 441 contributions from 59 different fields were found. When adding the tag 'forensic' to the query, only 41 contributions were found.

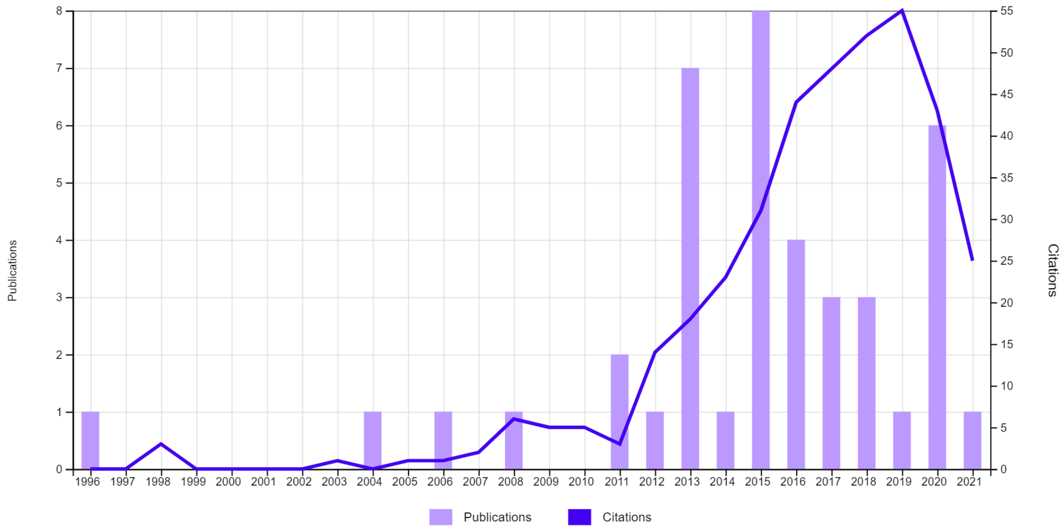


Figure 8.3: Barplot showing a review of all forensic contributions through the number of papers published in scientific journals and citations in the field of dynamic signatures.

8.2 Proposition 2 - The Methodological Changes Warranted by Dynamic Signature Examination

The analysis of the best evidence available is essential for reliability and admissibility of evidence. Dynamic signatures are files containing large amounts of complex multivariate data. Traditional pattern matching techniques largely ignore the dynamic and kinematic measurements made during the recording of the signature. Previous methodology does not permit the use of the dynamic signature to its full potential and therefore arguably does not constitute the 'best available evidence'. A probabilistic evaluation methodology for dynamic signature assessment provides a framework for the analysis of this additional information and may support examiners in quantifying evidential value, as well as in communicating reliable and justifiable conclusions. Its adaptation to dynamic signatures is an essential step to guarantee coherence and admissibility.

Recent advances in forensic science have concerned the study and the implementation of mechanisms of inference and the transmission of the value of scientific information (see chapter 1). A logical framework is essential for a valid assessment procedure and can help justify conclusions, as well as demonstrate their reliability and improve credibility of the domain accordingly. In addition to the fundamental need to utilize a logical and coherent framework to produce reliable science, a more immediate and practical concern exists for dynamic signature evidence. The particular issue is evidence admissibility and the concept of the 'best evidence available'. In most legal systems, judges and the jury are able to freely appreciate or even reject evidence. American courts have evidence admissibility hearings and several rules they apply to determine whether a piece of evidence is reliable and admissible. The legal courts need to evaluate scientific validity and credibility of scientific expertise in these cases. A lack of transparency, good methodology and procedures or even the incomplete use of evidence may cause the exclusion of said evidence. Dynamic signatures have been shown to differ from 'wet' signatures in many aspects. Even if traditional methods of signature analysis and comparison still apply to examination [392], not exploiting all available data may lead to exclusion of evidence from trials. All over the European continent, there is a strong current supporting a probabilistic approach for evidence evaluation.

A prominent example is provided in the AFLAC v Biles et al. [586] decision.[3] The case opposes the American Family Life Assurance Company of Columbia and the family of the late David Biles. Essentially, David Biles signed a life insurance contract with AFLAC on a Topaz signature pad (dynamic signature), through the insurance agent Brendan Hammond. Biles died a few months later, leading Biles' family to suspect foul play from Kenneth Ashley, the deceased's life partner, and Brendan Hammond, who had a prior narcotic-related conviction. Biles' family claimed that Ashley and Hammond conspired to fraudulently create the life insurance policy by forging Biles' signature. Ashley supposedly contributed to Biles' death, by either misdosing David Biles' medication on purpose or by omitting to help during an episode of Biles' condition. The family thus contested Ashley's claim to half of the policy and wanted to reverse AFLAC's payment, which should have been withheld because of the doubtful nature of the policy and death of David Biles.

The case revolved around a questioned dynamic signature, made on the Topaz signature pad. A graphical summary of the case can be found in Fig. 8.4 Both parties hired handwriting experts, Mr William Flynn for AFLAC and Mr Robert Foley for the Biles family. Both experts reached opposing conclusions, with Flynn's conclusions supporting the signature's authenticity, while Foley concluded it to be 'probably' a forgery[4]. Both parties proceeded to motion for exclusion of the opposing expert's testi-

[3] The complete court decision [586] can be found online, or briefly summarized by Harralson [242].

[4] Note that the position here presented is a fallacious 'transposed conditional' conclusion. Here the expert expresses an opinion on the hypothesis of interest for the court, the forgery, and not a value for the evidence

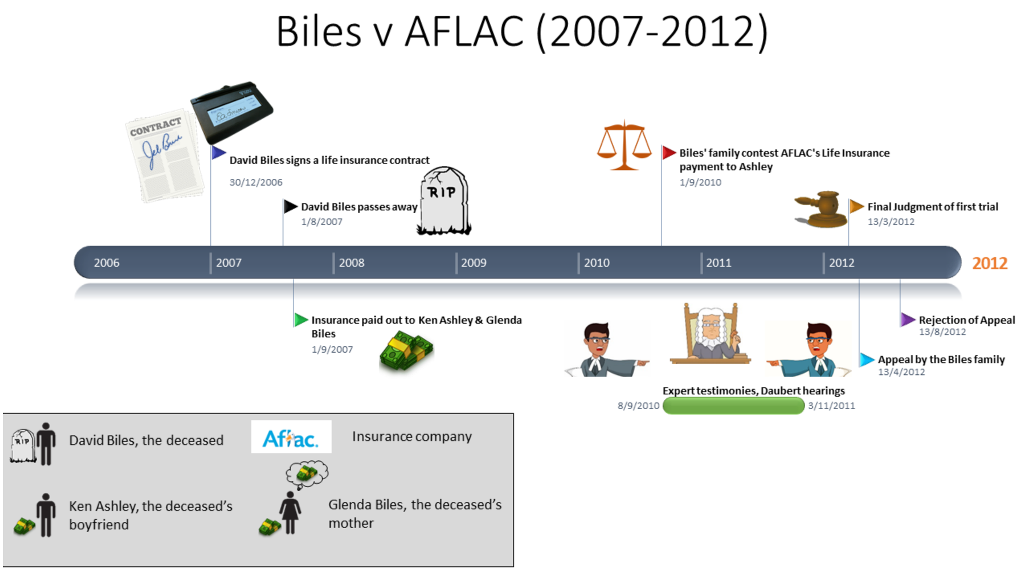


Figure 8.4: A timeline of the Biles vs AFLAC Case

mony. The court finally retained Flynn's testimony, because of methodological differences and technical knowledge. Foley was unfamiliar with electronic signature data and examined the signature as he would examine a copy of a handwritten signature. He relied on pattern matching and low resolution images on printouts to do so. Flynn had acquired knowledge about electronic handwritten signatures through his career and studies. He requested the digital signature data and worked from the raw pen data file. He most likely employed a 'wet' signature examination methodology, based mainly on pattern matching [392]. However he had knowledge on the kinematic parameters, as well as high quality visualizations of the signature. The court decided that Flynn, rather than Foley, had worked on the best evidence available and therefore retained Flynn's testimony. The following quotes summarize the court's decision.

> In addition, AFLAC moved to strike Mr. Foley's affidavit, contending his opinion therein was unreliable because it did not take into account the electronic nature of David Biles' signature on the application and arbitration acknowledgement form and was not based on an accurate representation of the actual signatures as made on the tablet but was instead based on low resolution images of the signatures which did not provide sufficient detail for an accurate analysis or the formation of a reliable opinion as to authenticity. [...] While Mr. Foley maintained that he provided the most reliable opinion he could, based on the evidence that was made available to him, he agreed with AFLAC's expert that the captured signature data would be the best evidence of the actual signatures on the documents in question, and that his opinion therefore was not based on the best available evidence. [...] The court thus concludes that Mr. Foley's affidavits, offered by defendants in response to AFLAC's summary judgment motion, are not reliable under applicable Daubert standards and that the motion to strike is therefore well taken.
>
> Unlike Mr. Foley, AFLAC's expert William Flynn based his analysis on a comparison of the signatures created based on the raw captured signature data and the known exemplars of Biles' signature. [...] It is manifest from Mr. Flynn's testimony that defendants' challenge to the reliability of Mr. Flynn's opinion is without reasonable basis; and therefore, the motion to strike/exclude will be denied.

(dynamic signature) under competing hypotheses (forgery, genuine).

The previous case exemplifies the consequences unawareness of static and dynamic signatures may have on forensic practitioners. Although the case primarily sets a precedent in American legislation, it sends a universal message to examiners. In Switzerland, judges can freely appreciate and evaluate evidence, instead of relying on evidence hearings. This principle, called the 'libre appréciation de la preuve'[5] permits the judges to consider or discard pieces of evidence according to his personal, justified judgment on their reliability, credibility and relevance. Failing to detect the electronic (or dynamic) nature of the signature and failure to use the entirety of the data may have dire consequences, even in continental European legal systems. Loss of credibility or unproven validity may also lead to the misapprehension of the inferential strength of signature evidence, culminating in the exclusion of the evidence. It is thus paramount that examiners are capable of detecting handwritten electronic signatures, know about challenges and hardware, as well as equip themselves with knowledge to methodologically examine and compare them.

On a more technical note, the AFLAC v Biles trial illustrates a difference in knowledge between examiners. Foley lacked the knowledge or tools to handle electronic signatures, while Flynn had previous experience dealing with them. As a result, their comparison material quality, as well as the available information extracted from the signature, were radically different. As a result of this gap in knowledge, Flynn had higher quality materials to work with than Foley. By extrapolation, the same reasoning could be applied for dynamic features and quantified data in signatures. An examiner focusing solely on shape information may produce less valid results than one using both shape, dynamic and measurement information. Hence, it appears logical to rely on the more eclectic method. Methodological differences and knowledge in data acquisition and data treatment may create the same, although less 'graphic' difference of 'evidence quality'. Dynamic signature data presents new challenges to examiners, which require careful revision of case circumstances, such as the type of hardware, the state of wear of the materials and the maintenance records. It is consequently essential to explore methodological ways to acquire, analyze, compare and evaluate the novel dynamic signature data.

This thesis proposes a methodological framework, exploiting the frequently available and quantifiable data in dynamic signatures. The methodology uses measurements[6] as global features. This data is complementary to pattern matching data and extends the forensic examiners panel of information at his disposal. A statistical framework, under the form of a parametric Bayesian model operating on empirical data was used (see chapter 11). Through a validation study, it was shown that this framework can be applied to casework and conforms to recommendations for evaluative reporting in forensic science [145]. Several qualities of the proposed model were highlighted and are summarized in tables 8.3 and 8.4. This aspect of the methodology should be integrated into a larger signature examination methodology, as it aims to provide complementary information (cf Figure 8.5). It intervenes during the analysis and comparison of the signatures and results can then be jointly interpreted with other relevant information. It is important to stress that the proposed methodology is only one of many tools at the disposal of the examiner. The success of the methodology depends on checking its applicability and its inherent assumptions, before its use. It also requires one to carefully weigh case conditions and limiting factors to choose the relevant data, to produce adequate and justifiable results. Further, the method in its current form makes use of global features, which are generally little used by FHEs and have great potential to complement their palette of techniques.

[5]'Sovereign consideration of evidence' - Freely translated by the author.

[6]In this case averages, maxima and variances of local features.

Finding	Conclusion
Dynamic signatures are immaterial, electronic evidence. They contain complex, multivariate data and resemble recordings.	Dynamic signature data far surpasses graphical information contained in signatures. Quantifying and measuring the additional data contained in dynamic signatures creates new opportunities for their description and deepening our understanding of signatures. They have their specific properties and challenges.
The specificities of dynamic signatures, as well as the added value due to signature kinematics, may result in challenges of admissibility and scientific examination procedures.	Dynamic signatures should be approached as a specific kind of electronic and biometric evidence. A revision and update to examination methodology is necessary to ensure admissibility and adequacy. Examination methodologies should use as much of the registered kinematic and graphical information as possible.
Probabilistic models respect the criteria of scientific acceptance, coherence and transparency. They can be tested for their robustness, reproducibility and accuracy.	Probabilistic methodologies can improve current examination practices, as well as reinforce confidence in signature examination by using empirical data and statistical models.
The Bayesian approach is compatible with both data-driven and the examiner's subjective approaches. Many examiners already work in this fashion, but lack a formalization and clear documentation on the subject.	Bayesian procedures, such as the one proposed in this research can be easily adopted by examiners and use both their experience and data. It is an extension and formalization of currently applied methodologies, which provides a strong and justifiable basis for conclusions while using available knowledge on signatures and forgery.
The method can be used with a variety of features and feature sets. It can be used on any signature type and style, as well as used in different case circumstances by choosing relevant data.	The method shows a high adaptability and flexibility. It is sufficiently generalizable to serve as a methodology for all types of cases, as long as its basic assumptions are fulfilled. Through its flexibility, even specific effects such as ageing or disease can be handled transparently. The only major downside for this adaptability is the necessity for case-specific datasets. Consequently, it applies to real casework conditions, even if it requires some additional effort in data collection.
The method is conceptually tailored to identity verification tasks. It calls for the use of case-specific data.	The methodology has a strong logical and conceptual basis, as well as a firm grasp on the signature problem. It respects the logical dependencies imposed by forgery and the logic in signature examination. The method can be further adjusted via the choice of propositions and the selection of alternative populations.

Table 8.3: Summary of results from the validation study supporting proposition 2

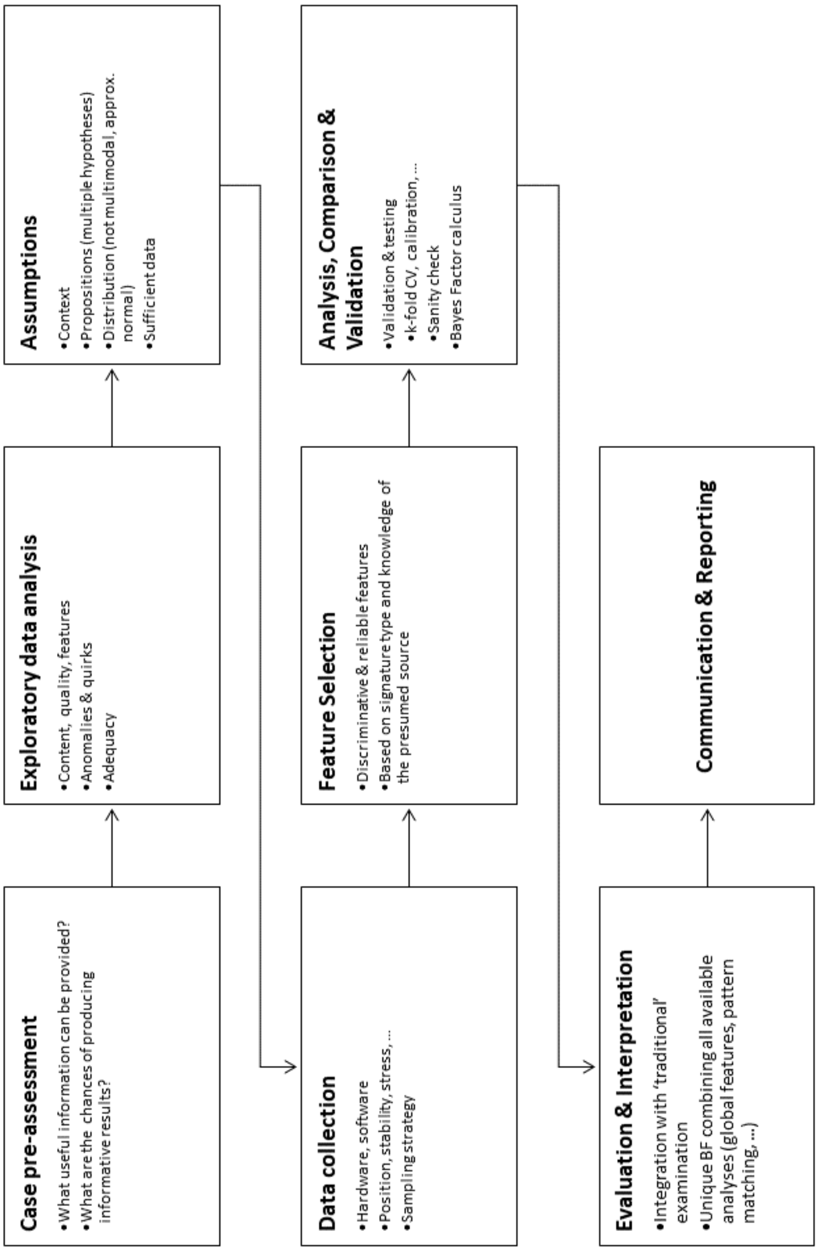


Figure 8.5: Summary of steps of the proposed model for integration into a forensic examination workflow

Finding	Conclusion
The method is fully explainable. The statistical assumptions are clearly stated. The data used in the model is empirical and described in detail. It relies on reproducible and explainable statistic procedures and provides a way to combine empirical data with subjective assessments.	The method is transparent on the data used, the model developed and its assumptions and limitations. Those aspects can be expressed and communicated clearly.
The value of evidence is expressed through Bayes' Factors. More than 95% of BF values support the correct hypothesis, even with little available reference materials (<10 signatures).	The probabilistic procedure is accurate in circumstances resembling current forensic casework. The Bayes Factor calculation is transparent and its result can be combined with the examiner's shape-based assessments.
The probabilistic method accounts for feature correlation and provides a way to combine features in a single evaluation step.	The proposed approach respects the conceptual differences between identity verification and identification tasks. It coherently combines signature features, which is a clear requirement in signature examination [244, 247, 300, 312, 396]. Through these considerations, it avoids over- or understating the evidential value.
The method produces little variance in accuracy when more than 10 signatures are used. Reproducibility, as well as performance of the system can be measured and transparently presented.	The statistical procedure is reproducible given a reasonably sized sample. It can be used to make recommendations for the minimal number of signatures necessary to make assessments.

Table 8.4: Summary of results from the validation study supporting proposition 2, continued

8.3 Proposition 3 - The Specificity of (Dynamic) Signature Examination

Signature examination requires a different probabilistic approach than other forensic fields, like fingerprint or DNA evidence. The identity verification scheme is different from identification tasks. Further, simulated signatures are a type of mimicry of a 'target' signature. Dealing with these forms of impersonation requires a probabilistic model tailored to the signature problem and its specific assumptions. A novel model for source-level signature evaluation is necessary to avoid misapprehension of the problem and erroneous interpretation of evidence.

First, a comparison of the case assessment model in three disciplines is proposed as a way to view the vast differences in every evaluative field. Here, three forensic fields, namely forensic genetics, handwriting examination and signature examination, are illustrated. Forensic genetics (non-mixture, simple profiles) has stood as the forensic gold standard [67, 410, 444] for both data use and statistical models. Signature examination and handwriting examination are not only behavioral characteristics, but their examination involves insincerity. These three disciplines are very different, with signature examination and genetics being almost polar opposites and handwriting examination reuniting elements of both problems. The conceptual description in table 8.5, aims to highlight the complexity for evidence evaluation in both handwriting and signature examination. The description concerns only standard, simplified assessment situations. In order to keep complexity to a minimum, various genetic concepts (such as chimerism, co-sanguinity and (sub-)population effects, or drop-in and drop-outs) have not been considered in the table. It is a representation of a simple case, analogous to many other identification-type disciplines in forensic science.

Element	Genetics* (Simple Profiles)	Handwriting	Signatures
Standard Propositions	H_1 : Mr. A is the source of the stain y H_2 : Someone else is the source of the stain y (not related to Mr. A)	H_1 : Mr. A is the source of the text y (genuine or disguised) H_2 : Someone else is the source of the text y (genuine or forged)	H_1 : Mr. A is the source of the signature y (genuine or disguised) H_2 : Someone else forged the signature y
Population Assumption	Stain y and control material x may or may not be from the same population	Text y and control material x may or may not be from the same population	Signature y and control material x may or may not be from the same population
Type of Assessment	Rarity of genetic features (STR, SNP) in a population	Rarity of handwriting features in a population & Complexity of reproducing or disguising features	Capacity for reproducing or disguising features
Within Source Variation Assumption	No variation	Writer-specific variation	Writer-specific variation
Sincerity Assumption	Assumes sincerity under H_1 and H_2	Considers both sincerity and insincerity in both cases	Assumes insincerity under H_2
Independence Assumption	Assumes independence of evidence and control materials (within the described parameters, except for related people)	Considers both dependence and independence of evidence and control materials	Assumes dependence of evidence and control materials

Table 8.5: Comparison of Forensic Genetics, Handwriting Examination and Signature Examination evaluation models. *The genetics aspect illustrated in the table is an oversimplification of most real cases and used as an example only. It excludes the possibility of having co-sanguinity or co-ancestry (population) effects, as well as specific cases which introduce within-person variation (chimerism, allele drop-ins, allele drop-outs). It mirrors the simplest possible case of evidence evaluation in forensic genetics. In case of signature examination, only the forgery scenario is considered.

This conceptual and theoretical difference is discussed and translated into statistical terms in chapters 6 and 11. The most important change in statistical terms is the independence assumption between the questioned and reference materials[7]. This assumption changes the formula and computation of the Bayes Factor for signature examination from the traditional 'rarity' models to an 'insincerity' model. In the comparative study in table 8.5, there is a focus on two aspects: the difference in statistical models and their propositions, as well as the importance of using adequate population data for inference. After the analysis of the conceptual differences, a more adequate statistical model was proposed (chapters 6 and 11). In chapter 11, a gradual development and description of the statistical signature model is proposed. First, a description of four multivariate statistic models proposed for the fields of glass examination, handwriting examination and signature examination is provided. Second, a comparative analysis of the performance of the four models applied to signature examination is proposed, in order to analyze the implications of applying either model to signature examination. To ensure comparability between the methods, all case assessment is carried out on the same[8] mock cases. The accuracy, variance, as well as the comparative analysis of inferential strength, are proposed to evaluate the adequacy of these models to the signature problem. An illustration of the achieved performance using the *"best"* feature set for every signature is illustrated in Figure 8.6. A discussion on the implications of the diverse assumptions and the theoretical adequacy is also proposed. A summary of the results can be found in table 8.6. Overall, it appears that theoretically non-adequate models do not truly express evidence that is applicable to the question at hand, here the verification of identity. Hence, these approaches often overestimate the value of the evidence in comparison to those produced by the specific signature model.

In forensic science, transparency and communication are vital. Good (forensic) science can be useless or even counterproductive when it is miscommunicated. In order to truly assimilate the results, having some knowledge of the underlying model and assumptions is important. Models are an imperfect repre-

[7]This assumption may also be broken in forensic genetics, because of different reasons. In particular population and subpopulation effects may create a correlation between the trace and reference materials. Nevertheless the relationship in signature evaluation is that of a direct copy instead of a distant co-ancestry.

[8]Meaning that the models were assessed on the same training and testing data sets.

Finding	Conclusion
The models for handwriting evidence and for signatures had the best overall accuracy. The current signature model shows comparable accuracy to the much more complex handwriting model.	Model specificity, complexity and adequacy, as well as within- and between-writer variance assumptions (WWV, BWV) have an effect on accuracy. Models for signatures should ideally incorporate the dependence assumption from the signature model, as well as relax assumptions concerning WWV and BWV.
The handwriting model is more complex than the signature model. It considers BWV and allows for variable WWV. The signature model currently only considers constant WWV. The higher model complexity resulted in higher accuracy when enough data is available. The accuracy of the handwriting model was lower when little training data was available.	Model complexity must be adapted to the amount of available control material. Additional complexity resulted in higher accuracy, especially with large amounts of materials. Nevertheless, increased model complexity caused a decrease in reproducibility when little data was available. A model with too many parameters to estimate and high complexity may not be reproducible and deployable in casework, where reference materials are often limited.
The signature model produced the lowest inferential strength of all the models. It also produced the weakest misleading evidence. The values obtained through the signature model correspond in scale to what is intuitively expected.	The signature model does not answer the same question as the other models and is not subject to the same assumptions. It uses two populations to calculate the BF. Distinguishing a specific forgery from a target genuine signature is a more difficult task than distinguishing two genuine signatures. Logically, the expected BF values for the forgery scenario should be lower than when differentiating two authentic signatures. This theoretical adequacy and conformity to expectations leads to justifiable and trustworthy BF values.
Variations in the inferential strength were observed for the three rarity-type models when exchanging the alternative population. Specifically the models produced lower BFs with the population of simulated signatures than with the genuine signature population as alternative. The proposed dynamic signature model uses both populations and is not subject to these effects.	Framing the problem correctly, and conceptually and theoretically examining the subjects is of utmost importance. Using adequate and case specific data is important for probabilistic evaluation of dynamic signature evidence. The variation and similarity between signatures in the simulated signature and genuine signature populations are different. Therefore, the databases used in the statistical model strongly affect the BF. Case-specificity is important for evidence evaluation in dynamic signatures.

Table 8.6: Summary of results from the model and methodology comparison supporting proposition 3. WWV = Within-Writer Variance; BWV = Between-Writer Variance.

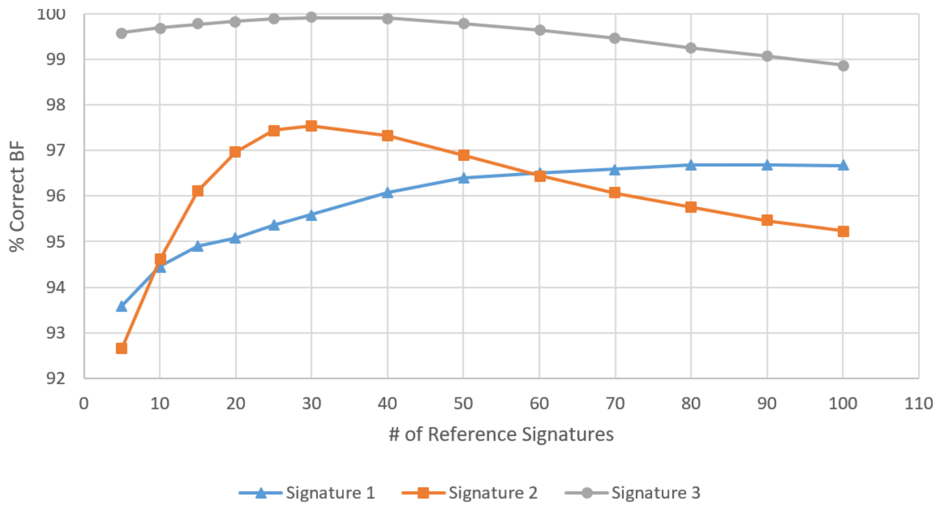


Figure 8.6: An illustration of the performance of the model for the three different signatures. Evaluation was carried out with feature sets specifically selected and different for each signature.

sentation of reality, some would even insist that every model is false by definition. Models, even simple ones, can however be useful to describe and analyse a given situation. As an example, take Newtonian and Einsteinian Physics. The theory of mechanics (Newton) has proved an oversimplification of the theory of relativity (Einstein), however it is still useful for most physical 'real world' interactions. A good model balances simplicity and accuracy, complexity is introduced when necessary. For example, the necessity for complicated models in signature examination stems from the common production processes (the signature movements and their constraints [72, 116, 598]), causing all dynamic features to be correlated. This is especially visible in the Principal Component Analysis results in chapter 13. The theoretical situation of insincerity, as well as the practical limitations caused by correlation of features, impose a particular structure of model with dependence assumptions, such as the one proposed in this thesis.

Concerning performance, the signature model, as compared to the simpler model proposed by Aitken and Lucy [6] for glass evidence shows significant increases in accuracy. Clearly, the signature-specific model and its assumption have a strong impact on BF accuracy. The model specific to the signature scenario performed almost as well as the more complex handwriting model. Both models have high BF accuracy and provide reliable information, but have their particular strengths and weaknesses. The signature model has fewer parameters and often performs well with little training data, while the handwriting model performs better with more material. The handwriting model overall has higher accuracy, possibly caused by its more complicated structure. The increased complexity in the model structure is due to the relaxed within-writer covariance assumptions. An amalgam of both models, or a decision rule when to use which model, may be the best way forward. While the signature model could be further improved by integrating the additional hierarchical layer for within- and between-writer variation from the handwriting model, the current model's relative simplicity has advantages. The proposed model is much faster than the handwriting model, as it does not require approximate Bayesian inference (MCMC and bridge sampling). It performed the best on small datasets, which are the most common for signature examination. It was also determined that the models do not answer the same question, as such the BF resulting from the signature model is more adequate for the problem at hand. Both types of models are important parts of forensic science and evaluative inference, but their application to a situation and their assumptions need to be verified before using them. In summary, it would appear that using a signature model changes the BF significantly, although data collection is more convoluted due to the specificity to a source.

8.4 Proposition 4 - Contemporaneity and Time in Dynamic Signature Examination

Contemporaneity is an important factor in forensic handwriting examination. Dynamic features are different from their static counterparts in terms of variation and stability in time. There is a need for long-term studies of dynamic signature variation to define a period of contemporaneity, as well as for validation of a probabilistic model in regard to template age.

In this thesis, dynamic signatures as a subject were explored and their methodological examination and probabilistic evaluation were described. An important aspect of signature examination is the impact of time, ageing and template age on the examination process. Signature examination (as well as most other evaluative forensic disciplines) are concerned with activities that happened in the past. In some cases, the facts are disputed much later and therefore there is an important duration between the collection of the questioned sample and the examination. Time intervenes in various ways in these cases. Signatures are behavioral biometrics, and behavior changes over time. Physically, people and their neural and motor system age and change, due to habits, accidents, medical conditions and age. Time is a difficult subject to study, as long-term studies and data acquisitions are restrictive, time-intensive and complex to plan.

Time and delays affect the quality of conclusions from the examination of signatures in particular. Often, signatures on important documents are contested years after the actual signature was made. Procuring adequate and contemporaneous materials can sometimes prove a difficult endeavor. Time is a limiting factor in examinations, and may especially lead to questioning the validity of results in changeable fields, such as signature examination. Practically, time can affect the signer through ageing, the movement model through evolution and the comparison through temporal distance (template ageing). In particular, the effects of signature evolution and template ageing may affect the forensic scientists in their strategy for sample collection and production, as well as in their conclusions. In order to quantify and describe the effects of ageing and time on dynamic signatures, a descriptive analysis of signatures over a duration of 18 months is proposed. The study presented in chapter 13 proposes a closer look at the variation of dynamic signatures within and between acquisition sessions, as well as its evolution through time. This exploration revealed a variety of points, summarized in table 8.7. In summary, the proposed probabilistic model does not seem strongly affected by small time frames, such as 18 months, as long as the source data is sampled through time and of sufficient size.

The purpose of this research is to propose and support the use of a probabilistic model as a coherent approach for the assignment and expression of inferential strength. Guest et al. [233] have also shown that some features may suffer from loss of discrimination, resulting in a loss of performance, due to temporal distance between samples. They proposed using time-invariant features to limit the impact of time on biometric systems. Long- and short-term effects of time do exist, as highlighted by the exploration study (see chapter 13). The extensive sampling in this thesis' data collection can therefore be used to investigate the effects of time on the BF stemming from a statistical model for signature specific evidence evaluation. The conditions and sampling strategies presented in the exploratory study (chapter 13) are reused in this study, in order to highlight the importance of choosing or extending samples, as well as determine the impact of contemporaneity. The second study is a validation study of signatures collected from an 18-month time period. The results of this study can be found in chapter 14 and are summarized in table 8.8.

Generally speaking, time and ageing have been shown to have an effect on dynamic signature data, both on the short- and long-term. Ageing and learning can cause signature features to gradually change. These effects are nevertheless very user- and age-group-related, as stipulated by the four-phase ageing model [244] and empirically corroborated [233]. The changes in signatures and the extracted features due

Finding	**Conclusion**
The comparison of signature data between acquisition sessions showed dissimilarities in averages and quantiles of dynamic feature quantification. There are differences in both the features values, and the extent of variation within sessions. These differences concern both the graphic (spacing, ...) and kinematic (time, speed, ...) features.	Individual acquisition sessions do not represent the 'true' variation of a signature over a time period. Produced reference materials for casework from a single data acquisition often misrepresent dynamic signature variation.
The observed between-session variation is higher than within-session variation. The results corroborate those of Evett & Totty [157] and Thiéry et al. [563].	Sampling strategies involving multiple sessions are indispensable for dynamic signature data collection and applications. Using produced reference data from several sessions is good practice to assess signature variation.
The signature data as a whole exhibited drifting and changing of the feature values throughout the sampling period. The evolution observed in the study was slow and progressive.	Signature features change over time. Both short- and long-term variation exist in dynamic signatures. In the absence of traumatic or fundamental changes, signature shapes and kinematics are stable in time over short periods. Longer periods (>5 years) may cause stronger changes [325].
Participants showed signs of learning and adaptation, especially during the initial phases of the data acquisition. Signature time and size decreased progressively during the data acquisition. Other features, such as writing speed, increased progressively.	Signatures made on paper and tablet surfaces present very different writing conditions. These observations also suggest that participants had to get accustomed to the new writing conditions. There was some (re-)learning or adaptation of their own signature movements on the digitizer. Therefore, the initial data acquisitions showed some non-representative signature behavior. These effects may prove problematic in an examination, especially with first time users of dynamic signatures.
Signature variation was different for each of the studied signatures. Our results corroborate that variation is personal, as previously noted by Guest et al. [233]. Signature variation occurs primarily over long time periods [323, 325].	Signature variation is person-specific. Every case requires a careful assessment on a case-by-case basis when ageing is involved. Such studies are only possible when observing a longer time period. No universal ageing model seems applicable to all writers. According to the published literature all writers age differently. Nevertheless, the period of contemporaneity appears to be wide for dynamic signatures, as changes occur over long durations.
Signature features are highly correlated, as observed in the principal component analysis. This was previously pointed out by Viviani & Terzuolo [598].	Feature combination assuming independence between the features used misrepresent evidential value. A statistical model which takes the dependences into account should be implemented to adequately measure the value of observations.
Feature distributions sampled randomly from the collected data showed high similarity to the 18-month distribution. Distributions obtained from individual acquisition sessions are often similar to feature distribution of the full 18-month period. Increasing the number of sampling points throughout time increases the representativeness of the time period. 8.7	Using low numbers of signatures (<10) or signatures from a single session may impact reliability of the BF and 'error estimates' such as the rate of misleading evidence or C_{llr}. 10-15 signatures scattered across a period are representative of the signature's variation over time. Otherwise, collecting data from at least 2-3 acquisition sessions is good practice for reliable data collection and guaranteeing robust statistical inference.

Table 8.7: Summary of the exploratory study of dynamic signatures.

Finding	**Conclusion**
The method has proved applicable, accurate and justifiable for all the experimental conditions.	A probabilistic framework and methodology is promising for dynamic signature examination, as it is capable of handling temporal variation and template ageing.
The accuracy and the inferential strength are only slightly affected by temporal distance (within 18 months).	Materials dating from within 18 months of the case can be considered contemporaneous for signers in a stable age group (18-65).
The BF value and reproducibility are more strongly affected by the amount of reference signatures than by contemporaneity.	Given a sufficient data collection with multiple sessions, the method produces robust and reproducible evaluations. Procuring enough reference materials is more important than their (time) provenance.
BFs produced while using a single session as reference materials were more variable than those obtained when increasing the number of sessions. The same observation is true for accuracy. Mixing signatures from two different sessions produces more reproducible and correct BFs.	Sampling from single sessions makes BFs non-reproducible and thus non-reliable (and sometimes incorrect). Collecting reference materials for a case calls for several data collection sessions. Two different sessions are sufficient for 'requested' materials (materials created post-case).

Table 8.8: Summary of the major results of our evaluative study on time and contemporaneity.

to ageing and natural variation are real and present. Inadequate sampling, especially when no or little material is available can lead to over- or underestimation of intra- and inter-person variation and cause imprecise conclusions.

Despite the reality of effects of time and ageing on the data itself, the proposed model's performance is little affected by time. Distances on a scale of 18 months have not strongly influenced the results obtained. Within this frame, it provides a robust and valid method to deal with signature evidence. Performance of the model was high, as stated for proposition 3 (see table 8.6) and varied little throughout the samples. The strongest differences appeared when comparing simulated signatures, which do not account for the signatures gradual evolution. Unfortunately, long-term effects of ageing can only be measured over longer periods, such as a minimum of 5 years [322–325]. As the changes and ageing effects appear strongly user-related [233], the results cannot be generalized. A longer, continuous and regular sample from a bigger population would be needed to further understand the complexity of time's influence on signatures. Further studies should focus on these aspects and use a sample containing a variety of signature styles and age categories.

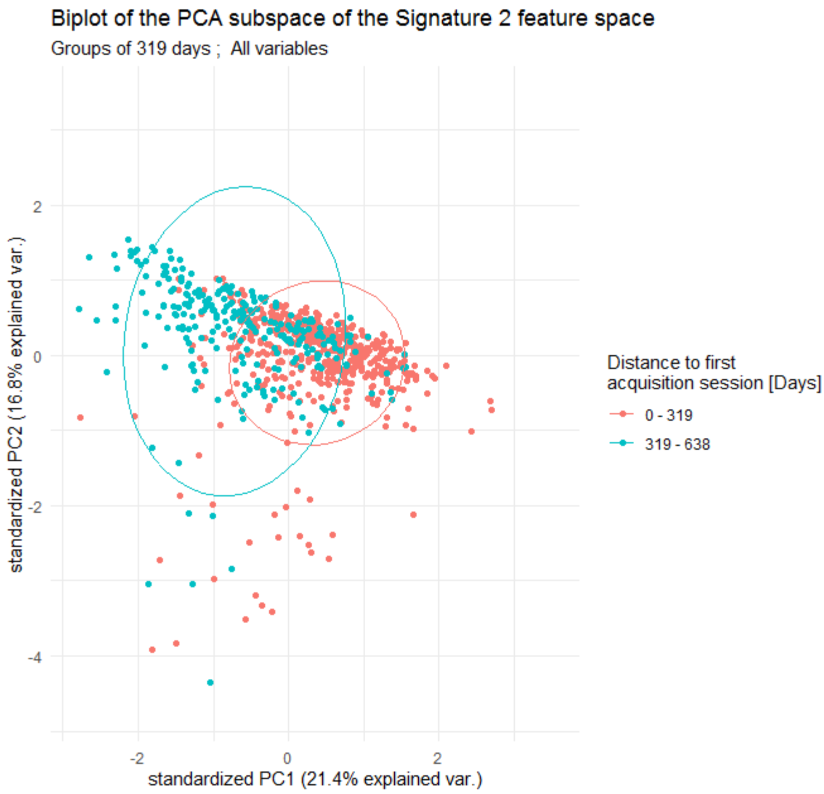


Figure 8.7: This PCA biplot shows the two principal components of the Signature 2 feature space. The PCA was conducted using all previously described features. The data was split into two groups of 319 days distance to visualize the temportal drift. The plot shows a gradual separation of signatures acquired in the first half, compared to the latter half of the acquisition period (21 months total).

9 Conclusions

9.1 Summary

This thesis' main purpose was to review and update the methodology of the signature examination field, in order to prepare for the increasing number of dynamic signature cases in the near future. Three essential aspects have been leitmotifs throughout the manuscript: knowledge sharing, methodology and transparency, as well as relevance and reality. Research that is not actively shared, or if it lacks relevance and practical applications, misses its point. Although the thesis is mainly concerned with methodological issues, practical applications in the field of signature examination show the feasibility and use of a multivariate approach in casework. These also highlight a comprehensive way to reconcile scientific accuracy, transparency and communication in forensic casework. An underlying concept of the thesis was to encourage synergies between signature examination and connected fields, such as biometrics, computer science and IT, as well as movement and neuroscience. This interdisciplinary approach strengthen the scientific basis of the field (see chapters 7 and 8) and introduces computational techniques and metrics necessary for measuring robustness and accuracy. Beyond these guiding principles, four main hypotheses have been addressed during the thesis. These deal with the nature of dynamic signature, methodology, data exploration and evidence evaluation.

First, the production, communication and distribution of knowledge to forensic scientists and lay people is crucial. Forensic scientists should research and publish actively in the field, raising awareness and integrating dynamic signatures into examination procedures and standards. Taking this approach is in my opinion a way to proactively prepare for the 'future' of signature examination. Many novel challenges related to the new medium for signatures, as well as signature modalities are bound to appear and can be anticipated. These include novel features characterizing signatures, variation of signature dynamics, hardware and other aspects. Through the review of currently available research (see chapters 4 and 10), the impressive advances in technical aspects, such as matching and probabilistic modelling have been highlighted. As a matter of fact, strong machine learning approaches for automatic signature comparison are on the rise, but they often lack specificity and transparency for forensic-use cases [445, 446]. Fewer publications have dealt with forensic methodology and statistical frameworks to evaluate the novel dynamic signature data. This document both provides an analysis on the aspects that remain to be studied and discusses, as well as proposing a probabilistic framework to improve on existing methodology.

Second, dynamic signature examination is currently lacking a probabilistic framework for evidence assessment. Most of the identified methodological shortcomings relate to the ways the evidence is assessed in forensic science. Forensic scientists world-wide have been witness to a gradual rejection of deterministic conclusions, as well as opaque procedures and reasoning. A methodology using a probabilistic model, tailored to the signature problem, improves upon transparency and rigor in this domain and helps forensic experts to establish their opinions based on data. Therefore, conclusions can be more easily justified. This step is in my opinion essential to assuage criticism to forensic signature examination. The Bayesian framework is a (possibly optimal) solution to formalize reasoning and improve transparency, while minimally affecting experts' examination process. It will help catch errors and produce justifiable conclusions. Checking the underlying assumption, clearly specifying limitations and the use of empirical data for the probability assessment (see chapters 1, 4, 11 and 12) are nevertheless necessary to further improve and support examiners. On a more practical note, the proposed framework permits the combination of both graphical and kinematic features of the dynamic signatures. Handling both data-driven and qualitative assessments inside the same framework is possible. In particular, data-driven approaches, with little used signature features can now complement examiners' expertise, while adapting to realistic, operational conditions (see chapters 12, 13 and 14).

Third, dynamic signature data has been shown to be a complex construct of multivariate data. The exploration of the complex structure of the data and correlations is a necessary step to fully understand and efficiently use it for forensic science purposes. Signatures present inherent natural variation as well as ageing effects, even in 'stable' age groups. Unlike 'traditional' forensic evidence, e.g. fingerprints and DNA, signatures are very variable and careful data acquisition and sampling are necessary for valid conclusions. Dynamic data appears to vary greatly, making its interpretation a difficult endeavor. Further research into the subject, as well as the comparison of local features is necessary to coherently use them. Nevertheless, it is certain that this added data is valuable and discriminative and bears great potential. Further, the availability of the measurements of these characteristics will allow for even more large-scale research, vastly improving the understanding and examination of handwritten signatures.

Finally, the evaluative results from the proposed method show that the statistical evaluation model attains high accuracy and good reproducibility. Probabilistic conclusions are expressed through a Bayes Factor, the measure of the value of the evidence. Through the presented results, it appears clear that the Bayesian approach for evidence evaluation is an adequate and useful tool for the forensic examiner. The proposal of a rigorous statistical framework, relying on empirical data, provides a strong foundation for examiners to rely upon when concluding and testifying. The Bayesian framework presents a coherent solution for evidence evaluation in signature examination [395, 591]. This project has demonstrated the use of a statistical approach for dynamic signature examination and evaluation. Hopefully, it has also provided a basis for collaboration and communication between disciplines such as biometrics, forensics and statistics.

This manuscript presents the first step towards the integration of a data-driven probabilistic methodology for evidence evaluation for dynamic signature examination. Through the different experiments, discussions and tables, the research hypotheses were corroborated, highlighting the feasibility and benefits of a probabilistic signature analysis. Based on these foundations, new research subjects will hopefully be built upon and solidify and extend our knowledge on signatures. Much room for improvement remains. Several elements in the field of dynamic signature examination surpassed the scope of this project. Local features are very promising, and are known to convey information complementary to that of global features. A complete revision of examination methodology, from the extraction, visualization and data treatment of dynamic data, to its evaluation and reporting would also greatly help practitioners. While the ENFSI recently provided a good primer [145], precise information and an easy to use library to calculate BFs could be essential for the field. Additional digital forensic aspects, such as metadata and timestamps associated with electronic signatures should be investigated. On an evaluative level, the probabilistic methodology should be extended to encompass disguise behavior in addition to forgery. Alas, due to the novelty of the field, many novel and unanticipated challenges may yet appear.

9.2 Research perspectives

Research in the field of dynamic signature examination is still in its beginnings. Many interesting research topics remain to be explored. This thesis acts as a primer and opened up an entire catalogue of different subjects for future research. A few of these are mentioned in the following paragraphs.

As a first possible research area, one may investigate the underpinning on how the Bayes' Factors are generated. Gaborini[1] indicates that the way parameters are compared[2] may need revising. This fundamental change may impact the robustness of the signature model and warrants a sensitivity analysis on the existing model. This alternative way of considering the 'test conditions' may affect the feasibility and evidence evaluation procedures in many fields. It is a major change and requires careful consideration.

A second area is focused on the data's distribution and the statistical model fit to the populations. For this thesis, a simple model was fit to the populations, having the advantage of speed and 'exactness' in its favor. This model may well underestimate the complexity of populations of writers with different within-writer variance, as was proposed in handwriting examination [56]. Different models, for example with hierarchical layers and modeling both within and between writer variation, and mirroring the actual complexity of the population should be tested. A more complex model may have advantages, such as more 'adequate' Bayes' factor values, but also present disadvantages such as need for more training data and increased computation time. Further research may determine which models are suited to which situations and applications and should be proposed to forensic examiners.

A third area is the revision of the hypotheses and the development of a multiple hypotheses framework for use in signature examination. Currently, few studies focus on disguise behavior as it is very difficult to obtain good quality samples. The statistical difficulties also involve calculus of Bayes' factors with multiple hypotheses and need of additional background data to inform the model. This topic is very relevant to forensic scientists whenever disguise cannot reasonably be excluded from the case context.

A fourth topic, along the same vein, is the integration of the probabilistic evaluation of evidence within a case assessment and interpretation workflow, developing data-driven pre-assessment among other subjects. Evidence evaluation, as developed in the current thesis, is only a small part of the case assessment and interpretation framework. The case pre-assessment for dynamic signature data may change in accordance to the digitizer specifics and may require a revision as to the previous procedures. Further, the decisions during an examination, as modelized by Found et al [187], require further attention and may also profit from statistical considerations. For example, the decision to collect more data, collect different data, determine error rates, perform a sensitivity analysis on the model, pursue or stop the examination, may all be studied in more detail.

On a more technical note, many other subjects are relevant to FHEs. The development of a solid data acquisition methodology, as well as a reliable data corpus are essential. Additionally, the influence of the writing position and circumstances (stress, sleep deprivation, posture, twist, duress, alcoholism, etc.) on the dynamic features of dynamic signatures are great unknowns, even though some information was provided in this thesis. Further studies should focus on their effect on local features. On top of this list, all the factors related to the hardware and proprioceptive feedback are equally important in this study. In the near future, interoperability and comparability of data from different devices may become an issue for forensic examiners. Many institutes and experts are equipping themselves with a multitude of devices already. A clear study of the influence of differing data acquisition conditions and data pretreatment would be of great benefit for the forensic examiner. Lastly, one cannot treat dynamic signatures without

[1] In his as of yet unpublished PhD thesis at University of Lausanne. Lorenzo Gaborini "Evaluating items of scientific findings in cases where the Defense says: 'It is my twin brother', SNF Number 170280

[2] Equal parameters vs 'free to vary', instead of expecting them to be unequal.

mentioning local features and score based likelihood ratios. Local features are a challenge to examiners in both examination and evaluation methodology. Their comparison to global features, and their general use is a great challenge that warrants much attention from the community. The comparison and further exploration of this subject, along with feature-based to score-based approaches, are interesting starting points for research in forensic science and collaborations with other fields. Finally, method validation and the measure and interpretation of calibration metrics, such as Empirical Cross-Entropy (ECE) and Calibrated log-loss (C_{llr}) needs to be further studied.

Part III

Original Research Articles

10

Dynamic signatures: A review of dynamic feature variation and forensic methodology

Published. Linden, J., Marquis, R., Bozza, S. and Taroni, F., 2018. Dynamic signatures: A review of dynamic feature variation and forensic methodology. Forensic science international, 291, pp.216-229.

Abstract

This article focuses on dynamic signatures and their features. It provides a detailed and critical review of dynamic feature variations and circumstantial parameters affecting dynamic signatures. The state of the art summarizes available knowledge, meant to assist the forensic practitioner in cases presenting extraordinary writing conditions. The studied parameters include hardware-related issues, aging and the influence of time, as well as physical and mental states of the writer. Some parameters, such as drug and alcohol abuse or medication, have very strong effects on handwriting and signature dynamics. Other conditions such as the writer's posture and fatigue have been found to affect feature variation less severely. The need for further research about the influence of these parameters, as well as handwriting dynamics in general is highlighted. These factors are relevant to the examiner in the assessment of the probative value of the reported features. Additionally, methodology for forensic examination of dynamic signatures is discussed. Available methodology and procedures are reviewed, while pointing out major technical and methodological advances in the field of forensic handwriting examination. The need for sharing the best practice manuals, standard operating procedures and methodologies to favor further progress is accentuated.

Keywords: Dynamic signature; forensic science; forensic handwriting examination; variation; evidence evaluation

10.1 Article motivation and structure

The field of forensic document examination has changed significantly over the last decades. The rapid development of computers, mobile devices such as Smartphones and tablet PCs, Smartpens, and other devices has given way to an explosive increase in connectivity and data generation. This has affected forensic document examination in many ways, including the domains of text processing and printing, imaging and image treatment, high fidelity reproduction and counterfeit detection. These developments have also led to novel skills and modi operandi for forensic examiners and criminals respectively. While it seems that the increase in use of computers should have resulted in the progressive abandonment of handwriting and signatures, the reality is quite different. The most common form of signature is still by far the handwritten signature. It is a behavioral biometric identifier linked to a physical entity, a given person signing, that serves as a sign of authenticity and intent. Handwriting and signatures are highly practiced, personal skills, which continually develop through years of practice. The dynamic signature, a digitized version of the analog handwritten signature, is becoming a common solution for businesses, concurring with paper-based signatures. With the rise of high-quality acquisition hardware such as connected pens and high-grade sensors, signatures can be written directly onto digital documents. This eliminates the need for printing and scanning paperwork, making the processes of signing faster and cheaper for businesses. Dynamic signatures are mostly embedded into documents by encryption. In this way, the signatures are strongly linked to both the signed document and to the signer, making them adequate replacements for paper signatures. The dynamic signature presents a challenge to the forensic document examiners because of the changes in data nature and volume. These changes mean that adaptations in evidence processing and evaluation are necessary as well. The field of handwriting and signature dynamics is of interest not only to forensic practitioners, but also to biometricians, medical practitioners and neuroscientists. Considering the recent developments in the field and the lack of a comprehensive summary for forensic science applications, it seems appropriate to critically review the current state of the art. This review should illustrate the purpose and methodology of examination, introduce common terminology and provide information about signature feature variation and influencing parameters. The present article defines the dynamic signature and its properties. In Section 10.2, questions of terminology are considered, in order to clear up difficulties originating from the ambiguity of diverse forms of electronic signatures. Section 10.3 elaborates on frequently available dynamic features and our knowledge about their variation in standard conditions. Subsection 10.3.1 deals with measurable dynamic features and their application to forensic purposes. The following subsections, 10.3.2, 10.3.3 and 10.3.4, review physical conditions, temporary states (e.g. intoxication) and hardware-related parameters respectively. In Section 10.4 a critical review of published methodology in forensic examination of dynamic signatures is given. Finally, a brief critique of current state of the art and some future perspectives are provided in section 10.5.

10.2 Defining the dynamic signature

The word signature, as defined by the Merriam-Webster dictionary, has as many as seven different meanings. Primarily, "signature" denominates the handwritten signature of a person, but it can also be seen as distinctive mark serving to set apart abstract entities, such as a corporation, a group or a project. While dynamic signatures conform to the first definition of a signature, many other types do not, such as cryptographic signatures ("electronic signature"), stamps or fingerprint-based signing. Curiously, the word signature designates both the act of signing, giving it a legal meaning, as well as the result of the signing process. Sometimes, the term signature is also used to refer to the writing process producing the trace. Many other meanings are associated with the concept [378].

Definition of signature:

- (a) the act of signing one's name to something
  (b) the name of a person written with his or her own hand

- [...]
- something (such as a tune, style, or logo) that serves to set apart or identify; also: a characteristic mark
- [...]

In the remainder of the present article, the word "signature" is used to describe the data resulting from the "recording" of the execution of the signing behavior. This data may take either the form of a graph (physical or digital ink), or numerical data, such as video recording, text data or images. Descriptors of the data will be referred to as either characteristics or features. Features related to the movements producing the signature will be referred to as dynamic features. Features related to the product of the executed movement, the graphical representation of the signature, will be referred to as static or graphical features. The signature is a "snapshot" of the individual's movement, given his state of health and mind, at a specific moment. Signatures are the result of a complex behavioral pattern, resulting from the activation of various regions in the human brain concerning functions such as linguistics, motor function and motor and visual feedback [424]. Not only is the human brain facing a demanding task when planning the movement, but it is also actively working during the execution of the signature. The signature movements need to be executed by the coordinated effort of different effector muscles situated in hand, wrist, fingers and shoulder [74, 326]. The multiple possible situations, physical states and deviations in movement control and execution create what is commonly referred to as "variation". The human body and brain are not as reproducible as machinery, which creates 'natural' variation within signatures from the same writer. Modified circumstances and state of the writer can create even stronger variation in the final product. Movement can be roughly separated into the planning and the execution stage. In both stages, there are many variables to be controlled and influence parameters to be accounted for. According to Huber and Headrick [268], this variation is due to parameters, called factors by the authors, which can be classified as being either intrinsic or extrinsic. Intrinsic factors are parameters over which the writer has some degree of conscious control. They are generally circumstantial in nature. Huber and Headrick further catalog several types of intrinsic factors, notably "imitation" (emulation of perceived writing characteristics), circumstantial factors (e.g. posture, writing substrate, writing instruments, situational constraints), temporal states (e.g. induced states from alcohol and substance abuse) and educational factors. Several categories, such as substance abuse might require further subcategories relating to the effects of the substance. Common substance classes influencing handwriting and signature include depressants, such as alcohol, and stimulants, such as caffeine and nicotine, or various other effects from medication. The effect (e.g. slower/higher writing speed, lower/higher pen pressure variation) and effect size depend on the dosage and the "efficiency" of the consumed substance and the metabolism absorbing it. Extrinsic factors are conditions out of the writer's conscious control, for example age, infirmity and injury, and handedness. Dynamic signatures are a type of handwritten signature, characterized by a chronological sampling of the signature movement. They differ from physical signatures in their acquisition method and the recorded features. "Physical" signatures are acquired on a substrate, mostly paper, which carries the ink trajectory, called the signature. Handwritten electronic signatures are recorded by digitizers, which may or may not need a "substrate." Many different kinds of digitizers exist, but signature pads are the most common digitizers. These pads sample the analog signal of the movement of the pen on (and sometimes even above) the pad spatio-temporally and create a dynamic signature. Due to the sampling, loss of both spatial and temporal resolution is inevitable. On the one hand, the resulting signature is less detailed and less continuous than a paper-based signature. On the other hand, the dynamic signature records previously "inaccessible" data, such as precise measurements of speed, acceleration, pen pressure, stroke direction and timing information. Dynamic signatures are known under various names in research fields. Researchers have used many designations, such as "Biometric Signature", "Electronically captured signatures," "Digital

Terminology	Synonym	Description
Physical signature	Signature 'pen and paper' signature	The ink trace resulting from the recording of the signature movements on a physical substrate, such as paper
Electronic signature	None	The 'electronic proof of a person's identity' [92] For further definitions see [95, 96, 435] General category, designates all kinds of 'electronic' data providing proof of authenticity.
Digital signature	Cryptographic signature	The 'electronic proof of a person's identity involving the use of encryption used to authenticate documents [91]
Handwritten electronic signature	None	Specific category Designates handwritten signatures containing only graphical data (static) or including temporal and movement data (dynamic)
Static signature	Offline signature	Digitalized version of a handwritten signature containing only graphical information (e.g. scanned signatures, image file of dynamic signatures)
Dynamic signature	Online signature	Digitalized version of the signature movement signal Chronological list of data points

Table 10.1: Summary of terminology

Dynamic Signatures," "Handwritten Electronic Signatures," "Online Signature" or "Dynamic Signature" [179, 242, 289, 290, 344, 355]. While the previous expressions all designate the same object, some names like "Electronic signatures" or "Digital signatures" are very similar, but refer to larger concepts or different kinds of signatures. Specifically, the term "digital signatures" refers almost exclusively to cryptographic signatures based mostly on Public Key Infrastructure (PKI), while the term "electronic signatures" is a legal term, often encompassing both cryptographic, handwritten and any other kind of "computer-based" signature. A visual representation of the taxonomy is proposed in Figure 10.1. Forensics has not found a consensus on terminology yet, but the biometrics field often refers to either "online signatures" or "dynamic signatures" [101, 127, 204, 233, 333, 366, 503]. A summary of common terminology with a brief description can be found in table 10.1. The term "dynamic signature" differentiates the signature from a scanned handwritten signature (or static signature) that does not contain any of the information related to the execution of the signature (e.g. timing, pressure, speed), while at the same time omitting reference to its digital nature. Any reference to "online," "electronic" or "digital" may cause confusion, as they may relate to cryptographic signatures. Further confusion arises because most digitizers use cryptography to guarantee safety of the biometric data and digitizer authenticity. Dynamic signatures also often use digital signatures in order to create a secure link between the signature and the digital document, as well as prevent tampering. The authors highly recommend using the "dynamic signature" designation, as it is logical, coherent and short. This proposed terminology corresponds to the one included in Harralson's work [242]. It minimizes the potential for misunderstandings as no computer-related words are included and it stresses the essential properties, the dynamics of the signature movement. Whatever choice of terminology will be adopted in the future, it is essential to define the expression and scope coherently in order to avoid misunderstandings and sharing misleading information.

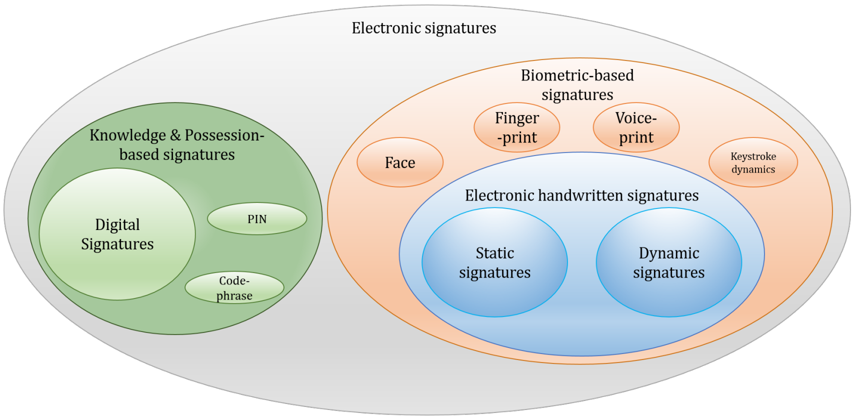


Figure 10.1: Taxonomy of electronic signatures

10.3 Digitizers, Dynamic Data and Variation

Dynamic signatures and the related dynamic data depend on multiple parameters, all influencing the signature and causing variation through different ways. Some of these parameters are linked to the hardware and measurement instruments, while others are more closely linked to the interaction between the writer and the writing implements; lastly, some are intimately linked to the writer. A review of a selected few parameters and their impact on the signature dynamics are presented in the next subsections.

10.3.1 Dynamic feature variation and discriminative power

Dynamic signatures record dynamic data such as timing and pressure information during the signing process. This information is not recorded quantitatively in physical or even static signatures and can only be inferred from the signature's graphical features, such as the line width or the groove depth created by the pen. Authors have considered several methods to provide better approximations for inference of dynamic features [151, 234, 429, 585, 612] on physical signatures, but there is much uncertainty involved in the determination. Having accurate measurements makes statistical treatment and data evaluation worthwhile and may lead to a more rigorous signature examination process. The field is very new to forensic science and has not yet received much interest from the forensic research community. Thus most of the available data is provided by research into biometric verification systems. Literature on dynamic signature variation is scarce, as most biometric literature focuses on classifier performance. Data on individual feature performance, as well as reasons for erroneous classification are often left for further research. Currently, the most commonly used dynamic features are pressure, velocity, acceleration, jerk, and pen angles (e.g. tilt, altitude, and azimuth), timing information and in-air (or pen-up) measures. Up until now these features have mostly been used through qualitative features, having been inferred from an ink trace. Forensic handwriting examiners (FHEs) have used concepts such as shading, tremor, fluidity, line quality, and tapering to describe these various dynamic qualities of signatures and handwriting. Table 10.2 describes the correspondence between measurable dynamic signature data with the qualitative features used by FHEs. Quantitative descriptors of dynamic features have been used in biometric verification systems, with little research from a forensic science point of view. Eoff and Hammond [147], for example, achieved an 83% rate of correct-writer classification in a population with 10 writers providing genuine samples using only speed, pressure and pen tilt. The discriminative powers of the diverse dynamic features have often been inferred through study of correct verification rates. Very few authors have studied separation of the variables, clusters or distributions in detail.

Dynamic feature	Related FHE feature
Timing	None
Positional derivatives (Speed, Acceleration, Jerk)	Tapering, flying starts & ends line quality, fluidity & tremor ink quantity, line width
Pressure	Shading, relative pressure ink quantity, line width
Pen angles (tilt, altitude, azimuth)	Shading
In-air features	None

Table 10.2: Linkage between quantitative dynamic features and qualitative features used by FHEs

The temporal derivatives of position, meaning velocity, acceleration and jerk, are mostly recognized as discriminative and useful features. They are often included in biometric verification systems [273], as well as used as a criterion in forensic examinations through fluidity, shading and line quality evaluations. Inversions in velocity (NIV) [218] are often used in movement and neuroscience to determine motor control, movement efficiency and automation. Teulings et al. [559] use the time-integrated squared jerk to characterize smoothness of movement. This feature has been used in many “therapeutic” or “developmental” handwriting studies to characterize movement quality. It is also included in commercial software such as the MovAlyzeR suite [418] and may prove useful for the forensic examiners for detection of tremor, a common sign of disfluidity [268]. Several studies indicate that velocity and acceleration, as well as their variations are interesting features in simulation detection [197, 337, 388, 390, 420]. Velocity and acceleration in simulations are often lower than in the genuine signatures. Jerk appears to be lower in text-based genuine signatures [388, 390] in comparison with the simulations, although the other signature styles do not show significant differences. Many studies have used velocity, acceleration and jerk as useful features for signature verification, as biometric literature reviews show [27, 273]. The writing pressure is a more controversial feature, as results differ between studies. This may in part be due to the different measurement schemes, as well as hard- or software-related problems [197]. Hook et al., Tytell, as well as Ostrum and Tanaka [260, 429, 585] all found in their respective studies that pen pressure is a stable feature in a genuine signature. Forensic practitioners have been citing pressure as a discriminating feature for simulations for nearly a century [97, 247, 427]. Unfortunately, very little empirical evidence for this claim has been produced. Kholmatov and Yanikoglu [303], however, appear to provide support against the usefulness of pen pressure. Other more recent studies do imply that the mean pressure is a discriminating feature [2, 388, 390, 408, 514] for forensic purposes. Some of their results [2, 388, 390] challenge a long-established theory, which predicts an increase in the pressure average in simulations. This theory involved the idea that there is increased stress on the person while creating a simulated signature [209, 214–216, 256], which affects movement execution. Pressure may also be of use as a local feature, using the continuous data provided by digitizers, rather than a summary in the form of the mean. Caligiuri and Mohammed [72] cite the absence of variation in the pressure signal as a feature of simulated signatures. Pressure variation and dynamics may carry just as much information as the mean value. Pressure is promising for use in forensic purposes, but has suffered from measurement reproducibility problems in the past [197]. It must also be noted that pen pressure may be measured axially or orthogonally to the surface, which complicates data comparison when two distinct digitizers were used. Many studies support pressure as a good feature for signature comparison. Even though recent results have been encouraging, the examiner should check carefully the measurement method and reproducibility for the digitizers involved. In-air pen movements constitute a category of features that is exclusively available in dynamic signatures. Terminology varies from "Pen-Up Movements," "Pen lifts" and "non-inking strokes" to "In-air Movements". These terms designate the strokes when the pen (or writing instrument, generally speaking) is not in contact with the digitizer tablet (or writing medium). Dewhurst et al. [114] studied these movements for forensic purposes and found that signature movements stay fluid and continuous even when the pen leaves the writing surface. In-air movements are often not straight linear movements from endpoint to starting point, but are curvilinear, continuous movements defined by said end and starting points of the successive on-surface strokes. Dewhurst et al. also found that in-air data (such as the trajectories) are as discriminating as on-surface data and might even be better suited for forensic purposes, as they cannot be seen on a signature image. Sesa-Nogueras et al. [514] found that both in-air and on-surface movements contain information relevant to handwriting and writer discrimination. The study also showed that information between in-air and on-surface strokes is not completely redundant. By combining information from both stroke types, better results could be obtained. Dròtar et al. [133] also showed that handwriting in-air movements contain different information than the on-surface strokes. In his study, he found that in-air movements can be used to effectively distinguish healthy control groups and Parkinson's Disease patients. In-air movements produced better accuracy than on-surface movements and the combination of both types of movements only marginally improved the in-air movement results. Other related data, such as the number of pen lifts, have often been used in "global" feature-based biometric systems and do not necessitate the recording of dynamic data while the pen hovers above the surface. The available studies suggest there being high information content and potential for in-air features for forensic and medical purposes. Pen tilts and angles are relatively rare features, as they are only recorded on a few digitizers or when using special accessories. The Apple iPad when used with an Apple Pencil or the Wacom Intuos Pro tablet, are examples of digitizers capable of recording these features. Franke [197] studied pen tilt (on a Wacom Intuos Pro) in a population of 30 writers and found out that a majority of people present a pen tilt between 50° and 60°. Many people differ in their pen-angle behavior, but a majority of people exhibit a range similar to these standard values. As for Pen-azimuth, Franke [197] noted that left and right-handed people have different azimuths. According to her study, pen azimuth values are more heterogeneous. Still, some values are more frequent, with a population mean value of 140° and a standard deviation of around 20° for most writers. Research done by Lei and Govindaraju and Fierrez-Aguilar et al. [180, 329] suggests that these features destabilize the verification system and lead to poorer discrimination. Some other authors observed that pen tilts and angles have improved the verification rate in their systems [626]. Zareen and Jabin [626] have observed that a false acceptance rate on a mobile device decreases steeply when integrating pen-tilt features. Franke describes pen tilt and pen azimuth as discriminatory features for writers signing their names [197]. Sesa-Nogueras et al. [514] found that pen tilt and azimuth have lower entropy (thus lower information content) in on-surface strokes, but contain more information for in-air movements. Pen angles have not often been used in a forensic context, so their efficiency remains to be determined. Most of the dynamic features need to be further explored in the forensic context. For instance, many questions regarding the features' long-term stability and short-term variations in time need to be researched. Additionally, some studies and models of the signature movements, such as the work done by Plamondon et al. [124, 125, 456, 459, 461, 462], imply that there are strong correlations between the signature trajectory and the associated dynamic features. These correlations between features mean that evaluation of univariate features may be inadequate for evidence evaluation in forensic science. In order to better approximate the strength of evidence, multivariate data evaluation may be necessary.

10.3.2 Physical conditions - age, health and posture

Movement does not depend solely on the planning and effectors of the movement, but it is also influenced by the physical state of the writer. Casework has led FHEs to research the effects of body posture, age and infirmity, as well as medication, etc. Most of these parameters are intrinsic (as described by Huber&Headrick [268]) and often circumstantial in nature. While the effects of such conditions on the static signature trace have been studied and documented, they may have far-reaching consequences for signature dynamics as well. Sciacca et al. [511] investigated the effects of posture on signature characteristics in order to gain insight into the comparability of handwritten documents and graffiti. Their results strongly suggest that within the signatures of one writer, variability is not different whether the person is in a sitting or a kneeling position [512, 513], as long as the writing surface is horizontal. More variation and changes have been observed when the writing surface is vertical [513]. Equey et al. [148] investigated the changes in width, height and aspect ratio when signing in multiple positions. They tested four different conditions: first sitting on a chair with the writing substrate on a horizontal table, second standing up with the writing substrate against a vertical wall, third standing up while holding the writing substrate on a hard board and fourth standing up with the writing substrate on a horizontal table. Their results showed that signature size strongly varies between the positions, leading to increased variation of aspect ratios. The greatest changes were observed when the person had to hold the writing substrate on a board while standing. The authors suggested that the instability of the writing substrate may actually cause more variation than the position itself. An older study performed by Evett and Totty [157] states that for handwriting, variation between sessions may have more effect than the studied effect. This complicates the interpretation of the position studies, which had not considered this. Thiéry et al. [562, 563] suggest that some of Equey's and Sciacca's results may have been misinterpreted, as their follow-up study showed no clear impact of position in a classification task. They further suggest that pressure strongly depends on posture and thus may be a good indicator for inferring the writer's posture. Overall, the authors conclude that a "sampling session effect," rather than an effect of position, is the cause of most of the variation observed in the previous studies. They are unable to give a generalized answer on the importance of the writer's position in writing variation, when signatures are not acquired during the same session, as the inter-day variation seems to have bigger effects than variation due to position. Finally, the influence of body position appears to be limited, but inclination and stability of the writing surface are high impact parameters. Forensic handwriting examiners are frequently working on wills or on dated signatures. In such cases, circumstances are often such that no new reference materials can be produced. The signer may have passed away, his handwriting may have degraded or changed in the mean time, obliging the forensic scientist to work with the available contemporaneous reference materials. Along with aging, comes an increased risk for illness or infirmity. Case-specific context information about the long-term evolution of the signature due to an illness or injury is necessary. If no such material is available, acceptance tolerances in the comparison process must be adapted to compensate for the lack of more adequate reference materials. In order to work in the presence of these difficult conditions, studies on illness, aging and degradation of motor function have been a priority for forensic examiners. It is known that aging is accompanied by a significant decline in cognitive functions in mammals [239]. "Executive function, which includes processes such as cognitive flexibility, cognitive tracking, set maintenance, divided attention, and working memory, is a cognitive domain impaired in aged humans and monkeys and is thought to be one of the first functions to decline with aging (Moore et al., 2006; Rapp and Amaral, 1989)." Naturally, this decline also affects handwriting and signature behavior, as was very well summarized by Caligiuri et al. [69, 72]. Normal aging effects include increased reaction time, decreased speed, increased movement time, increased variability and reduced grip strength. These signs are not exclusive to aging and may also be a result of diseases or medical conditions [72]. For many conditions in handwriting, such as aging, the effects are often strongly dependent on the individual, as shown by Galbally et al. [207] for aging dynamic signatures and other authors in biometrics [180, 289]. In forensic casework, access to adequate and contemporaneous reference material is highly recommended [187, 268, 396]. Galbally et al. [207] conducted a long-term variation study for 15 months, with six signing sessions and 46 signatures per user. His results showed that dynamic features vary more strongly than static features. In summary, the writer shows less variation in the spatial representation than with the execution dynamics, which may be subject to change [123, 207, 261, 606]. The authors of the study also noted that dynamic features are more strongly affected by “aging” and that these effects influence system performance in verification tasks. The downside of the work by Galbally et al. is the low number of samples per session as well as the absence of any training before signing, allowing the user to get accustomed to the signature pad. It is doubtful that four signatures are sufficient to represent a signature’s variation. Sciacca [511] for example recommends at least eight repetitions for words and letters when evaluating handwriting evidence. Mergl et al. [375] found that younger individuals write faster and more fluidly than older individuals. Guest [233] corroborates these results. He found that with age, signatures tend to be written more slowly, but noted no decrease in reproducibility with age. He did report that features related to execution time and pen dynamics were significantly different in his three different age groups. Age and contemporaneous material are important parameters in forensic handwriting examination, even more so when dealing with dynamic signatures. Many medical conditions (e.g. Parkinson’s Disease, Obsessive-Compulsive Disorder, Attention-Deficit/Hyperactivity Disorder, Alzheimer’s Disease, Huntington’s Disease and depression, [53, 70, 72, 164, 348, 454]) affect the handwritten signature. However, a detailed discussion of the various effects and causes surpasses the scope of this review. The forensic examiner should be aware of the effects of these conditions on dynamic information. Particular attention has been paid to neurodegenerative disorders [71, 72, 133, 134, 169, 213, 248, 249, 454, 606] and their symptoms, in particular Parkinson’s Disease, due to their frequent appearance in elderly people. Those conditions are often relevant for the forensic examiners when a will is being contested. Medical conditions such as Parkinson’s Disease are apt to change motor planning, inter-limb coordination and writing size, having large impact on dynamic features. Teulings et al. [559] for example cite movement control problems, slowness, reduced movement amplitudes and prolonged deceleration phases as classic signs for Parkinsonism.

10.3.3 Temporary states - sleep deprivation, alcohol and intoxication

Handwriting and signatures being a “snapshot” of the writer’s current state, many other parameters influence the “natural” execution of the signature movements. These may be linked to his physical state (e.g. injury, fatigue, sobriety, effect of medication/drugs/substances, recovery from illness) or to his emotional state (e.g. stress, anxiety, emotion and depression). Emotional states of people are notoriously difficult to study, making physical state research more popular. Research on such states is relevant to forensic casework if the parties claim special circumstances, such as intoxication, extreme fatigue or stress (e.g. due to coercion). Several studies have been conducted on the impact of sleep deprivation and fatigue on handwriting. Durmer and Dinges [137] describe the adverse effects of sleep deprivation on psychomotor performance and motor control. While these effects apply for long tasks, participants are generally able to gather their attention for short tasks (below 10 minutes) according to Bonnet and Rosa [52]. Huber and Headrick [268] summarize several studies and sources on sleep deprivation, notably Roulston’s and Remillard’s unpublished studies. Both conclude that lateral expansion was found to increase with fatigue. Remillard also noted bigger writing size and slower writing speed. Conduit [94] provides a summary of more recent existing literature on the subject. In his article, he critically analyzes several other studies, including that of Tucha et al. [583] on the same subject, mainly for methodological weaknesses due to the small sample size and the biasing effect the experiment order might have had. Tucha et al. [583] found that writers increased spacing slightly in handwriting when deprived of sleep. Bigger differences were observed in the handwriting kinematics. The study found that handwriting under sleep deprivation showed lower writing times, higher maximum velocity of ascending strokes and a decrease in numbers of velocity and acceleration inversions. Increases in maximum accelerations (positive and negative), as well as maximum velocity in descending strokes were weaker. Conduit [94] only investigated the spatial features of handwriting in his study, but found that these remained consistent, except for word and letter spacing. The author also found that the increase in spacing when subjects were sleep-deprived also extended to handwritten signatures. A study by Jasper et al. [292] investigated the effect of fatigue on handwriting and tested the effect when sampling within fixed time periods. The results of this research showed differences in handwriting features according to a cycle, the circadian rhythm. This rhythm is a kind of biological clock, working in a 24-hour cycle, with recurrent periods of fatigue or alertness, depending on the time of day. The authors show that handwriting fluency, quality and signature speed are not affected by sleep deprivation, while handwriting kinematics do vary according to the "fatigue" level of the participants. In another article by the same authors [291], invariance in signature execution is attributed to the lower complexity of the signature task, as compared with handwriting. Handwriting is a compound function involving not only motor control of the arm, wrist and fingers, but also the use of syntactic, semantic and lexical processing. Unfortunately, the study by Jasper et al. suffers from the small sample size, and the reliability of their results remains questionable. Most studies agree that short duration and highly automated tasks such as signature may be performed normally even when subjects are sleep-deprived [52, 291, 583]. Another subject of interest in forensic handwriting examination is substance consumption. Different kinds of substances may produce varying effects on the motor planning and/or execution. Some substances are psychoactive, while others affect the effectors (e.g. muscles). For the sake of illustration, psychoactive substances can be classified according to their effect on the nervous system, being either depressants (slowing the function of the nervous system), stimulants (accelerating the function of the nervous system) or hallucinogens (altering perception of reality, space and time). Many other substances may be contained in medication, narcotics, beverages and food. As substances have different effects on the system, effects on the handwriting and signature are expected to be different as well. Few authors have studied these effects on dynamic signatures, but several articles on the subject mention changes in signature dynamics subsequent to consumption. Caligiuri and Mohammed [72] provide a summary of medication and substance abuse effects on the handwriting and signature movements. Alcohol is a substance than can have temporary or permanent effects. Alcohol works as a depressant on the central nervous system and may cause euphoria and intoxication in individuals. Huber and Headrick [268] summarize some studies treating the influences of alcohol on handwriting, finding a total of seventeen effects on handwriting, all the while being critical about the applied methodologies and reliability of the studies. Alcohol influences longer writing tasks more strongly than shorter writing tasks such as signatures, much in the same way that fatigue does. Huber and Headrick cite the irregularity or increase of pen pressure due to high Blood Alcohol Concentration (BAC), as well as a decrease in writing speed and an increase in grammatical and orthographic errors, erratic movements and tremor. Phillips et al. [448] looked into the mechanisms producing the impairments caused by alcohol consumption by using handwriting recorded on a Wacom digitizing tablet. Their results indicate longer stroke lengths with stable stroke duration, a shift toward acceleration to deceleration imbalance, with alcohol intoxication being related to longer acceleration phases and pressure inconsistency for the non-alcohol-dependent group after consumption of an alcoholic beverage. The most notable changes are the prolonged acceleration phases and longer strokes, while stroke duration stays comparable. This indicates a change in writing behavior, specifically velocity and acceleration, while being under the influence of alcohol. The authors also state that mean pressure is not affected by alcohol consumption, although the pressure variations are affected. This was observed through the decrease of standard variation in the sample, implying a "flatter" pressure profile. Huber and Headrick [268] also state that effects on alcohol-dependent subjects may differ from regular people, as their features may become less variable when alcohol is consumed. Results by Phillips et al. [448] corroborate Huber's statement, although they are based on a very small writing sample, containing only four occurrences of the cursive letter "l". Asicioglu and Turan [24] investigated the handwriting of 73 people after consuming alcoholic beverages. Breath alcohol content was measured for every participant and effects of dosage and consumption were described. The study uses qualitative assessment of discrete features to compare handwriting in sober and non-sober conditions. The results corroborate the previously cited studies. The authors observed increases in "casualness" and "sloppiness," letter height, word length, grammatical and orthographic errors, spacing and number of tapered ends. The tapered ends are argued to be indicative of high-speed execution. The authors also argue that while execution of strokes is often faster while inebriated, pauses may be longer and thus may compensate the overall execution time in some cases. Interestingly, the authors have also observed the inverse effects, but less frequently. Overall, increases in writing times are more common (˜70-80% of the sample), decreases being more uncommon (˜20-30%) and no changes being the rarest phenomenon (˜1%). The authors also state that the level of breath alcohol may not be a good indicator for effect size, as some participants with low breath alcohol levels showed large impacts, while others with high breath alcohol levels showed little impact. Still, the authors found a correlation between breath alcohol level and height, angularity and tapered ends (and thus speed). Shin and Okuyama [515] used a dynamic signature verification system with several writing conditions to determine effects of alcohol on verification performance and to find features useful for detection of alcohol intoxication. They noted that the effects of alcohol fluctuate in time and were most pronounced 35 minutes after consumption. The effect progressively increases and decreases, as would be expected due to the progressive metabolizing of the alcoholic beverage. The authors propose four features to detect alcohol intoxication, namely average time needed to complete a signature, average pen pressure, pen velocity and stroke angles. Pen pressure and average time drop with alcohol consumption (by 30% and 8.8% of the original values), while velocity and internal angles increase (by 22% and by 10%). Velocity is the only feature to show a sharp drop from 110% to 80% of the original value between the signature session at 45 and 55 minutes after consumption. The study by Shin and Okuyama corroborates information pointed out by the other studies in respect to the increased variation of dynamic data after consumption of alcohol. Signature dynamics and verification rates are both affected by alcohol consumption, but effects depend highly on the metabolizing of the ingested alcohol. Other frequently consumed substances are caffeine and nicotine. Tucha et al. [584] investigated the effects of caffeine, a widely used stimulant, on motor performance. He administered controlled doses of caffeine (from placebo to 4.5 mg) with a caffeine-free coffee substitute to 20 right-handed adults. The experience investigated performances on a Wacom digitizing tablet using a short German sentence. The author looked at individual characters and sentences, while considering the metabolizing of caffeine. The study showed that only high doses of caffeine significantly affect writing behavior, with levels that could cause nausea in some of the participants. The parameters that were most affected were writing speed and acceleration on the individual elements, while the speed on the entire sentence remained relatively unchanged. Tucha et al. [582] conducted a similar study on the consumption of nicotine as per nicotine chewing gum, with doses of 2 and 4 mg respectively. His study, performed on 38 smokers and 38 non-smokers showed that nicotine consumption produces higher movement velocities and shorter writing times in both groups. According to the authors, the effects are in proportion with the dosage of nicotine, although they underline a lack of significance observed in their results. In both studies, no mention of pen pressure was made. Caligiuri and Mohammed [72] also review the very limited literature available on the effects of cannabis on handwriting. In their own study, conducted on a sample of five individuals, they highlight trends indicating movement fluency disruption and movement prolongation. Psycho-motor slowing appears consistent with other cannabis consumption studies. Average pen pressure was found to increase notably for four out of five individuals. The authors also note that sensitivity to the substance might be an issue, due to the fact that one of their subjects was very strongly affected, while the others were less so. We must point out that the study was of very limited scope and that results should be considered with great care.

10.3.4 Hardware-related conditions

Many types of digitizing devices exist, including gloves, mobile devices, camera-based devices and dedicated signature tablets [65, 166, 367, 449, 495, 497, 507, 604]. Every digitizing mode has its own particularities and a different set of recorded measurements. By far the most widespread digitizing devices are dedicated signature tablets and tablet PCs [356], which have been the focus of most forensic studies. These devices can (or must) usually be used with a pen, creating familiar conditions and permitting "natural" signature behavior up to a point. Writing conditions on paper and digitizer are quite different [14, 84, 218, 607], as the surface of the digitizers is smoother and there is less friction between pen and surface [218]. As a result, pen movement on this surface may be perceived as 'slippery' or 'too fast' and may require adaptation of the signature movements to the new substrate. Alamargot and Morin [14] tested this on young children in order to see differences in writing behavior to check the influence of visual feedback for writing and proprioceptive information in different learning stages. He generally found that the older children tried compensating for the more slippery surface by applying more pen pressure but still end up writing faster and less legibly. Gerth et al. [218] pursued this experiment with an adult population, comparing paper and ballpoint with plastic pen and tablet surface conditions. Results indicate that adult writers are also influenced by the tablet surface and tend to increase the pen pressure, letter size and writing speed to compensate for the different surface type. The authors also observed that experienced writers adapt to the new substrate in as few as 10 repetitions and were able to decrease writing pressure and other conditions to "normal" levels. The authors also noted that people adapt more quickly if the task they perform is an "automatized" movement process. These results affect recommendations for reference signature collection in dynamic signature cases. People who are not used to signing on tablet screens may need a longer time to adapt their writing and for their variation to stabilize. Thus, multiple or split sessions might be required to get used to the signature pad conditions and produce "natural" signatures. Another hardware-related issue is the writing space available for signatures and writing. There has been an ongoing discussion on how and to what degree signatures are affected by external constraints, such as predefined signature boxes, lines, display size, etc. These circumstances may force a writer to adapt his signature and thus affect its dynamics. Phillips et al. [447] found that the size of the handwriting can produce extensive changes in the dynamics. Downsizing handwriting requires the writer to adapt stroke number and size, as well as change acceleration and deceleration patterns. Upsized handwriting showed greater accelerations than medium-sized handwriting. The conclusions of Phillips et al. are in agreement with the hypothesis of Teulings et al. regarding spatial invariance [560, 561], rather than timing invariance. Signatories may also decrease overall speed and add strokes in order to fit their signature into the constraints, which also results in higher signature times. These constraints are traditionally boxes or lines, but may also be windows or prompts to sign, a specific interface or the digitizer screen itself. Fazio [170] investigated size constriction effects on a Wacom digitizing tablet with a compatible Wacom inking pen, while writing on a sheet of paper fixed to the tablet. Her study corroborates the previously cited results. Additionally, her study shows that the effect of constraint appears to be highly variable. Some individuals' signatures are not influenced at all, while other signers adapt strongly to fit into the constraints. This can sometimes go as far as changing features and parts of their signatures. Notable changes were found in velocity, jerk, overall length, ascender length and descender length. Velocity and jerk showed a decreasing trend with increased constraints. The only dynamic feature mentioned that did not vary with the size constraint is pen pressure. Impedovo et al. [274] also found the velocity is dependent on size constraints. It is worth noting that context may create circumstances that force people to adapt their signature and produce an obligatory change in signature. Fazio [170] explains the case of the Canadian passport, which contains a box-shaped size constraint. Any signatures touching the box's

Parameter	Main effect	Side effect	Summary
Age	Increase in time and decrease in writing speed	Decrease in fluidity	Aging effects strongly depend on the individual and affect dynamics differently Age group differences have been observed
Health	Depends on condition	Depends on condition	Diversity in diseases and effects is enormous Effects range from movement planning disruption to effector problems Specific literature on the encountered condition needs to be consulted
Posture	No notable effects	No notable effects	Existing studies have been unable to show significant changes due to posture
Fatigue	No notable effects	Increased spacing	Effects are only observed in tasks requiring concentration for long amounts of time Signatures are not affected
Alcohol	Increase in writing speed and acceleration increase in variation of dynamic features imbalances in dynamics	Diverse effects depending on dose and individual	Alcohol has strong effects on dynamics, but the effect strength depends on time of consumption and quantity ingested
Intoxication	Dependent on substance type, dose administered and metabolization	Dependent on substance type, dose administered and metabolization	Effects are varied due to the diversity in substances and effects on the brain and effectors
Writing surface angle	Various effects in extreme cases (e.g. vertical surface)	No notable effects	Smooth surfaces may cause higher writing speed and higher pressure Effects appear mainly while becoming accustomed to the tablet or in comparison with writing on paper
Writing surface size	Dependent on the individual, may reduce speed and shrink writing to fit	Pressure is slightly affected by constraint	Type of constraint and severity of not respecting constraints may be important Effect highly depends on the individual's choice and adaptation to the constraint
Writing instrument	Higher variation in graphical and dynamic features when using a writing implement the person is not used to (e.g. finger)	No notable effects	Signing with a finger rather than a pen with nibs introduces more variation Effect strength depends on interaction between writing surface and writing instruments

Table 10.3: Parameter effect on handwriting

border are invalid, forcing individuals to respect the size constraint and create a stronger effect than in ordinary circumstances. Diversity of writing instruments and signature pad surface characteristics are also of interest in the examination of dynamic signatures. As the instruments are not actually transferring ink to the tablet, interactions with the surface differ from regular writing instruments. Most dynamic signatures currently use either pen/stylus or the person's fingers as writing instruments. Signing with the finger conserves the motor programs used for signing with a pen, in accordance to the motor equivalence principle [268, 424, 615]. This means that the execution of the signature is still highly similar, no matter what instrument is being used. Nevertheless, signature reproducibility decreases strongly when a finger is used as writing instrument. Prattichizzo et al. [465] carried out a comparison of handwriting and drawn shapes using either fingers or pens and found that higher precision was achieved when using a pen. Tolosana et al. [572] recently tested finger, stylus and mixed signature verification. In his study, he found that error rates were lowest for stylus-made signatures, followed by finger-made signatures and mixed comparison. Stylus-based signatures performed significantly better concerning the equal error rates presented, which might indicate better reproducibility. Pens may show different behavior because of the material used, contact surface (pen-tip size), rigidity (deformation of the material), type of nib (or absence thereof) and their cross-sectional shape and diameter. Goonetilleke et al. [226] state that writing speed is not affected by pen shape, but accuracy is highest when using slim, circular pens. With the rise of Smartphones and mobile devices, signing with fingers is expected to become more common. The variety in writing instruments poses a challenge, as output from instruments such as the finger and pen do not provide the same precision and accuracy. Many different parameters have been shown to influence the dynamic features of signatures. Table 10.3 summarizes the major effects on handwriting and signature dynamics reviewed in the preceding paragraphs.

10.4 Methodology in Dynamic Signature Examination

Forensic science serves the justice system by providing assistance to the decision maker when specialized or scientific knowledge is required to evaluate evidence. The evaluation of scientific evidence often concerns the freedom and fortune of individuals. Forensic science needs to withstand cross-examinations and criticism from scientists, judges, juries and lawyers. For this reason, quality standards (and in some cases gatekeeping measures, such as admissibility hearings) are required to guarantee the necessary scientific rigor in forensic handwriting examination. Furthermore, forensic scientists cannot provide unprocessed analytical results. They need to identify the client's needs, respond to the relevant questions and report these answers in a comprehensible way. This means the standards of quality are not exclusively concerned with the results of an examination, but they do require that examiners provide detailed information about the applied methodology, standard operating procedures, validity of techniques used and the means of communicating the results. In order to be able to meet the standards of quality, accurate and well-tested methodologies are required. Many authors have published methodologies for physical signatures examination [187, 268, 312, 334, 382, 396] or have tried to describe and define the examination of handwriting [532, 542–545]. All of these publications have several things in common: A clear definition of the scope and goals at the beginning of the examination, rigorous analysis and comparison procedures, and a coherent evidence evaluation phase. A solid knowledge of the studied evidence, its features and their variation are necessary to frame and guide the process. Research on handwriting has greatly advanced since the early stages of handwriting examination [97, 247, 427, 482], in response to the severe criticisms laid upon the forensic handwriting examiner's activities [385, 481, 482, 484, 485] and forensic science in general [112, 386, 410, 444, 487]. Thus, the processes causing variation in signatures, as well as the extent of variation have been studied. FHEs have strengthened their knowledge of the origin of signature movements and the copying processes. Handwriting examination has delved into neuroscience, movement generation on a processing and planning level [72, 128, 214, 424, 461, 463, 492, 541, 615], exploring the dynamics of the movement process. It has also incorporated elements of biology and movement science to comprehend movement execution on a muscular (effector) level [128, 197, 255, 421, 465]. Additionally, forensic science has gained insight from the findings in biometrics, with several authors developing movement-modeling techniques [124, 125, 456, 459, 461, 462] to represent handwriting movement. There has been a research effort towards descriptor development and feature selection in dynamic signatures [344, 355, 438, 480, 573, 581] or adapting automated comparison systems to forensic purposes [413, 470, 471, 473]. Furthermore, the criticism has led forensic scientists to provide evidence for expert opinion reliability [43, 189, 190, 345, 517, 518], to reconsider the identification and unicity paradigms [87, 566] and to publish their methods [187] and best practices [144, 145] . Franke and Srihari have advocated for "computational forensics" [199, 524] as future development in forensic science, emphasizing the benefits of adding computational techniques to traditional forensic expertise. Their computer-assisted framework might help forensic science progress in a variety of ways, notably speeding up examination of large volumes of data, performing large scale testing and calculating performance and reliability, synthesizing new data sets, as well as standardizing work procedures. This framework is especially useful when dealing with quantitative data such as dynamic signatures. Many parallels can be drawn between the examination methodologies for physical signatures and dynamic ones, as both types of signatures are recordings of the same process. For example, examination order and structure, as well as qualitative examination procedures can be applied to physical and dynamic signature examination. These similarities might be one of the reasons why few researchers have been concerned with dealing with the more specific aspects of dynamic signatures. Nevertheless, several authors have noticed these methodological gaps and tried to fill them. Harralson [242] proposes a "methodology", which is actually a decision tree based on signature types (cryptographic, dynamic or static), capture reliability and "sufficient" captured features. The decision tree is certainly useful for the examiner, but it is too limited in scope to serve as an examination methodology and withstand an admissibility hearing. No other publications mentioning complete examination methodology for dynamic signatures can be found. Harralson does, however, mention important concepts, such as digitizer metadata [199, 242, 508, 594], sensor and capture reliability [196, 197] and "sufficient" features, which have been further addressed by other researchers. Other problems such as device interoperability [18, 455, 572, 573, 581, 593] and measurement compatibility [196, 197] have come to researchers' attention and have been studied. Some other publications are focusing on feature selection, reliability and data treatment for forensic examiners [185, 388, 389, 420]. Articles of more technical nature treat device interoperability and verification performances [572, 573, 581, 593] on different devices and conditions. Data quality and examination reproducibility are increasingly important to researchers and practitioners. There is also specific literature on examination and comparison methods in signature examination. Linden et al. propose an approach for defining a match based on whether or not its measured values are inside the variation observed from the known source signature sample [337]. Several authors have underlined how automation can help forensic examiners [49, 198, 199, 524, 526, 532]. Examination and analysis techniques for time-function features and parameter features have been used in biometrics and can be transposed to forensic science [27, 320, 347, 408, 438, 440, 468, 480, 523, 549]. Technical advances have been achieved in comparison techniques, using algorithms such as Longest Common Sub Sequence (LCSS) [440], Dynamic Time Warping (DTW) [273, 337], models such as Hidden Markov Models (HMM), Gaussian Mixture Models (GMM) or neural networks (NN) [273, 393], as support for the signature comparison process. New visualization tools and comparison procedures are being developed [251, 317, 337, 420]. Some commercial toolkits (e.g. Wacom Signature Scope, NeuroScript MovAlyzeR, signotec e-sig Analyze, Topaz SigAnalyze and SigCompare) are already available for forensic examiners, while others are under development by researchers [250, 503, 540]. These programs often offer extended visualization and analysis capabilities, helping the forensic examiner in his examination process. Unfortunately, dynamic signature data formats are often proprietary and no universally applicable open-source toolkit helping with data treatment, visualization and comparison currently exists. While much progress has been made in data treatment, visualization, analysis and comparison, few authors propose ways to evaluate quantitative signature data probabilistically. Forensic science is undergoing a paradigm change [78, 87, 154, 158, 566], moving away from the uniqueness and individualization paradigm, distancing itself from categorical statements. The more recent probabilistic approach maintains that uncertainty is present in the inference process and should be handled with statistical tools and models. Recently recommended statistical evidence evaluation approaches rely on personal probability [37, 341, 343, 557], the Bayes theorem [9, 12, 13, 158, 557, 558] and to some extent decision theory [35, 36, 553]. Recommendations from European Institutes and American case law suggest that the Bayesian approach is more coherent and well adapted to forensic science purposes than categorical conclusions or purely technical information [31, 154, 156, 158, 281, 567]. Marquis et al. describe a static signature case, approached using the likelihood ratio approach and personal probabilities [361]. Gonzalez-Rodriguez et al. developed a way to use Likelihood Ratios in biometric systems, using Kernel Density Functions (KDF) [224]. Kupferschmid [317] also applies kernel density functions (KDF to either features or scores (obtained by DTW), and uses the estimated densities for Likelihood Ratio computation. Chen et al. [85] used a score-based (DTW) likelihood ratio framework to work on grayscale, width and radian information extracted from static signatures. Their approach is similar to Kupferschmid's, but they additionally measure calibration and performance of the system. Gaborini et al. [201] apply and adapt the method of Marquis et al., based on Fourier Descriptors [56, 358, 360, 555, 556, 562], and propose a multivariate evidence evaluation approach. As can be seen through the literature review in this article, dynamic signature examination has gained the attention of the forensic examiner community, which is actively working to fill methodological gaps and develop robust and reliable methods. This implies that there are still important steps to be taken to attain the demands set by documents such as the NAS [410] and PCAST reports [444]. Other forensic fields have already started adapting to the requirements set to forensic science by the legal system [146, 284, 416]. Forensic handwriting examination should also adapt, especially when dynamic signatures and quantitative data are involved. Examiners should try to obtain information from digitizer manufacturers, to qualify measurement reliability and avoid detrimental pretreatment of data at the acquisition step. Standard testing procedures should be developed to test the measurement reproducibility of the digitizer and identify acquisition errors in the data. In this way, information about digitizer reliability data could be shared among laboratories. Research should also be directed at rational ways of selecting relevant and discriminating features in questioned signature cases. Using all available data contained in a dynamic signature is at best difficult and, at worst, detrimental to inference process. A robust way of selecting interesting and discriminating features for casework would greatly assist forensic examiners, who have to repeat the process for each new case. Multiple research teams should test comparison procedures for subjective visual inspection and automated comparison in order to qualify reproducibility. This would entail that a description of comparison methodologies always include mention of the data, procedure, algorithm and features that are being used. Casework in forensic science is defined by many different circumstances. These circumstances may influence the dynamic features of signatures more strongly than the spatial characteristics. Ways to deal with particular parameters need to be found. Especially, temporal distance between questioned signatures and reference data introduces important variation in signature dynamics. Few studies with a focus on temporal distance's impact on dynamic features of signatures are known. Most importantly, the scope of examination must be clearly defined. This calls for careful consideration of the court's needs. It would be desirable for examiners to clearly state their inference process and conclusion type, as well as to clarify the limitations associated with the chosen method [25]. Forensic examiners are invaluable in this process, as cases differ strongly and require adaptation to the case circumstances. Nevertheless, statistics and computer algorithms could be an invaluable asset to the Forensic Handwriting Examiner. They may not only strengthen reproducibility and transparency of the process, but also speed up examinations. Assumptions, data sets and procedures would have to be formulated clearly. Forensic signature examination has not, as of yet, adopted a probabilistic framework. The currently used method, mostly subjective assessment of graphical signature features, could be complemented with a statistical method dealing with the quantitative signature data. In this way, experts could make the most of the dynamic signature data, use the "best available evidence" and increase reproducibility and confidence in their results. The Bayesian framework for evidence evaluation would enable examiners to formalize a logically coherent reasoning process. Additionally, the Bayesian framework would give handwriting examiners a common vocabulary to express strength of evidence. This would facilitate exchange between researchers and colleagues, making collaborative testing and proficiency tests more easily comparable. Additionally, by conforming to the ENFSI and AFSP recommendations [25, 144, 145], the handwriting examiner guarantees adherence to quality standard and reduces exposure to criticism. It is the authors' firm belief that research into these key aspects would greatly benefit the dynamic signature examination by increasing scientific and statistical rigor. This will hopefully lead to acceptance and recognition of forensic handwriting examination as valid scientific discipline and satisfy critics by providing both a structured approach and empirical data to back up the field's claims.

10.5 Conclusion

A state of the art of the forensic examination of dynamic signatures has been provided. The review focused on literature discussing dynamic features and their variation under multiple parameters, such as writing position, writing substrate, age or intoxication. It also provides a review of methodology for forensic dynamic signature examination, while pointing out gaps existing in the current way of evaluating findings. The review has shown that case circumstances may have strong effects on both graphical and dynamic signature features. A lack of awareness about these parameters may lead to overestimation (or underestimation) of signature variation. In mild cases, this may lead to unjustified confidence in results, while in the worst case this may lead to erroneous conclusions. The state of the art has shown that valuable information on dynamic features in signatures can be found. While technical and signature variation related knowledge continually appears, few studies deal with forensic evidence evaluation and inference. Evett [158] reminds forensic practitioners that forensic science is about reasoning, not only about technical advancement and empirical data. The technical studies provide the construction materials for good forensic science, but without a solid foundation in the form of a framework for reasoning, the conclusion cannot remain standing. Researchers should continue exploring dynamic data and its variation. Nevertheless, forensic scientists should not forget that empirical data is given meaning through an evaluation process, which necessitates a framework, method and statistical model for reasoning. The next step for forensic examination of dynamic signatures will be the development of a probabilistic model that is able to use empirical data to inform the Bayes Factor. This model will support the forensic examiners in their casework through reliable and reproducible inference.

11 Bayesian multivariate models for case assessment in dynamic signature cases

Published. Linden, J., Taroni, F., Marquis, R. and Bozza, S., 2021. Bayesian multivariate models for case assessment in dynamic signature cases. Forensic Science International, 318, p.110611.

Abstract

Dynamic signatures are recordings of signatures made on digitizing devices such as tablet PCs. These handwritten signatures contain both dynamic and spatial information on every data point collected during the signature movement and can therefore be described in the form of multivariate data. The management of dynamic signatures represents a challenge for the forensic science community through its novelty and the volume of data available. Much as for static signatures, the authenticity of dynamic signatures may be doubted, which leads to a forensic examination of the unknown source signature. The Bayes' factor, as measure of evidential support, can be assigned with statistical models to discriminate between competing propositions. In this respect, the limitations of existing probabilistic solutions to deal with dynamic signature evidence is pointed out and explained in detail. In particular, the necessity to remove the independence assumption between questioned and reference material is emphasized.
Keywords: Dynamic signatures, Questioned documents, Handwritten signature evaluation, Bayesian multivariate models, Bayes' factor.

11.1 Introduction

Forensic science has been strongly criticized in recent years. The highly debated NAS and PCAST reports [410, 444] in the United States of America especially focus on the lack of statistical approaches in many pattern matching disciplines, as well as the lack of method validation. Many disciplines, handwriting and signature examination included, are indirectly concerned by these criticisms [385, 397, 401, 410, 444, 482, 485, 501, 502], while their progress towards validated and well documented approaches has been noted. Pattern matching fields such as fingerprint examination, shoeprint examination, speaker recognition and handwriting examination have a reputation of being highly subjective and prone to human error and cognitive bias. A general lack of statistical procedures in practice has been underlined [444]. In the last decade, this situation has led to a number of publications on standard operating procedures, reporting guidelines, but also methodological and fundamental research. Some researchers look for 'objectivity' by way of automated comparison systems. Automated systems often use a dissimilarity measure generated through a complex matching algorithm, which can be 'translated' into a statement about evidential value [106, 146, 194, 224, 254, 380, 384, 400, 402, 415, 417, 471, 473, 505, 506]. These systems offer more standardized and reproducible procedures [199], but are less versatile and 'adaptive' than the human examiner's methods. They require a rigorous framing of the forensic question to be answered, as well as an extensive collection of data. Currently, the court's question 'how much does a reference sample coming from a known source support that a given person is the source of a questioned signature?' is not answered in a coherent way. Evidence assessment methodology requires a clear, domain-specific definition of the propositions the Court is interested in, extensive theoretical knowledge for feature selection and model justification. Researchers have realized that in addition to data collection, there is an urgent need for a framework for evidence evaluation. Probabilistic procedures [13, 31, 37–39, 64, 80, 81, 154, 246, 379, 397, 399, 401, 551, 554] for evidence evaluation have been suggested as solutions. These data-driven methods provide the means for thorough validation, as well as implement a practical model to assess the value of observations made by the expert in the legal context.

Forensic handwriting examination automated systems, e.g., Flash-ID, WANDA-FISH, Graphlog, CEDAR-FOX, GRAWIS [105, 505, 528–530] were designed exclusively to deal with handwritten text. Their purpose is limited to text identification by reducing the pool of potential sources. The systems focus on rarity quantification of features in a given population. In the context of signatures, forensic literature on automated systems is sparse. Most of the available research is oriented to biometric tasks, where the objectives (identification or verification of the source), are incompatible with the role of the forensic scientist who should supply information as to the value of each piece of evidence under a set of mutually exclusive propositions. These are put forward by the parties at trial. A forensic scientist should inform a decision-maker (e.g. judge, jury) with evidential value statements enabling them to make their own decision, based on available contextual knowledge and on their own assignment of the undesirability of adverse outcomes (i.e. a false identification or a false exclusion). Recent recommendations made by the European Network of Forensic Science Institutes (ENFSI) and the Royal Statistical Society (RSS) promote forms of probabilistic reasoning to evaluate and communicate evidence [25, 144, 145, 283, 488]. These recommendations also noticed that key issues – that are 'those aspects of a case on which a Court, under the law of the case, seeks to reach a judgement' (ENFSI 2015 at p. 21) – provide the general framework within which requests to the forensic scientists and propositions are formally defined. Unfortunately, there are currently no methodologies dealing with dynamic signatures available to the forensic signature examiner that follow these recommendations. A specific evaluative methodology for dynamic signature examination is developed and illustrated in this paper.

Dynamic signatures (also called "online signatures") are a type of handwritten electronic signature [335]. They are essentially a "recording" of a signature, rather than an image. In this recording, both spatial and temporal characteristics of the signatures are acquired simultaneously and continuously throughout the signature. They provide novel, extensive multivariate quantitative data on signatures. Probabilistic approaches have been developed in various forensic fields like fingermark [416], glass evidence [6, 103], voice comparison [403], handwriting and static signature examination [56, 201]. The aim of this work is to highlight the limitations of current approaches to infer authorship in the presence of dynamic signatures and to propose a framework based on reliable assumptions.

Section 11.2.1 reviews the existing probabilistic models for forensic multivariate data, while a proposal for questioned dynamic signatures is illustrated in Section 11.2.2. The conditions and methods tested are presented in 11.3, while results and performance of the model are discussed in 11.4. Section 11.5, finally, concludes the paper.

11.2 The Bayesian approach – framing the question of unknown source signatures

Consider the following scenario: John Doe receives a bill from an Insurance Company 'Insurance Inc.' regarding a life insurance contract signed on January 1, 2019. John Doe is unwilling to pay the bill and denies having signed a contract with 'Insurance Inc.'. Given the amount of money at stake, 'Insurance Inc.' attacks John Doe in court. The questioned signature on the contract was signed on a digitizing tablet and is a dynamic signature. 'Insurance Inc.' claims that John Doe signed the said document, while John Doe claims someone else must have forged his signature. The court designates a forensic handwriting examiner in order to assess the value of the findings under the propositions put forward by the parties at trial. The examiner's responsibility is to provide assistance through the use of a coherent statistical framework.

In order to build up a statistical framework suitable for dynamic signature examination, a first look at existing Bayesian multivariate models in forensic science applications (including the field of handwriting examination) is proposed. We discuss assumptions and limitations of these models and move on to why these assumptions are no longer justifiable for questioned signature cases; these models are adapted to deal with situations involving dynamic signature features.

11.2.1 Multivariate statistic models for forensic evidence

Several models for the evaluation of the evidence in presence of multivariate data can be found in the literature [6, 7, 56, 194, 201, 224, 624], although few publications deal with handwriting data [56] or signature data [201] specifically. Questions in handwriting examination are most often related to writer identification, which is analogous to determining the source of a fingerprint or a glass fragment. The existing underlying statistical models rely on the assumption of i) independence between sources, ii) the sincerity of the signature and iii) all specimens are from the same relevant population. These assumptions are suitable for most physiological biometrics and for physical evidence. Such models inherently focus on the inherent variability of features. For example, in handwriting examination [56], comparison between questioned document and control documents originating from a known source rely on both the within-writer variability and on the between-writer variability.

The possibility of 'insincerity' of the available findings, i.e. the wilful alteration or imitation of characteristic information, is not taken into consideration. While this assumption may be sound for glass evidence or handwriting examination, it is difficult to justify in the context of questioned signatures. The possibility of forgery and disguise, forms of mimicry [202] to approach a 'target signature', breaks the independence assumption. While glass fragments, for example, can be described by a refractive index, which is an intrinsic property, signatures are not physical 'properties' of a person but a result of a complex behavioral process and a behavior can be changed consciously up to a certain degree (i.e. it is subject to insincerity). This fact will have an impact on the choice of the relevant population and the comparison materials, as forgeries show different variation than genuine signatures.

Three existing statistical models for the evaluation of evidence in the form of multivariate data are presented in the following subsections. These models focus on the notion of rarity and variability of characteristics in a population. The propositions of interest can be formalized as follows:

H_1 The questioned and reference materials originate from the same source

H_2 The questioned and reference materials originate from different sources

All models provide the examiner with Bayes' Factor (BF, for short), a measure that quantifies the degree to which observations support, in one way or another, competing propositions.

The first model taken into consideration is the one proposed by Aitken & Lucy [6, 7] in the context of glass evidence evaluation. Such probabilistic models allow one to deal with data showing two levels of variation: that within sources (e.g. within measurements on glass fragments originating from the same window) and that between sources (e.g. between measurements on glass fragments originating from different windows). This model can also be implemented for handwriting evidence, as this type of evidence also presents two levels of variation (i.e. there is variability in handwriting features within and between writers).

Consider a database $\{Z_{i,j}\}$, $i = 1, \ldots,$ m and $j = 1, \ldots,$ n, where there are p (>1) collected features (e.g. chemical composition of glass features, measurements describing handwritten character loops) from m sources with n repetitions for each source. The data is multivariate and assumed to be normally distributed, with a mean vector within sources θ_i and a within source covariance matrix W (eqn. 11.1). Note that the within-source mean vector θ_i (eqn. 11.2) varies between sources $i = 1, \ldots,$ m, while the within source covariance matrix W is assumed to be constant. This might be a reasonable assumption in some contexts like glass evidence evaluation, but less so in others such as signature examination. Following Aitken & Lucy [6, 7], a Normal prior probability distribution is taken to model uncertainty about the mean vector within sources θ_i. The two-level model (that will be called MVN, shorthand for Multi-Variate Normal) can be represented as follows:

$$Z_{i,j} \sim \mathcal{N}(\theta_i, W), \tag{11.1}$$

$$\theta_i \sim \mathcal{N}(\mu, B), \tag{11.2}$$

where μ is the between source mean vector and B the between source covariance matrix.

The evidence is defined by the occurrence of features from the questioned material and from the reference material originating from a given source. Measurements are denoted by y and x, respectively. The corresponding mean vector within sources are θ_y and θ_x, respectively. The data is assumed to be normally distributed:

$$(y|\theta_y, W) \sim \mathcal{N}(\theta_y, W), \tag{11.3}$$

$$(x|\theta_x, W) \sim \mathcal{N}(\theta_x, W). \tag{11.4}$$

The value of the evidence is computed as

$$\text{BF} = \frac{f_1(y, x|H_1)}{f_2(y, x|H_2)}, \tag{11.5}$$

where $f_1(y, x|H_1)$ is the marginal likelihood under hypothesis H_1 and $f_2(y, x|H_2)$ is the marginal likelihood under hypothesis H_2. If the questioned and reference materials originate from the same source (H_1 holds), then the mean feature vectors are equal, $\theta_y = \theta_x = \theta$. The marginal likelihood $f_1(y, x|H_1)$ from eqn. 11.5 can be obtained as shown in eqn. 11.6:

$$f_1(y, x|H_1) = \int_\theta f(y|\theta, W) \times f(x|\theta, W) \times f(\theta|\mu, B) d\theta \tag{11.6}$$

Vice versa, if the questioned and reference signature originate from different sources (H_2 holds), then the mean vectors within sources will be different, meaning $\theta_y \neq \theta_x$. The marginal likelihood $f_2(y, x|H_2)$ from eqn. 11.5 can be obtained as described in eqn. 11.7:

$$f_2(y, x|H_2) = \int_{\theta_y} f(y|\theta_y, W) \times f(\theta_y|\mu, B) d\theta_y \times \int_{\theta_x} f(x|\theta_x, W) \times f(\theta_x|\mu, B) d\theta_x \tag{11.7}$$

Note that in this latter case it is assumed that feature vectors originating from reference and questioned materials are independent. In the case of handwriting evidence, it means assuming the handwritten material has been produced without any intention of reproducing someone else's writing style.

There are instances where a Kernel density distribution may more appropriately model the between-source variability, for instance in presence of asymmetry or multimodality, where the Normal distribution does not fit well. In this latter case, the distribution for the between-group variability in 11.2 may be estimated starting from the available database $\{Z_i j\}$ as follows

$$f(\theta|z_1, \ldots, z_m, C, h) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{K}(\theta|z_i, C, h). \tag{11.8}$$

where $K()$ is the Kernel function, $z_i = \frac{1}{n} \sum_{j=1}^{n} (z_{ij})$ are the group means and h is the smoothing parameter. Aitken and Lucy [6] propose using a Normal (Gaussian) kernel centered at the group means z_i with covariance matrix $h^2 B$. This model is denoted MVK (shorthand for Multivariate Kernel). The marginal likelihoods $f(y, x|H_1)$ and $f(y, x|H_2)$ are obtained as in (11.6) and (11.7), respectively, where the Normal distribution for the between-source variability $f(\theta|\mu, B)$ in 11.2 is replaced by the kernel distribution $f(\theta|z_1, ..., z_m, B, h)$ in (11.8).

For both the MVN and MVK models, the marginal likelihoods can be determined analytically [6]. Parameters μ, B and W are being estimated using available background data (e.g., Aitken & Lucy [6] made reference to a database collecting measurements on the refractive index of glass fragments from m=62 windows).

Further research has been conducted on handwriting evaluation, a subject more closely related to the one of signature evidence tackled in the current study. Bozza et al. [56] observed that while a constant within source variability is reasonable for glass evidence, it is less sound for handwriting data. As such, handwritings differ not only in their feature mean values, but also in the inherent degree of variability. Forensic handwriting examination literature has long asserted that variability is personal and plays a large role in examination and evidence evaluation processes [268, 313, 396]. The extension proposed by Bozza et al. [56] allows non-constant within source variability by introducing an inverse Wishart distribution to model the within-source covariance matrices W_i. The extended model (Multi-Variate Normal Inverse Wishart, MVNIW in shorthand) is detailed below: $Z_{ij} \sim \mathcal{N}(\theta_i, W_i)$ $\theta_i \sim \mathcal{N}(\mu, B)$ $W_i \sim IW(U, v)$, where U is the scale matrix of the inverse Wishart distribution and v is the number of degrees of freedom of the inverse Wishart distribution, while the other hyperparameters have been introduced earlier.

Parameters μ, B are being estimated using available background data, while the scale matrix U and the number of degrees of freedom ν must be elicited. Bozza et al. [56] suggest to choose the scale matrix U so that the prior distribution is centered on the common within-source covariance matrix W, that is itself estimated from the available background data. On the other side, a large(small) number of degrees of freedom ν will allow to reduce(increase) the variability of the prior distribution.

The questioned and reference material are assumed to be normally distributed:

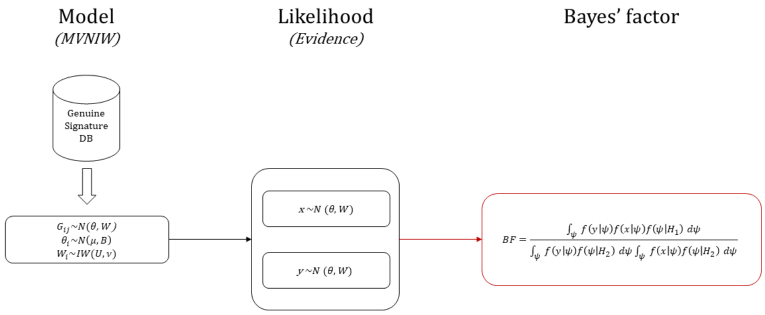


Figure 11.1: Graphical Representation of the MVNIW Model. A database (DB) of genuine material $\{Z_{ij}\}$ is available to estimate model parameters.

$$(y|\theta_y, W_y) \sim \mathcal{N}(\theta_y, W_y), \tag{11.9}$$

$$(x|\theta_x, W_x) \sim \mathcal{N}(\theta_x, W_x). \tag{11.10}$$

The value of the evidence is given by the ratio of two marginal likelihoods under the competing propositions H_1 and H_2.

If the feature vectors originate from the same source (i.e. H_1 holds), the within-source mean vectors and the within-source covariance matrices are equal, that is $\theta_y = \theta_x = \theta$ and $W_y = W_x = W$, and the marginal likelihood $f(y, x|H_1)$ can be obtained as follows:

$$f(y, x|H_1) = \int_\psi f(y|\psi) f(x|\psi) f(\psi|H_1) d\psi, \tag{11.11}$$

where $\psi = (\theta, W)$ and $f(\psi|H_1)$ is a compact form for the prior probability distribution of parameters under hypothesis H_1. If the feature vectors originate from different sources (i.e. H_2 holds), then $\theta_y \neq \theta_x$ and $W_y \neq W_x$, and the marginal likelihood $f(y, x|H_2)$ can be obtained as follows:

$$f(y, x|H_2) = \int_\psi f(y|\psi) f(x|\psi) f(\psi|H_2) d\psi. \tag{11.12}$$

where $\psi = (\theta, W)$, with $\theta = (\theta_y, \theta_x)$ and $W = (W_y, W_x)$, and $f(\psi|H_2)$ is a compact form for the prior probability distribution of model parameters under hypothesis H_2. Note that, as with the MVN and MVK models, questioned and reference material are assumed to be independent: the sincerity of the questioned material is undisputed. The increased model complexity allows one to better model the variation for handwriting behavior but also presents a further difficulty. The marginal likelihoods in (Eqns. 11.11) and (11.12) are no longer available in closed form since the integrals do not have an analytical solution. Bozza et al. [56] proposed to derive the marginal likelihood from the output of a Markov Chain Monte Carlo (MCMC) procedure, using Chib's method [86]. Other techniques such as bridge sampling [230] or importance sampling [222] can alternatively be used. A schematic representation of the MVNIW model can be found in Figure 11.1. Note that the available background data is given by sincere, genuine material and is denoted by $\{Z_i j\}$.

It must be underlined that there is another aspect that needs to be taken into account in signature examination, which refers to its 'sincerity'. A 'physiological' feature, like the glass refractive index can be assumed to be always sincere, as it can only be altered by unusual conditions. A 'behavioral' biometric like a signature can be affected by both conscious and unconscious factors and conditions. Glass cannot alter its properties willingly, while handwriting and signatures are a product of a conscious behavior and can be willingly altered to some degree. These alterations may be minor variations such as a small tremble or different sizing, but can also be major changes, such as shape and directional changes, or the use of different allographs. When we calculate the BF for signature evidence using one of the probabilistic solutions described above, we make the underlying assumption of sincerity of the signature, as we consider two signatures made by different people (i.e. when H_2 holds) to be independent. This point is mandatory because the marginal likelihood at the denominator (equation 11.7 for models MVN/MVK and (Eqn. 11.12) for model MVNIW) is obtained by multiplying the two marginal likelihoods of y and x that are assumed to be independent. The Bayes' factor that is given by

$$BF = \frac{f(y, x|H_1)}{f(y, x|H_2)} = \frac{f(y|x, H_1)}{f(y|x, H_2)} \frac{f(x|H_1)}{f(x|H_2)} \tag{11.13}$$

, simplifies to

$$BF = \frac{f(y|x, H_1)}{f(y|H_2)}, \tag{11.14}$$

because of the assumption of independence between sources when proposition H_2 holds, so that $f(y|x, H_2) = f(y|H_2)$, and moreover because features on the reference material are independent of the proposition, so that $f(x|H_1) = f(x|H_2)$. The approach to casework in handwriting and signature examination is fundamentally different because of the purpose of the handwritten element under scrutiny; whereas signatures are designed for identification, handwriting is designed for communication. As a result, the context for their examination is different, and justifies a difference in statistical model. Handwriting cases often involve an identification task, while signatures are more likely to be used in verification. The assumption of independence between sources can be justified for handwriting evidence evaluation under the condition that one writer does not have an interest in disguising his handwriting features or reproducing the features of another subject. As a consequence, it does make sense to assume that the questioned sample and reference sample belong to the same population and therefore share the same kind of intra- and inter-variability within sources. This amounts to saying that the same background population can be used to estimate model parameters under both competing propositions.

In most signature cases, however, the writer specifically wants to produce features similar to and inspired by a 'target'. It is similar to making a 'copy', which necessarily depends on the original signature. The difference stems here from the fact that the forensic examiner has to deal with a real signature, with underlying movements and variation, that is informed by a given source's signature. As such, it may be agreed that simulated signatures do not follow the same movement and writing mechanisms as genuine signatures, as the simulator may need to work outside his writing habits. This may produce a focus on copying the shape and eye-catching features of a signature, rather than the movement dynamics. In turn, this copying or drawing process would lead to either exceptionally narrow variation or a very wide range of variation (in case the simulator lacks the necessary skill to reproduce the features). The estimation of model parameters relies on the availability of a background database $\{Z_(i, j)\}$ that is considered to be representative of provided target features (i.e. signature characteristics). It appears incoherent to estimate model parameters from the same populations independently on which proposition holds. This necessarily leads to the collection of additional ad-hoc background data with signatures reflecting disguise and simulation behavior. Further details will be provided in Section 11.3.2.

11.2.2 A proposal for questioned signature casework

In John Doe's case, as in most cases involving signatures, the court's question pertains to the authenticity of the signature. Authenticity is linked to the notions of source and "originality". The underlying question is to provide information about the authenticity of a questioned signature. In many European countries, examiners are being encouraged to provide this assessment with a measure of the associated uncertainty [145]. The approaches used to provide this information are almost exclusively probabilistic. Forensic document examiners are thus asked to measure the extent of the support a series of observed features provides to competing propositions if the signature originates from a given source and it is not a copy. In the majority of signature cases, the examiner's task is similar to a "verification" task as defined in biometrics. In verification, a presumed identity is provided and the focus lies in classifying the questioned (unknown source) signature as either a genuine or an "impostor" signature. Although parallels to signature verification exist, forensic and biometric approaches differ. In contrast, forensic examiners aim to provide comprehensive and transparent information for the court, refraining from taking decisions. They aim to report probabilistic assessments of the evidence under the possible scenarios. Signature disguise has also become an important matter of fact. Only under particular circumstances or assumptions can this possibility be safely ignored. Unusual circumstances, such as body position, intoxication or illness and ageing are also limiting factors. Forensic science intervenes in cases long after the actual act of signing, so it has to deal with incomplete information and temporal delays between sample acquisitions more frequently. In this section, we propose a probabilistic approach to providing evidence assessments for signature examination.

In routine casework, it is generally alleged that a signature is either simulated or disguised rather than having a randomly made match to a genuine signature. Insincerity is the most common defence for the presumed source. Insincerity implies that there necessarily is a 'target' signature, which the insincere signature is similar to. Insincerity thus implies a conditioning on a set of signatures, usually a subset of the presumed sources' signatures, which serves as 'learning' material for the simulator. The relevant question corresponding to those allegations would be 'how likely it is that somebody could reproduce the presumed source's signature?' rather than 'how likely is it to find a genuine signature from a person different than the presumed source matching the questioned one?' This first consideration will have a major impact on the statistical model for the evaluation of signature authenticity. The previously described independence assumption, as well as the propositions specified are therefore inadequate and will be redefined.

To begin with, a clarification as to how handwriting and signature cases differ is needed. This clarification specifically concerns the problem of insincerity for signatures cases. Handwriting cases, unlike signature cases, fall into one of two categories, namely identification and verification. The first type generally involves a genuine (or disguised) handwriting, whose source the forensic scientist wishes to determine. The second type involves a presumption of identity, which the examiner wishes to verify and might involve either genuine or insincere (simulated) handwriting. Identification type tasks are very uncommon in signature cases, however most of the probabilistic models used in forensic science are tailored to these specific issues. This stems from the roots of the models, which often originated from identification areas, such as fingermarks, DNA or glass evidence. Within identification fields, very few cases feature dependencies between samples, and it is routinely assumed for sources to be distinct and unrelated. In verification-type tasks, a presumed identity exists, and the plausible alternative is that a third party tried to imitate this source's features. The impostor necessarily tries to mimic existing features and therefore depends on some source material. As a result using the identification models is a logically coherent and defensible approach for handwriting examination, but appears not to be for signature examination. Another important difference between handwriting and signatures needs to be noted. Handwritings need to stay within a codified form, so that a reader can decipher it, while this is not required for signatures (in most countries). Because of the lower level of constraints and conventions in signatures, the inherent variation between handwriting from different individuals is much lower than in signatures.

Signature cases almost implicitly exclude random matches to other genuine signatures by providing a 'presumed identity' for the signer. Except for people with identical names, insincerity would be the only reasonable explanation if the signature was contested. Signatures are distinctive signs of identity and reflect the signer's writing habits and movement patterns. They are also said to be highly automated movements ("overlearned"). As a result, they can be executed with little concentration, just like a simple movement such as throwing a ball. As a consequence, their variation should be less affected by concentration-related factors, such as alcohol intoxication or fatigue. Signatures can have different styles, shapes, directions, inclinations, flourishes and ornaments. Signatures should logically present greater variation between sources, as they are more diverse than handwritings, and lower variation within source, as they are highly automated. Simulated signatures should be more similar to the presumed source and present a smaller between-writer variation. We would argue that for signatures, propositions allowing only the genuine and sincere signature alternative arguably do not reply to the question of interest. This second consideration naturally leads to our proposal to reformulate the existing proposition and to adapt the Bayes' factor calculation accordingly.

11.2.2.1 Propositions for questioned signatures

As recommended by the ENFSI guidelines on evaluative reporting, the starting point of an evaluative process is the definition of (at least) one pair of propositions [145, 283, 488].

Before delving into the definition of the propositions, we would like to clarify the scope of application of the current proposal. Forensic Handwriting Examiners (FHE) and biometricians have classified simulated signatures into several categories. Researchers have adopted the terminology of 'Presentation Attack' [277, 503, 577] for instances where a simulator physically reproduces a signature. For the interested reader, a summary of these presentation attack types can be found in [351]. Naturally, other attacks which do not require a 'physical presence' exist. Such attacks involve creating a synthetic signature, or altering and reinjecting existing signature data [172, 203–205, 208]. These attacks surpass the scope of this article, but may pose a great challenge to the FHE community. This article will only address the more traditional 'Presentation Attacks'. To address signatures, the FHE community has adopted a specific vocabulary, which is chosen carefully as to not imply any legal meaning. For example, forgery is a legal term and is unfit for use by forensic examiners. For the remainder of this document, only the term 'simulated signature' will be used to designate a forged signature. The article will discuss the theoretical possibility of modelling disguise behavior into the evaluative model. A signature is only valid when a physical person can be linked to the document and signature. As such, there must legally be a presumed source. The questions of interest for a court of justice are those related to the source of the signature. This implies in particular that the relevant alternative population is made of specific forgeries of the presumed source's signatures. Further, forgeries are believed to have different movement mechanisms with respect to genuine signatures and thus exhibit different variation than genuine signatures. This makes the evaluation task more demanding, because the relevant population to be taken into consideration under the competing propositions will differ.

The specification of the first proposition in this case is straightforward. It appears natural to propose that the questioned signature was written by the presumed source. The presumption of identity is created through personal information, such as a name and address. This proposition, i.e. H_1 in Table 11.1, mirrors 'Insurance Inc.'s version of the facts, which assumes that John Doe actually is the real signer. If the presumed source did sign the document, two different scenarios are plausible. In the first one, $H_{1,1}$, the presumed writer produced a sincere genuine signature, as he is expected to do. This would result in a completely 'standard' signature, the simplest case for a signature examination. In the second one, $H_{1,2}$, the signer tried to produce a disguised signature by consciously altering his signature features. The acknowledgement of a possible disguise process makes the explanation of discordant measurements on the questioned signatures less obvious and may affect the evidence evaluation significantly. In both cases, the presumed signer did actually produce the questioned signature, making them both genuine signatures. The problem lies in the "sincerity" of the signature. The disguised signature is insincere and can be confused with a simulated signature produced by a third person.

The second proposition, H_2, according to the previous scenario, would be John Doe's version of the facts. He generally denies signing the document. The alternative proposition requires the definition of possible alternative scenarios. The straightforward way is to select the negation of the first proposition, meaning that the presumed source did not produce the question signature. However, the remaining possibilities if the presumed source did not sign need to be defined. Two possible scenarios can be considered under this proposition. First, we assume that someone else has produced the questioned signature, but did not have any intention to simulate the presumed signer's signature. This would be just a random similarity between signatures, much like a random forgery. This proposition will be denoted $H_{2,1}$. Second, we consider proposition $H_{2,2}$ when someone else has produced the questioned signature while trying to simulate the presumed signer's signature. In most cases, this would implicate that the forger has had access to some sample of the target signature or at least has some knowledge about it. We could further detail the alternative propositions in order to distinguish the method of forgery used (e.g. freehand, tracing), but this may add unnecessary complexity. The scenario described by $H_{2,1}$ involves a sincere signature, while the one described by $H_{2,2}$ involves an insincere signature. As most important documents contain not only a signature but also other identifying information such as name, address, date of birth among others, a random match with another person's sincere signature seems far-fetched in most cases. This is true at least for Western Europe, where signatures often differ from handwriting and do not necessarily depict the signer's name. The situation may be different in other linguistic regions and cultures. This very same fact also justifies anchoring the alternative proposition on the presumed source's signature. The simulated signature is not any random simulated signature, but should mimic specifically the presumed source's signature. The specificity (or absence thereof) in this statement is the key to choosing an adequate database for the signature evaluation. In practice, we may choose either a database of many different simulated signatures, or specifically simulations of the source's signature. Having a non-specific simulated signature population would eliminate the need for data acquisitions for every case, though this does not take the specific signature's complexity into account. The genuine random match scenario and the simulated signature scenario require a different relevant population to be taken into consideration. In the case under proposition $H_{2,1}$, the questioned signature should be evaluated using a model whose parameters are estimated using a genuine sincere signature population. Vice versa, in the case under proposition $H_{2,2}$, a specific database of simulated signatures of the presumed source's signature should ideally be used, or alternatively a population of non-specific simulated signatures. The BFs derived in Section 11.2.1 according to models MVN, MVK and MVNIW, provide a numerical representation of the impact of available measurements to compare proposition $H_{1,1}$ with proposition $H_{2,1}$. These solutions are based on the assumption of independence between questioned and control material under the alternative proposition. However, it is felt more appropriate to remove such assumptions for signature examination, for the reasons explained above. The proposed approach will be described in Section 11.2.2.2.

The alternative proposition could also be refined if the relevant population is reduced to only a few relevant suspects. One could compare the questioned signature to both the genuine signatures made by

Proposition	Sub-Proposition	Scenario	Wording	Shortened name
H_1	$H_{1,1}$	Sincere	The questioned signature is a sincere signature made by John Doe	genuine signature
John Doe made the questioned signature	$H_{1,2}$	Insincere Disguised	The questioned signature is a disguised signature made by John Doe	disguised signature
H_2	$H_{2,1}$	Sincere Third Party	The questioned signature is a sincere signature made by a third party	randomly matching signature
Someone else made the questioned signature	$H_{2,2}$	Insincere Simulated	The questioned signature is a simulation of John Doe's signature made by a third party	simulated signature

Table 11.1: Generic Propositions for signature evaluation

the presumed source and the simulated signatures made by one or multiple suspects. This, however, bears the risk that the suspects might disguise their simulated signatures. If the control material is unreliable, so are the evaluative results. The problem of behavioral characteristics is that reference materials can also be altered willingly. Moreover, the relevant question here is not whether the signature is a genuine or simulated signature, but rather who is more likely to have produced the questioned signature.

11.2.2.2 The questioned signature model

As mentioned before, the existing proposals for handwriting examination (Section 11.2.1) are not designed to be used in cases where the 'sincerity' of the features is in question. This in particular is the case when a simulator is trying to intentionally recreate the features of someone else's signature. For simplicity, let us assume that John Doe has no reason to disguise his signature. Considering that the personal information filled into the contract matches and identifies John Doe, a sincere random signature can also be excluded. The sole propositions of interest are thus $H_{1,1}$ and $H_{2,2}$ (Table 11.1), meaning the signature is either sincere and was made by John Doe or somebody else has made a simulated signature resembling John Doe's. The Bayes' factor in (11.14) cannot be used to reply to the question of interest here, as it is based on the assumption of independence between sources at the denominator that is no longer reliable here. It follows that one must compute the integral in equation (11.15).

$$BF_{Sig} = \frac{f(y, x|H_1)}{f(y, x|H_2)} = \frac{f(y|x, H_1)}{f(y|x, H_2)}. \tag{11.15}$$

Two different background databases are now needed to inform model parameters under the competing propositions. The first one is a database $\{Z_{ij}\}$ of genuine signatures, as previously seen, the second one is a database of simulated signatures, denoted $\{S_{ij}\}$, where again

$$Z_{ij} \sim N(\theta_i, W_i) \tag{11.16}$$

$$S_{ij} \sim N(\theta_i, W_i). \tag{11.17}$$

Let us now consider a simplified model where a conjugate Normal-Wishart prior distribution is introduced for (θ_i, W_i), that is

$$\theta_i \sim N(\mu, \kappa W_i), \tag{11.18}$$

$$W_i \sim W(U, v), \tag{11.19}$$

where prior beliefs about the population mean take the variability of the observations into account. Parameter κ can be thought as the prior sample size for the mean vector θ. It formalizes the size of the sample from a Normal population providing an equivalent amount of information about θ. The hyperparameters μ and U can be elicited making reference to different background databases (i.e. either of genuine or simulated signatures). The data are distributed according to a Normal distribution:

$$y \sim N(\theta_y, W_y) \tag{11.20}$$

$$x \sim N(\theta_x, W_x). \tag{11.21}$$

The model is slightly different from the previous ones, as the between-source variability is not modeled in this case. Consider first proposition H_1, according to which the questioned signature is a genuine, sincere signature from a given source. It appears logical to choose prior information originating from genuine, sincere signatures. This means using a genuine background population $\{Z_ij\}$ described in Section 11.3.2, to elicit the prior probability distributions. The conditional distribution $f(y|x, H_1)$ is obtained as:

$$f(y|x, H_1) = \int_{\psi} f(y|\psi, H_1) f(x|\psi, H_1) f(\psi|\phi, H_1) d\psi \tag{11.22}$$

where $\psi = \{\theta, W\}$ and $\phi = \{\mu, U\}$. Note that the prior distribution $f(\psi|\phi, H_1)$ has been informed using hyperparameters elicited from a genuine signature population $\{Z_{ij}\}$ (see Section 11.3.2 for a detailed description). The posterior predictive distribution $f(y|x, H_1)$ is available in closed form, as distributions are conjugate and it turns out to be a Multivariate Student t distribution [409]. Consider now proposition H_2, according to which someone forged the questioned signature. A simulated signature will in fact be conditioned on source's signature features. This subsample allows the forger to create a signature that would have a strong, intentional resemblance to the source's. Note that the prior information does no longer come from the authentic signature database $\{Z_{ij}\}$, but rather from a specific simulated signature population $\{S_{ij}\}$. This population contains only simulated signatures of a given source made by many different authors ('forgers'). Detailed information about $\{S_{ij}\}$ are available in Section 11.3.2. The conditional distribution $f(y|x, H_2)$ is obtained as:

$f(y|x, H_2) = \int_{\psi} f(y|\psi, H_2) f(x|\psi, H_2) f(\psi|\phi, H_2) d\psi,$

where $\psi = \{\theta, W\}$ and $\phi = \{\mu, U\}$. Note that the prior distribution $f(\psi|\phi, H_2)$ has been informed using hyperparameters elicited from a specific simulated signature population $\{S_{ij}\}$. The posterior predictive distribution $f(y|x, H_2)$ is available in closed form, as it is a Multivariate Student t distribution [409]. The Bayes' factor then becomes

$BF_{Sig} = \frac{\int_{\psi} f(y|\psi, H_1) f(x|\psi, H_1) f(\psi|\phi, H_1) d\psi}{\int_{\psi} f(y|\psi, H_2) f(x|\psi, H_2) f(\psi|\phi, H_2) d\psi}.$

A schematic representation of the model can be found in Figure 11.2.

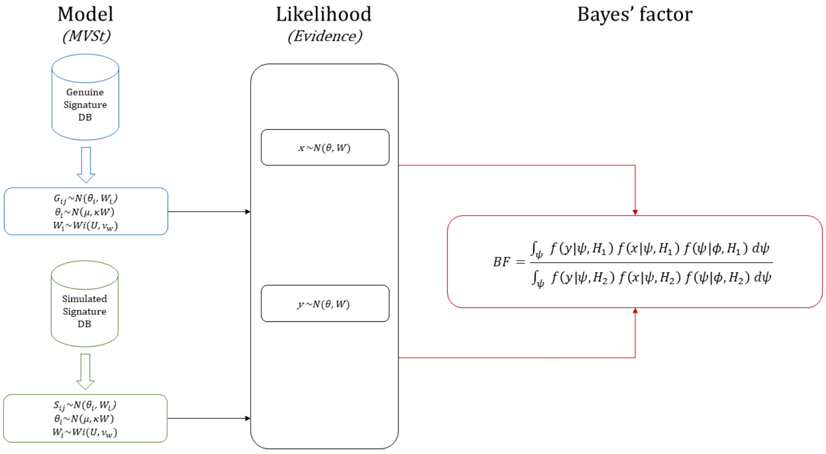


Figure 11.2: Figure 2 - Schematic representation of the MVSt Model. $\{G_{ij}\}$ and $\{S_{ij}\}$ are the genuine and simulated signature databases (DB).

11.3 Materials and Methods

11.3.1 Data acquisition

Signature acquisition was conducted on a Wacom DTU-1141 signature pad, connected to a PC running Windows 7 SP1. Drivers and software for the tablet are the associated Wacom products. The tablet has a writing surface of 283 x 210 mm, with a spatial resolution of 2540 lpi and temporal resolution of 200 Hz. Pen pressure is measured axially and quantified using 1024 levels, while neither azimuth nor altitude are measured. The pen movements above the writing surface are recorded, when close enough to the surface. Data was recorded in a proprietary format, then decrypted with the Wacom software suite and finally the pen data was extracted. The pen data is a chronologically ordered sample of points from the signature, taken approximately every 5 milliseconds. Every data point contains four measurements and some technical input data. Signature recordings include both static (graphical) and dynamic features. Data treatment, visualization and probabilistic evaluation were all carried out in the R statistical software package [466]. Signatures were acquired from several participants who were asked to sign in identical writing conditions. Participants were asked to sit down at a desk with the signature pad in front of them on the horizontal surface. The signature pad could not be inclined (vertically), although participants could rotate the signature pad for comfort. Participants were seated on an adjustable office chair, which they were allowed to freely change. The experiment was done using the equipment provided with the DTU-1141 tablet. The pad surface, stylus and nib were not changed during the experiments. In the present study, two types of signatures were collected: genuine and simulated signatures. One group of 23 individuals was asked to sign their genuine signature 20 times, while a group of 3 people was asked to sign their signatures for 18 months on a regular basis. The simulated signatures were collected through a competition with a prize, in order to provide an incentive to forgers. Simulators could choose to simulate one or multiple signatures of the three reference materials. As for the simulated signatures, forgers were not given any instructions on how to simulate the signature, they were free to choose the 'modus operandi'. Nevertheless, almost all forgers chose to do freehand simulations. Only one participant chose to trace the signatures.

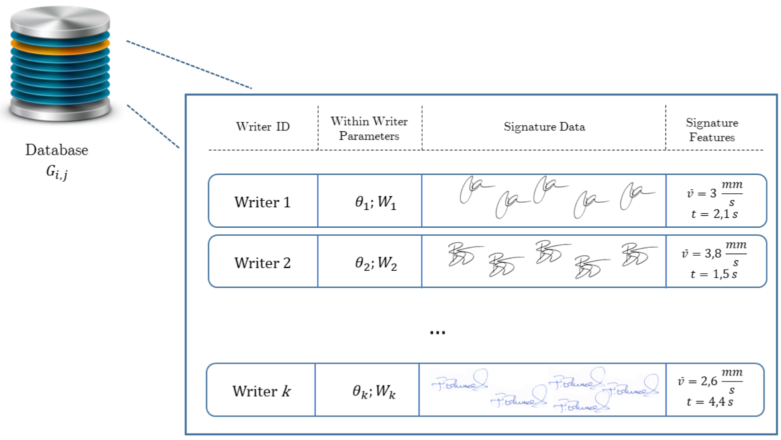


Figure 11.3: Figure 3 - Schematic representation of a signature population. The features $\bar{v}$ and $\bar{t}$ are the average pen speed and the signature duration respectively.

11.3.2 Evidence and background data

The first data source is the 'questioned' signature. Its feature vector is denoted y. This signature is the part of the evidence and the main point of dispute in the case. The purpose of the examination process is to inform about the authenticity of the questioned signature. It is an unknown source specimen, generally a single signature. The second data source is the reference (or control) material from the presumed source. Its feature vector is commonly noted x. The reference collection should be as extensive as possible and follow well-established principles concerning relevance, quantity and contemporaneity. The reference data x is part of the evidence. For the present study, reference materials have been collected from 3 individuals during a 18-month period, with regular acquisition sessions approximately every three weeks. The considered statistical models to infer authorship rely on available background databases that can be used to elicit hyperparameters. A first background database contains genuine signatures that are case unrelated and it is given by a set of authentic signatures collected from m individuals with a total number of n signatures for each one, $\{Z_{(i,j)}\}$, with $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Such a database should include signatures with varied styles and complexities to reflect the general population. In this study, 23 people produced 20 samples of their signature each to serve as a genuine, sincere background population. The purpose of the database of genuine signature features is to inform the prior distribution of model parameters under proposition H_1. The database is schematically represented in Figure 11.3. The second background database contains simulated signatures that are case related and is given by a set of known source simulations of the presumed source's signature from m simulators with a total number of n simulated signatures provided by each one, $\{S_{(i,j)}\}$, with $i = 1, \ldots, m$ and $j = 1, \ldots, n$. The simulations should exclusively relate to the signature of the presumed source. This means that the information contained in the database relates to the specific case only, by conditioning data acquisition on the presumed source. Simulators should aim to produce their highest quality simulations. They should be producing the simulations in conditions closely resembling case circumstances, regarding position, writing implement, substrate, etc. They should have access to several genuine signatures serving as models and time to train prior to acquiring the simulations. Having as many randomly selected simulators as possible should guarantee reflecting a population with both good and bad simulators. The simulations for the alternative propositions can only be collected after occurrence of the case. They are less subject to bias if people unrelated to the case produce the simulations, as they have no stakes in the case and are not likely to underperform intentionally. This database is used to inform the prior distribution of model parameters under the alternative proposition.

Population	Content	Role
y	Measurements from unknown source signature	Evidence
x	Measurements from known source signatures made by the presumed source	Evidence
$\{G_{(i,j)}\}$	Measurements from known genuine signatures (case-unrelated)	Background Information
$\{S_{(i,j)}\}$	Measurements from known simulated signatures (presumed source related)	Background Information
$\{Z_{(i,j)}\}$	Measurements from known genuine signatures from writer Z	Used to draw x

Table 11.2: Summary of all data sources

# of genuine signatures $\{Z_{(i,j)}\}$	670	590	600
# of forgeries $\{S_{(i,j)}\}$	280	400	160
# of forgers	28	40	16
# of forgeries per forger	10	10	10

Table 11.3: Summary of available data per signature; Specific datasets $\{Z_{(i,j)}\}$ and $\{S_{(i,j)}\}$.

All data sources (recovered and control materials), as well as the genuine signature database $\{Z_{(i,j)}\}$ and the simulated signature database $\{S_{(i,j)}\}$ are summarized in Table 11.2. The number of simulated and genuine signatures, as well as the number of distinct forgers for each of the studied signatures can be found in Table 11.3. Every forger provided 10 forgeries.

11.3.3 Methods and experimental conditions

In this article, four probabilistic models are considered and compared in terms of rate of misleading evidence. The first two models are those proposed by Aitken and Lucy [6, 7], with Normal (MVN) and Kernel (MVK) distribution to model the between-source variability. The third model is the one proposed by Bozza et al. (MVNIW), that extends the MVN model to allow for a non-constant within-source variation [56]. Finally, a fourth model (MVSt) has been proposed in the current work. Note that the independence assumption between questioned and reference samples under proposition H_2 is removed only when calculating the BF using the latter model.

The test procedure will rely on mock cases based on known-source data. The cases are specific, which means they use data specific to one presumed source. For example, the genuine data exclusively comes from the population of case-unrelated genuine signatures, the references exclusively come from signer 1 and all forgeries or simulations of signature 1. A description of the test procedure is provided below.

First, a set of 200 questioned signatures is selected by randomly drawing 100 signatures originating from a specific writer from the genuine signature population $\{Z_{(i,j)}\}$, and an additional set of of 100 signatures from the specific simulated signature population $\{S_{(i,j)}\}$. Then, r reference signatures are drawn from the remainder of the genuine dataset $\{Z_{(i,j)}\}$. These reference signatures are identical for all 200 cases within one random trial. Every fictional case is thus composed of a questioned signature (either genuine or simulated) and a set of r reference signatures made by the presumed source. A Bayes' factor is therefore calculated for each fictional case featuring a randomly drawn set of questioned and reference signatures. As a result, 200 BFs will be obtained, and their values and 'nature' (misleading or not) will be recorded. A Bayes' factor greater (smaller) than 1 is in fact expected whenever the questioned material is genuine (simulated). Note that the selected signatures are eliminated from the background data (either genuine $\{G_{(i,j)}\}$ or simulated $\{S_{(i,j)}\}$) in order to estimate the model parameters that are needed for the BF computation. Every experimental condition is repeated k times, to ensure that results

Feature set	Feature 1	Feature 2	Applied to
1	Average speed	Signature duration	All signatures
2	Signature duration	Pressure variance	Signature 1
3	Average pressure	Vertical pen speed variance	Signature 2
4	Time spent with pen lifted	Maximum distance to centroid	Signature 3

Table 11.4: Feature set summary.

Feature Sets	# of random trials per condition k	# of reference signatures r	Background Population
1	10'000	{3,5,8,10,15,20,25,30,50,75,100}	{Genuine,Simulated}
2,3,4	100	{3,5,8,10,15,20,25,30,40,50,60,70,75,80,90,100}	{Genuine,Simulated}

Table 11.5: Experimental condition summary.

are reproducible.

The selected signature features are summarized in Table 11.4. Features were selected as follows. A first feature set (Set 1) given by average speed and signature duration was considered for all signatures, as its features are commonly available and produced acceptable performance over all three signatures. Other feature sets (Sets N°2-4) were selected and tested according to the different signatures. We exclusively used bivariate feature sets in the study. Feature selection itself is a complex endeavor and surpasses the scope of this article. The experimental conditions were varied as summarized in Table 11.5. For each experimental condition, there have been multiple random trials. For feature set 1, 2'000'000 BFs were calculated per experimental condition, while for the other feature sets, 20'000 BFs were calculated in total. Results take the form of averages and variances of all the random trials conducted per experimental condition. Note that while in order to compute the Bayes' factor in eqn. 11.2.2.2 using MVSt both background databases are used, only one database is used to obtain the BF when models MVN, MVK and MVNIW are implemented. Intuitively, the selected database should be one of genuine signatures, as the assumption of dependence between questioned and control material under hypothesis H_2 is not taken into account. A further experiment is performed to investigate the impact, in terms of model accuracy, of the choice of a simulated background database.

11.4 Results

This section is structured by signature. The results are summarized through several figures representing the performance of different models in terms of accuracy (% of correct BFs; 1 - RME). Consider first signature 1 with feature set 2 (i.e. signature duration and pressure's variance) with an increasing number of reference signatures used to specify the control material. The performance of both MVNIW and MVSt models is greater than 80% (Figure 11.4). Overall, the MVNIW obtained the highest accuracy, no matter the amount of control materials available. As far as signatures 2 and 3, model performances are reported in Figure 11.5 and Figure 11.6, respectively. It must be observed that while model performances obtained with signature 2 are in accordance with what observed with signature 1, this is not valid anymore with signature 3, where model MVSt clearly performs better.

A further experiment had been conducted to check the impact on model accuracy once choosing a database of simulated signatures instead of a database of genuine signatures using models MVN, MVK and MVNIW. Figure 11.7 reports the results in terms of accuracy for signature 1. It can immediately

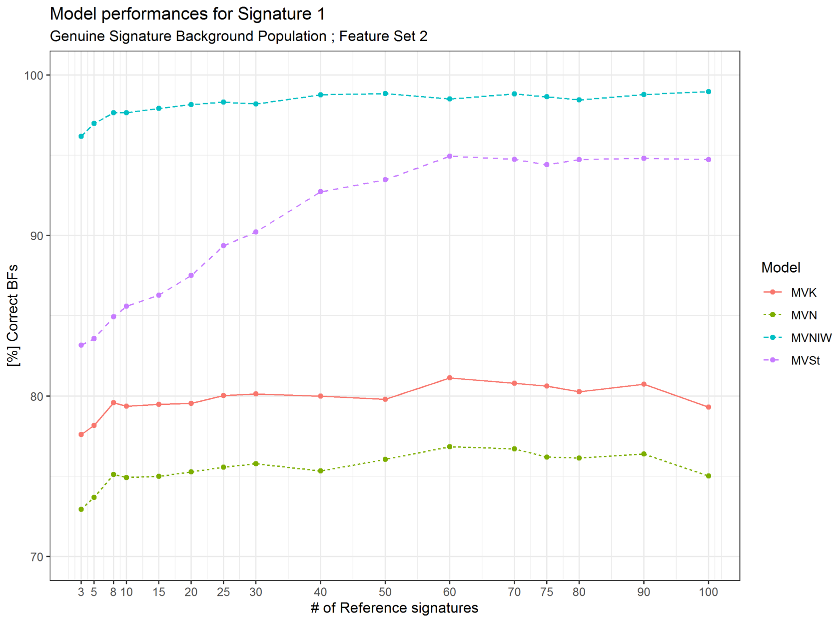


Figure 11.4: Accuracies for Signature 1 and feature set 2.

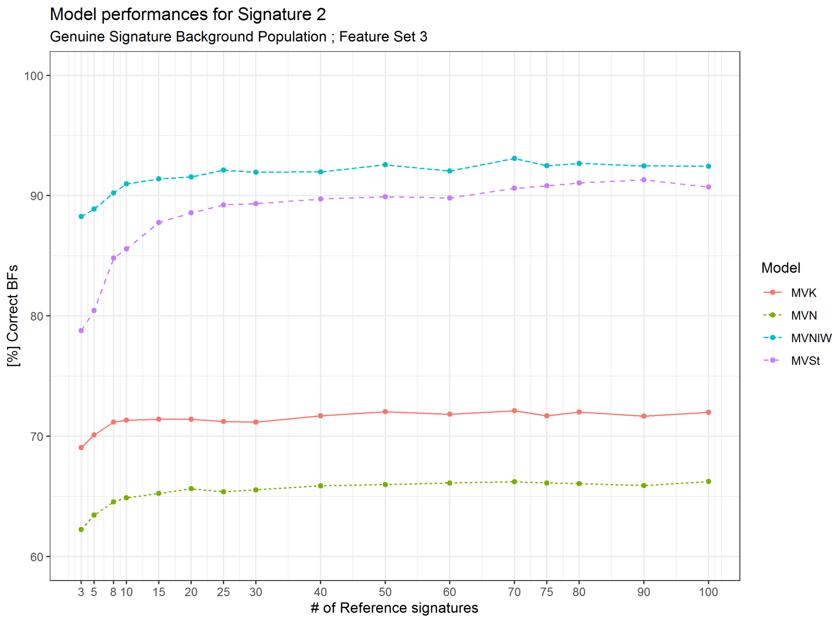


Figure 11.5: Accuracies for Signature 2 and feature set 3

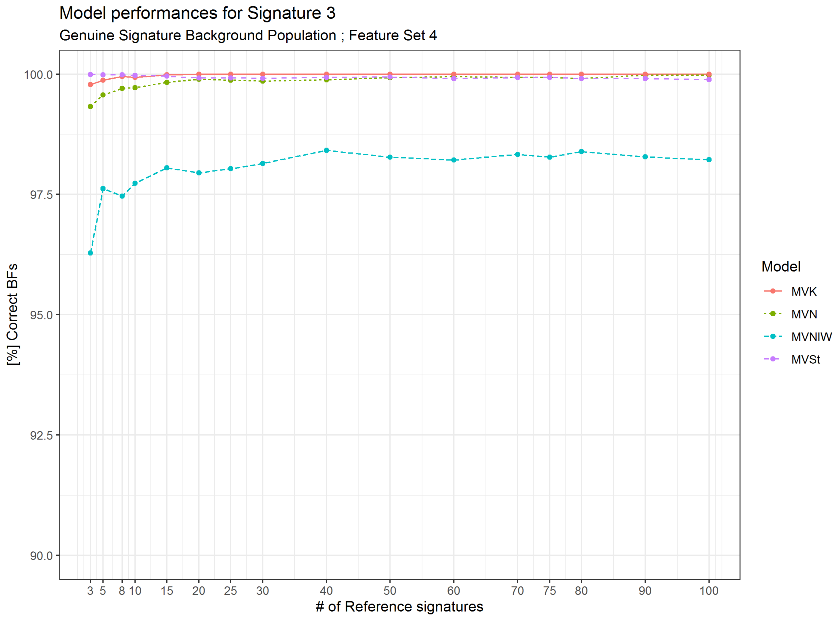


Figure 11.6: Accuracies for Signature 3 and feature set 4

be observed that the performance of such models decreases when changing the background population to simulated signatures. This drop in accuracy is expected, as the previously cited models rely on the assumption of independence to calculate their BF values. The closer the mean and variance of the genuine and alternate population, the less the assumption is able to hold. For the specific simulated signature population, only the MVNIW model, which is 'finer' in its within-writer model, is able to cope with the more challenging population. The MVSt model uses two populations as competing models and therefore does not rely on the independence, but rather on the difference in mean and variability to calculate BFs. In this capacity, it is more robust and specific to signature examination casework.

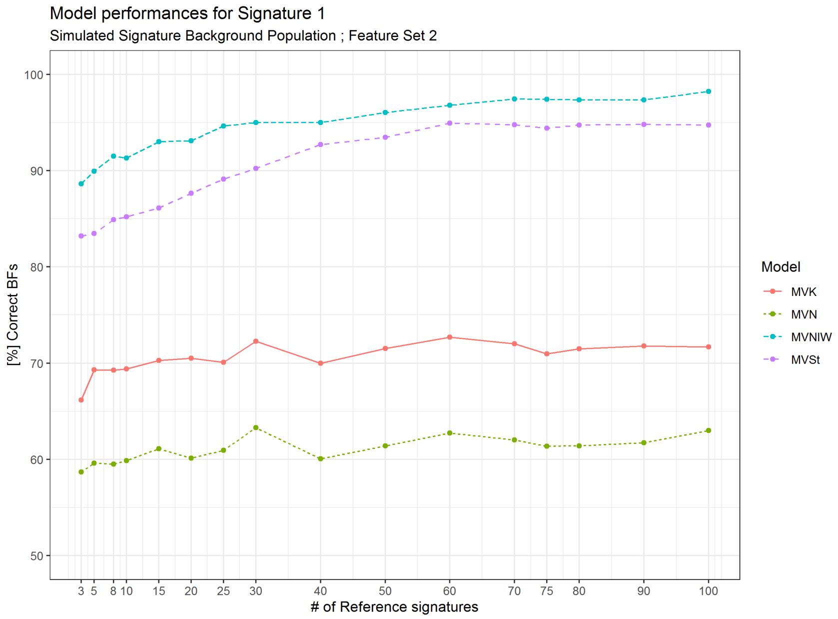


Figure 11.7: Accuracies for Signature 1 with the s_x background data and feature set 2

11.5 Discussion and conclusions

Questioned dynamic signatures represent an emerging topic for forensic document examiners. Signature acquisition by means of digitizing devices allows one to collect several features (e.g. the average speed, or the signature variation) that make it possible to describe a signature in the form of multivariate data. The question of interest is whether selected feature sets are amenable to discriminate between competing propositions related to the origin of the source. Existing Bayesian statistical models for the evaluation of evidence in the form of multivariate data have been taken into consideration, and their performances with reference to dynamic signatures have been explored. Discrimination among competing propositions has been conducted by means of a Bayes' factor, a rigorous concept that provides a balanced measure of the degree to which the evidence is capable of discriminating among competing propositions, as recommended in the ENFSI Guideline [145]. A signature acquisition process/study has been performed, where selected signatures have been used to serve as reference and control material in a hypothetical scenario involving disputed signatures, while the rest has been used as background data to inform prior distributions about model parameters. A different level of accuracy has been observed with reference to different models, background data, feature selection and reference signatures. Results are, however, encouraging, suggesting that selected features collected from dynamic signatures can be discriminative for the purposes at hand.

Model selection clearly represents a key issue. The Bayes' factor computed starting from the reviewed existing statistical models is based on the assumption of independence between sources under the alternative hypothesis. While this assumption is sound in many forensic frameworks, such as those tackled by the reviewed statistical models, this is not so for signature evidence evaluation, where a forger will likely try to reproduce a target signature. A simpler Bayesian statistical model (called MVSt) has been proposed, where the two-level dependence structure is not taken into account, and the marginal likelihoods under competing propositions are available in closed form. Starting from this latter model, a Bayes' factor has been obtained, where the marginal likelihood at the denominator of the ratio takes into account the dependence between sources. At this purpose, it has been necessary to collect a database of simulated signatures, in addition to the one of genuine signatures previously collected. Simulated signatures do not follow the same movement and writing mechanisms as genuine signatures, as the simulator may need to work outside his writing habits, and thus exhibit different variation. This represents a novel aspect tackled by the current proposal. Another key element is the choice of the background data necessary to inform prior distributions about model parameters under the alternative proposition. A FHE could either refer to a generic database of case-unrelated signatures, or to a specific forgery database using the presumed source's signature as model. By using simulated signatures of multiple, case-unrelated signatures, only general statements about whether or not the questioned signature shows any sign of a generic simulated signature can be made. A specific simulated signature population however allows for case specific conclusions. Non-specific information intuitively seems to be less adequate for this purpose than using the case-specific information, but is advantageous in an operational sense. Collecting specific datasets means that a data collection needs to be organized for every single case. Cost-benefit and adequacy of using specific or general a population should be investigated in a further study.

As a future step, a distinct modeling of the within-source and between-source variability should clearly be taken into consideration in the model. These changes may however come with a trade-off between accuracy and data requirements. The simplicity of the MVSt model is also one of its strengths. The reduced number of parameters to be estimated and its simpler structure appear to be the reason for good performances whenever poor background data are available. This is particularly important in forensic examination, where trustworthy material is often scarce.

Feature selection has a large impact on the model performances and must be further investigated. Specific feature selection adapted for the signature at hand is essential in signature examination. Finally, we would like to point out that this study relies on only three genuine signature types. While much intra-writer variation information was collected, simulated signature count and forger populations were limited. Large-scale data acquisition on both model signatures and forgers are future avenues of study worth looking into.

11.6 Acknowledgments

The authors would like to thank Wacom Europe GmbH for providing the software and hardware necessary for performing this study. The authors also thank all volunteers that have participated in this study and agreed to invest their time and effort. The authors would also like to thank the Ecole des Sciences Criminelles for providing the prizes and support for the study.

12 Bayesian evaluation of dynamic signatures in operational conditions

Published. Jacques, L., Silvia, B., Raymond, M. and Franco, T., 2022. Bayesian evaluation of dynamic signatures in operational conditions. Forensic Science International, p.111173.

Abstract

Forensic handwriting examiners (FHE) activities are focused on comparative analysis of handwritten objects such as signatures. Their role is to provide and evaluate evidence for and against the authenticity of a questioned signature. In recent years, cases involving handwritten signatures captured on electronic devices are becoming more commonplace. These so-called ‘dynamic signatures’ (also known as ‘digitally captured signatures’) are much different from paper-based signatures. Not only the medium of recording differs, but also the type, volume of data and features are different from the pattern-based evidence that are paper-based signatures. Recent developments in forensic science – including signature examination - have led to the adoption of evaluative probabilistic methodologies in many disciplines [see, e.g. ENFSI 1915 Guidelines]. In the current paper, a probabilistic model to evaluate signature evidence in the form of multivariate data, as proposed and described in [338], is adopted. Topics like data sparsity, joint evaluation of multiple features and feature selection are investigated. Performed experimental studies showed an accuracy rate above 90% even when a limited number (5) of reference signatures are available. The performances of a multivariate approach are compared with those characterizing a so-called multiplicative approach where variables (features) are taken as independent and the Bayes factor (BF) is obtained as the product of univariate BFs associated to each selected feature. The simplicity of this latter approach is however accompanied by severe issues about the reliability of results. The use of a multivariate approach is therefore highly recommended. Finally, the evidential values in correspondence of alternative feature sets are compared. Results suggest that discriminative features are writer-related and necessitate a case-specific selection.

Keywords: Online Signature, handwritten signature, Forensic Science, Questioned Document Examination, Bayes’ factor, Feature Selection, Biometrics, Multivariate data.

12.1 Research Context

Dynamic signatures are handwritten signatures acquired on digitizers, which capture both temporal and spatial information during the acquisition. Recently, an increasing interest in dynamic signatures has been noted [275, 536, 604], due to many economic advantages linked to paperless approaches [330, 333]. Along with the increasing popularity of dynamic signatures, forensic document examiners have increasingly been faced with cases involving such products and simulated signatures. The role of the forensic scientist is to assess the value of the measurements of signature features under several mutually exclusive scenarios of judicial interest. Forensic scientists have to provide this information in a transparent and intelligible way [141, 284, 489], with many authors advocating probabilistic frameworks and the use of the Bayes' factor (BF) [5, 13, 145, 153, 489, 556].

Recent progress in handwriting examination has been fundamental in restoring trust in a severely critiqued discipline. Much progress was made through the meticulous description of methodology, as well as the validation of expert performance. This was achieved through population studies, as well as the application of computational techniques to strengthen pattern matching examination [105, 106, 199, 254, 293, 362, 384, 505, 506, 524–529, 532, 538, 547, 555, 562, 589, 597]. Reliability claims were supported by statistical analyses related to mock cases built from large datasets containing known source writing [43, 189, 345, 346, 517]. There have also been many advances in methods for validation criteria [240, 380, 472].Further progress in feature characterization could be achieved through computational techniques [199, 287, 351, 379, 381, 413, 473, 524, 570]. Forensic handwriting examination has been able to take advantage from this research [195, 210, 211, 505], but no major methodological breakthroughs related to dynamic signatures have been achieved in recent years.

Despite the existence of a solid methodological framework for handwriting examination [187, 268] and the availability of guidance for communicating conclusions to courts of law [13, 99, 153, 489], only few attempts have been devoted to the quantification of the value of dynamic signatures [194, 201, 531]. Research in this field is often focused on technical aspects such as features' acquisition rather than methodological ones [335]. Experts have been criticized because of the lack of objectivity of their conclusions [482, 485, 499, 501, 502]. Probabilistic models for handwriting evidence evaluation do exist [56], and can be adapted to be used for the forensic examination of digitally captured signatures [338].

In the presented research, three experiments have been performed to investigate the effect of data sparsity, the problem of joint evaluation of multiple features and the feature selection when dealing with the evaluation of signature evidence. Operational conditions often do not permit a large-scale collection of samples and forensic scientists face the problem of handling low volumes of data. Moreover, dynamic signatures offer a large choice of novel features whose correlation and discriminative power is still unexplored and requires investigation. The dataset used in the study is described in Section 12.2.1, while the probabilistic model is summarized in Section 12.2.2. An overview of the selected features, as well as a detailed description of the experimental design is presented in Section 12.2.3 and 12.2.4, respectively. Data analysis and results are reported in Section 12.3 with attention to the model's accuracy and reproducibility face to (i) limited control material (Section 12.3.1), alternative approaches for the joint evaluation of multiple features (Section 12.3.2), and (iii) alternative feature sets (Section 12.3.3). A summary of the research findings and of the benefits of a probabilistic model for digitally captured features evaluation is discussed in Section 12.4.

12.2 Materials and Methods

In this section, the data used in the study (Section 12.2.1) as well as, the probabilistic model for the evaluation of features (Section 12.2.2) and the type of features used in the study are described. Finally, experimental designs for studying the impact of data sparsity, joint evaluation of multiple features and feature's selection are presented (Section 12.2.4).

12.2.1 Data collection

All signatures were acquired in standardized conditions. Participants were sitting at a desk at approximately 1m height and sat on an adjustable office chair. They were allowed to adjust chair height and position for their comfort. They were also allowed to rotate the digitizing tablet to a comfortable angle (on a horizontal plane). The tablet's inclination (vertical) was not changed during the trials; it had to remain flat on the table. A Wacom DTU 1141 signature tablet was used for the data acquisition. The sampling rate of the tablet is 200 Hz with a coordinate resolution of 2540 lpi and 1024 levels of pen pressure measured on the pen axis. Wacom drivers and software were used for data acquisition. Three distinct sets of data have been collected for this study and are classified into: 1) genuine non-case-related signatures, 2) reference signatures (genuine case-related signatures) and 3) simulated signatures (known source forgeries).

Non-case-related signatures For the non-case related signatures, participants were asked to sign their own genuine signature twenty times. Signatures had to be real full-length signatures, with no initials or shortened versions allowed. No selection based on style or complexity was performed. Twenty-three people participated in the collection and produced a total of 460 signatures. This set of signatures is used as background data and is denoted $z_{ij} = (z_{ij1}, \ldots, z_{ijp})$, where $i = 1, \ldots, m$ is the writer's identifier, $j = 1, \ldots, n$ is the number of signatures collected for each writer and p is the number of observed variables. This dataset is denoted by b_g.

Genuine signatures For the reference signatures, three signers who did not participate to the acquisition of non-case-related signatures, were asked to regularly sign during over the length of 18 months, in order to capture natural long and short terms variation. Signers where chosen so to have different styles: a "text-based" (signature 1), a "stylized" (signature 2) and a "mixed" (signature 3) signatures [64]. Their signatures are also characterized by a different graphical and dynamic complexity. These signatures were used to generate fictive cases.

Simulated signatures For the simulated signatures, fifty-seven volunteers (forgers) were asked to simulate at least one of the three types of genuine signatures. Thirty-nine forgers simulated a single signature, eleven forgers simulated two signatures and seven forgers simulated all three signatures. All of the forgers participated in a 'contest' with a reward for the best forgers. This condition was meant to provide incentive to produce the best forgeries possible. No instructions were given regarding forgery strategy. All forgers chose freehand simulations, with only one exception. They were free to train on both paper and tablet for 15 minutes prior to starting the forgery acquisition, starting from six specimens of the chosen genuine signatures. They could keep the reference signatures in their sight during the forgery acquisition. They were asked to deliver 10 simulated signatures for each chosen reference type, with the possibility to discard any of the attempts according to their personal judgment. The collected data for the three case-related signatures is summarized in Table 12.1. These signatures will serve as part of the evidence (so to have fictive cases where the proposition according to which the questioned signature is non-genuine) or as background data that is denoted $s_{ij} = (s_{ij1}, \ldots, s_{ijp})$, where $i = 1, \ldots, m$ is the forger, $j = 1, \ldots, n$ is the number of delivered attempts (forgeries) and p is the number

Type	Signature 1	Signature 2	Signature 3
# Genuine	670	590	600
# Forgeries	280	400	160
# Forgers	28	40	16

Table 12.1: Summary of available case-related data

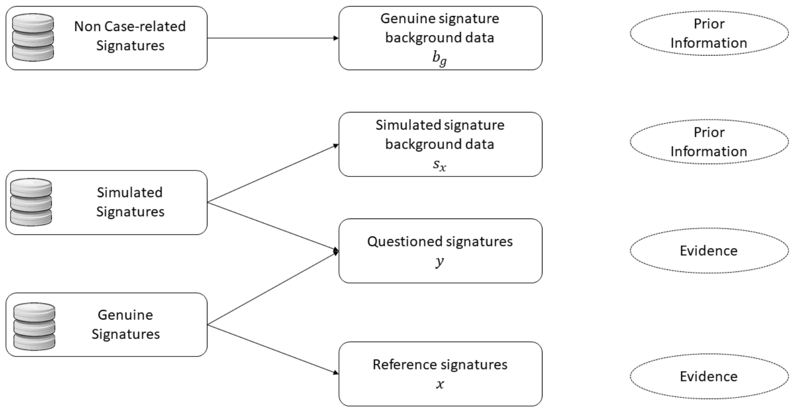


Figure 12.1: Datasets used in this study (column 1) and their role in BF computation (columns 2 and 3).

of observed variables. This information is shortened to s_x. Denote by x the measurements on selected features characterizing one of the three genuine signatures used to generate fictive cases. Denote by y the measurements on selected features characterizing a questioned signature randomly drawn either from the genuine reference signatures (a fictive case under the proposition of a genuine signature) or the simulated signatures (proposition of a non-genuine signature). Features describing the reference and the questioned signatures form the evidence $E = \{x, y\}$. A visual summary of the available databases is provided in Figure 12.1.

Every experiment involves the generation of 'random trials' from known source data. For every trial, there are drawn a total number of q=k+m questioned signatures and a total number of r reference signatures. Questioned signatures are drawn (without replacement) from both simulated signatures (k=100) and genuine signatures (m=100). The drawn signatures are excluded from their respective populations for the rest of the single trial, so that they cannot serve as reference and questioned signature at the same time. Reference signatures are drawn from the remaining reference signatures. The simulated and non-case related signatures that have not been selected as evidential material, form the background data that will be used to estimate the model parameters Φ (see Section 12.2.3). A Bayes' factor is then calculated for each one of the q fictive cases using the probabilistic model that will be sketched in Section 12.2.2. This process, representing one trial, is repeated 10'000 times per experimental condition to ensure the study of the empirical range of reproducibility. A visual summary is provided in Figure 2.

12.2.2 Multivariate dynamic signature evaluation model

In most questioned signature cases, the court's question is to determine the source of the questioned signature. Often, this process boils down to the signature being genuine or simulated. We therefore limited the alternative proposition to simulated signatures. The following pair of "default" hypotheses

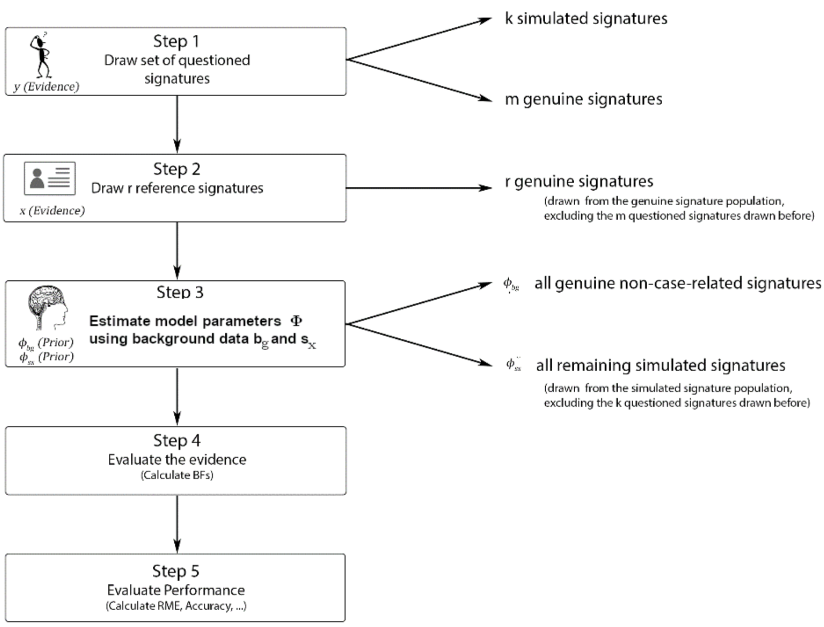


Figure 12.2: Procedural description of one random trial, focusing on sampling and the parameter extraction. One random trial can be seen as being constituted of q=k+m separate mock cases.

can be used in most signature analysis cases:

- H_1 : The signature on the incriminated text (e.g. a contract) is a genuine signature made by a given source;
- H_2 : The signature on the incriminated text (e.g. a contract) is a simulated signature made by an alternative source (i.e. someone other than the given source).

A Normal distribution is assumed for the background data Z_{ij} and S_{ij}, $Z_{ij} \sim \mathcal{N}(\theta_i, W)$ and $S_{ij} \sim \mathcal{N}(\theta_i, W)$, where θ_i is the mean vector and W the covariance matrix. A conjugate prior distribution is assumed for the model parameters (θ_i, W), that is

$$\theta_i|W \sim \mathcal{N}(\mu, \kappa W) \tag{12.1}$$

$$W \sim \mathcal{W}i(U, \nu) \tag{12.2}$$

, (see [338] for a detailed description). The parameter vector (θ_i, W) will be denoted by the greek letter Ψ. The hyperparameters (μ, κ, U, v) are denoted by the Greek letter ϕ. A subscript b_g or s_x will be added to specify whether the parameters are estimated using the background data related to non-case-related signatures z_{ij} (b_g , H_1 is true) or to simulated signatures s_{ij} (s_x, H_2 is true). The distributions of the measurements y and x on the questioned and reference signature are taken to be Normal, $(y|\theta, W) \sim \mathcal{N}(\theta, W)$ and $(x|\theta, W) \sim \mathcal{N}(\theta, W)$. The Bayes' factor can be obtained as:

$$BF = \frac{\int_\Psi f(y|\Psi, H_1) \times f(\Psi, \Phi_{b_g}, H_1) d\Psi}{\int_\Psi f(y|\Psi, H_2) \times f(\Psi, \Phi_{s_x}, H_2) d\Psi} \tag{12.3}$$

12.2.3 Features

In biometrics, features are classified into three broad categories [480]: global features, local features and segment features (see Table 12.2). According to Richiardi et al. [480] one is faced with global features

Feature Type	Data type	Implication
Global (or Parameter) feature	One measurement	Summarizes and reduces information from all measured points into an easily usable unit Loss of information due to the loss of local information
Local (or Function) feature	List of measurements, function (often time functions)	Contains data from every single data point, instantaneous and localized data Every measurement describes one specific point in the signature
Segment feature	One measurement or list per defined unit	The data is segmented according to a specific criterion. (e.g. strokes, inked trail between pen stops, . . .) Every section has their own 'global' and 'local' features. and can be compared unit by unit.

Table 12.2: Biometric feature type classification scheme

if 'one feature is extracted for a whole signature, based on all sample points in the input domain' ([480] at p. 1). If we paraphrase this statement, a global feature summarizes all available information from a signature into a single value that characterizes the entire signature, e.g. the average pressure, the pen velocity variance, the maximal velocity, the signature length or the signing time. A brief description of segment features can be found in Table 1. Local features are a chronological sequence of measurements and contain much more information. So, their treatment is more complex [74]. Segmental features are a hybrid case, which surpasses the scope of this article. This article focuses on global features only. There are several reasons for this choice. First, global features do not require any segmentation or algorithmic treatment. A global feature-based methodology can be extended to any signature. Global features can be measured on dynamic, static [58] and paper-based signatures and are universally applicable. Second, global features complement pattern matching methods currently used by forensic examiners. Forensic examiners traditionally focus on the visual information, such as the shape of the signature. Third, global features can be reproducibly measured.

A total of 60 global features were extracted from each examined signature. Table 2 contains a list of the 12 measured global features used in the study. Table 3 contains a list of 16 local features, which can be summarized into 48 global features by averaging, as well as calculating their variance and their maxima (e.g., the tangential speed dt1 becomes dt1_mean, dt1_var, dt1_max). Features were grouped into classes based on the type of measurement they originate from (e.g. distance, time, or velocity). This classification is useful to explain what feature class may be prevalent for discriminative purposes and useful in practice.

Features were not considered separately. The combination of features is referred to as a feature set. To select a list of relevant features, all possible combinations of features have been tested and their performance in terms of accuracy has been analyzed. Feature sets were ranked via a performance criterion that will be described in Section 12.2.4.

All of these features were used in multivariate feature sets. Features were not considered separately when performing analysis, but multiple features and their covariance were used simultaneously. We refer to the combination of features as a feature set. Put differently, a feature set is a multivariate dataset, defined as a vector of n features. To select features, we simulated all possible combinations of feature sets and tested them separately. Feature sets were ranked via a performance criterion defined below. This criterion takes into account two types of information: Overall 'error', through the sum of the rates of

Feature	Description	Feature class
Height	Height, measured vertically from minimum to maximum point	Expansion Feature
Width	Width, measured horizontally from left- to rightmost point	Expansion Feature
WHRatio	Ratio of Width to Height	Expansion Feature
Uplength	Length of in-air movement trajectory	Length Feature
Downlength	Length of on-surface movement trajectory	Length Feature
Totlength	Total length of trajectory	Length Feature
Totaltime	Time to finish for signature	Time Feature
Downtime	Time pen is touching the tablet	Time Feature
Uptime	Time pen is lifted	Time Feature
DownTot	Ratio of Down- to Totaltime	Time Feature
UpTot	Ratio of Up- to Totaltime	Time Feature
DownUp	Ratio of Down- to Uptime	Time Feature

Table 12.3: Measured global features

Feature	Description	Feature class
dt1	Tangential Speed	Speed Feature
dt2	Tangential Acceleration	Acceleration Feature
dt3	Tangential Jerk	Jerk Feature
dx1	Horizontal Speed	Speed Feature
dx2	Horizontal Acceleration	Acceleration Feature
dx3	Horizontal Jerk	Jerk Feature
dy1	Vertical Speed	Speed Feature
dy2	Vertical Acceleration	Acceleration Feature
dy3	Vertical Jerk	Jerk Feature
dp1	First-order derivative of pressure (time)	Pressure Feature
dp2	Second-order derivative of pressure (time)	Pressure Feature
dp3	Third-order derivative of pressure (time)	Pressure Feature
TVD	Angle of velocity to the horizontal axis	Directional Feature
TAD	Angle of acceleration to the horizontal axis	Directional Feature
P	Pressure intensity, measured axially (pen inclination)	Pressure Feature
XY	Distance to coordinate centroid	Expansion Feature

Table 12.4: Measured local features

misleading evidence, and the reproducibility of results, through the variance of the accuracy. The lower the value of the criterion, the higher the overall performance of the feature set.

$$PerfCrit = RME \times \sigma^2_{\text{Accuracy}} \tag{12.4}$$

12.2.4 Experimental Setup

Three separate experiments are performed in order to (i) investigate the effect of data sparsity, (ii) compare a multiplicative approach for the joint evaluation of multiple features to a multivariate approach, and (iii) analyze the discriminative power of different feature sets. Performances of the performed experiments can be studied through accuracy and reproducibility. Accuracy is expressed as the proportion of BF values correctly supporting the hypothesis known as true or, analogously, by the rate of misleading evidence (RME) [13]. If misleading evidence occurs, it can be said to favor either H_1 or H_2. Here, misleading evidence towards H_1 signifies falsely supporting the proposition according to which the signature is genuine. This is denoted by $RME(H_1)$ and is obtained as the ratio between the total number of BFs greater than 1 (i.e., supporting H_1) and the total number of fictive cases where the questioned signature is known to have been simulated (i.e., H_2 is true). If evidence is misleading towards H_2, the BF falsely supports the proposition according to which the signature is a simulation. This is denoted by $RME(H_2)$ and is obtained as the ratio between the total number of BFs smaller than 1 (i.e., supporting H_2) and the total number of fictive cases where the questioned signature is known to be genuine (i.e., H_1 is true). It can be reasonably requested that a model should have a high value of average accuracy across trials and a low value of the accuracy variance across trials (reproducibility):

$$\frac{s_A}{|\bar{x}_A|}, \tag{12.5}$$

where s_A is the standard deviation of accuracy values across trials and $\bar{x}_A$ is the average accuracy across trials. The lower the value of the coefficient of variation in 12.5, the higher the overall performance of the model. A limitation frequently encountered in forensic casework is data sparsity. In experiment 1, the number of reference signatures drawn for each trial varies from a minimum of 5 signatures to a maximum of 100 signatures. For every signature, the four-highest ranked tri-variate feature sets are retained and are used in 100 random trials per experimental condition.

The importance of correlation among features is investigated in addition to the role of model dimensionality. In experiment 2, the performances of a multivariate versus a multiplicative approach for jointly evaluating single global features are analyzed. Scientific literature on handwriting evidence [293, 589] proposed to consider features as independent and to calculate the value of the evidence by multiplying the single evidential value assignations (BFs). In signature evidence, however, features appear to be strongly correlated due to movement mechanics [139, 598] and the assumption of independence can be hardly justified. Features showing a better performance when considered singularly where retained. These features were then used to calculate the BF (i) by multiplying BFs associated to each variable (feature), and (ii) using the multivariate approach illustrated in Section 12.2.2. Computations were performed on the same case data (questioned and reference signatures) to ensure comparability. The effect of adding variables in a multivariate model is also of interest. Adding variables increases the model dimensionality, but it is also expected to improve performance. Another daunting problem is the selection of discriminative features for the detection of simulated signatures. A total of 34'220 unique feature sets of size 3 generated by combining the 60 features presented in Table 12.3 and Table 12.4. The model dimensionality is set equal to 3 variables to keep the computational time short. The feature sets' performance was evaluated through 1'000 random trials for each set. These computations were repeated for nine different reference set sizes, ranging from 10 to 160 reference signatures. For every experimental condition, the ten best performing feature sets were selected. Then, for each signature the percentage of cases where a specific feature is included among the best performing sets was calculated. This percentage expresses how useful a specific feature is for that particular signature.

12.3 Results

12.3.1 Experiment 1 – Validation Study and the Effect of Data Sparsity

The four best performing feature sets were selected for each reference signature. It must be underlined that they varied across signatures. The average accuracy of the selected feature sets for the three reference signatures is represented in Figures 12.3 to 12.5, while the accuracy variance is represented in Figures 12.6 to 12.8. Most feature sets have accuracy above 90%, with few exceptions where the reference signature is small (see Figures 12.3 and 12.4). As the number of reference signatures increases, a higher accuracy and a lower variability may be observed. This is not true, however, for very large sample sizes. In some cases, a modest decrement in terms of mean accuracy and an increment in terms of variability have been observed. See for example signature 1, where the feature set including the lifted pen trajectory length, the horizontal speed variance and the pressure variance, shows increased accuracy variance increases when more than 40 signatures are used for training. In general, a total number of 10 to 15 signatures appear as sufficient materials to produce reliable and reproducible results. A lower sample size may simply be too small to sufficiently represent the signature natural variation. The lower accuracy for very large sample sizes may be explained by a high variability of the reference signatures, that is not adequately modeled. The background data of non-case related signatures used to estimate model parameters under hypothesis H_1 might not be sufficiently representative. In presence of graphically different signatures, a viable alternative could be to divide them in comparable subclasses. Moreover, it must be added that the best performing feature sets selected for signature 2 are mostly given by dynamic features, which may be subject to greater natural variation. They may also be more sensitive to writing conditions such as posture and pen-pad-interaction or the writer's physical and psychological state (e.g. sickness, stress, threat, medication, narcotics, ...) with respect to static features. Larger efforts should therefore be devoted in the collection of background data.

12.3.2 Experiment 2 – Combining Feature Information through the Bayes Factor

Forensic examiners have always argued that no single element in signatures is sufficient for detection of simulated signatures [56], and that multiple features should be observed and combined. Intuitively, one may expect that using more information always yield better results. This is not always the case. A larger model dimension is not necessarily accompanied by performance improvements, as some variables may turn out to be redundant or meaningless. These increase model complexity unnecessarily without providing additional benefits. Moreover, the greater the number of variables, the larger will be the size of the background data needed for parameter estimation in a multivariate approach. Such datasets may not be available to the forensic examiner. The effect of increasing the dimensionality of the feature set can be seen in Tables 12.5 to 12.7. The incremental addition of features is generally accompanied by an increase in accuracy and reproducibility for all reference signatures. For signature 1, the average accuracy increases of about 2% when comparing univariate versus quadrivariate feature sets, while the

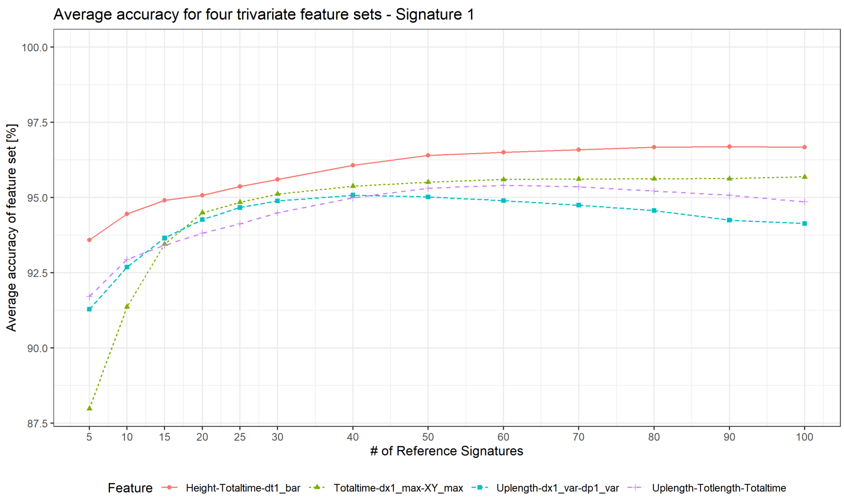


Figure 12.3: Average accuracy accuracy for the four best performing trivariate feature sets over the 100 trials per experimental conditions for signature 2. For more detail on the features, see tables 12.3 and 12.4.

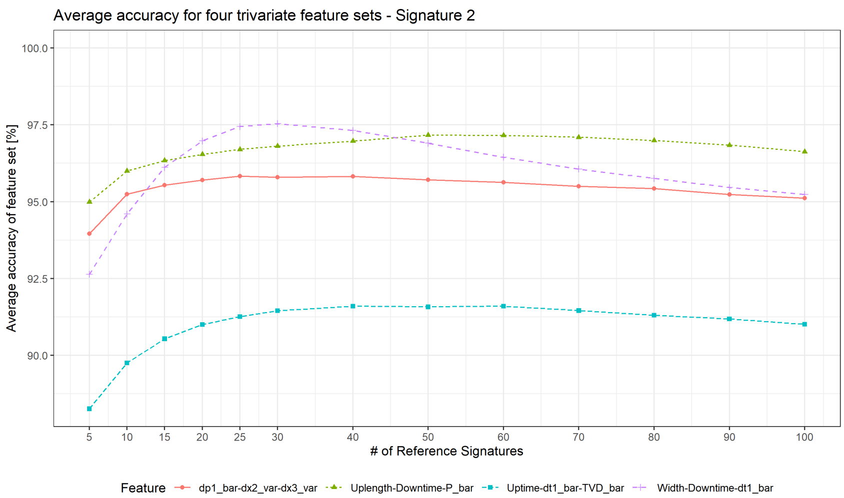


Figure 12.4: Average accuracy accuracy for the four best performing trivariate feature sets over the 100 trials per experimental conditions for signature 2. For more detail on the features, see tables 12.3 and 12.4.

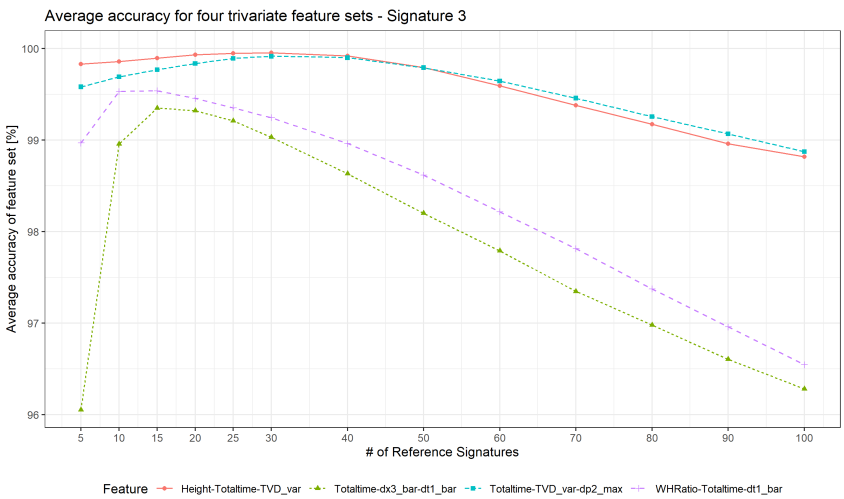


Figure 12.5: Average accuracy accuracy for the four best performing trivariate feature sets over the 100 trials per experimental conditions for signature 3. The four feature sets show a small increase in accuracy between 5 and 15 signatures. For more detail on the features, see tables 12.3 and 12.4.

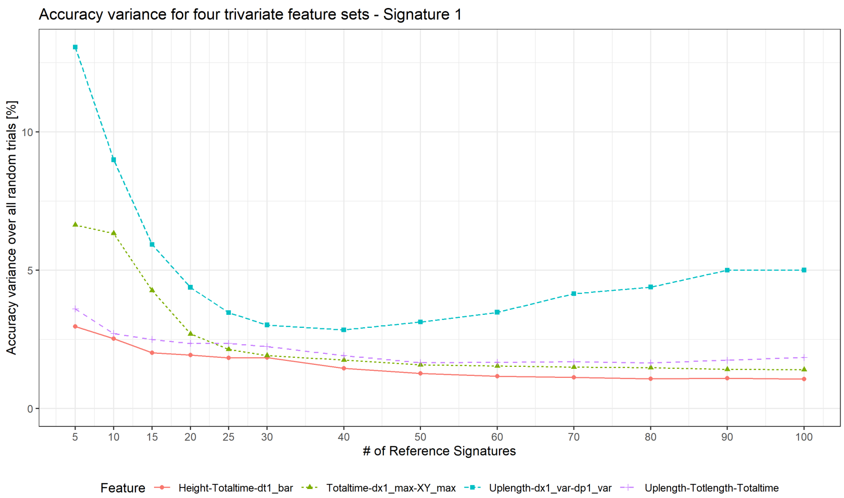


Figure 12.6: Accuracy variance for the four best performing trivariate feature sets over the 100 trials per experimental conditions for signature 1. For more detail on the features, see tables 12.3 and 12.4.

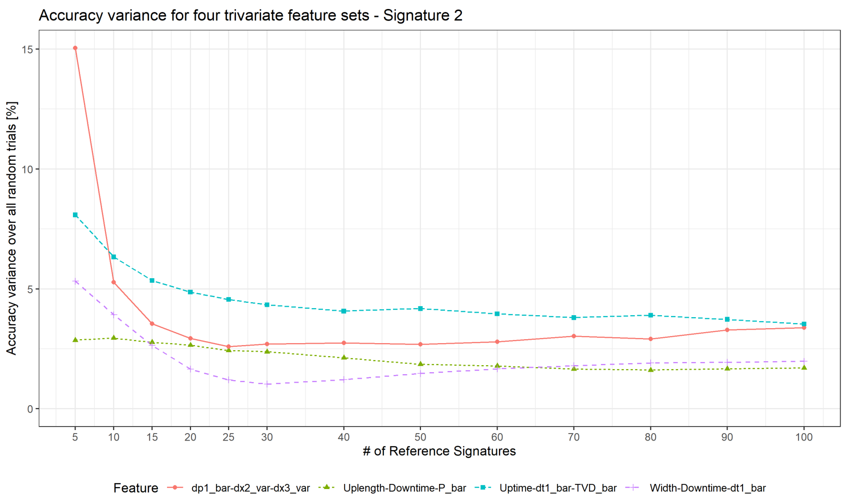


Figure 12.7: Accuracy variance for the four best performing trivariate feature sets over the 100 trials per experimental conditions for signature 2. For more detail on the features, see tables 12.3 and 12.4.

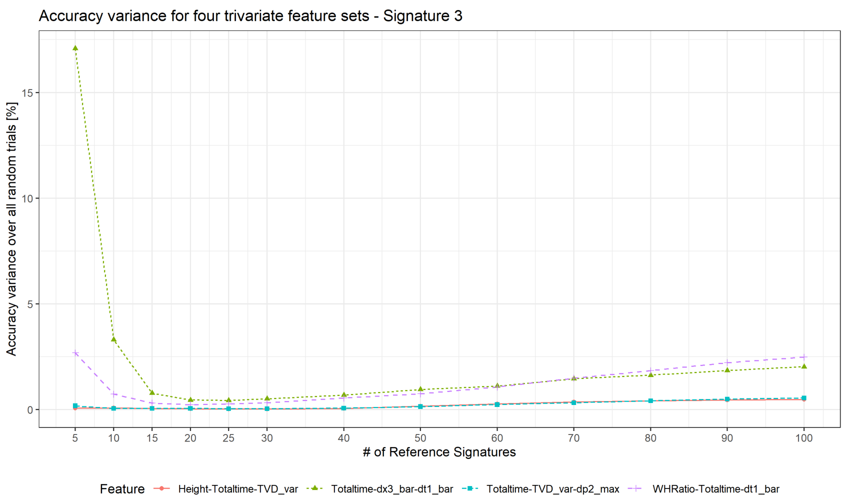


Figure 12.8: Accuracy variance for the four best performing trivariate feature sets over the 100 trials per experimental conditions for signature 3. For more detail on the features, see tables 12.3 and 12.4.

	Signature 1			
	Univariate	Bivariate	Trivariate	Quadrivariate
Feature set	Totaltime	Totaltime, P_var	Totaltime, dp3_max, XY_max	Totaltime, dt1_var, TAD_var, XY_max
Accuracy [%]	95.26	95.88	96.03	97.42
Acc. Variance [%]	0.96	0.80	0.69	0.50
RME_{H_1} [%]	8.80	7.59	7.65	4.35
RME_{H_2} [%]	0.49	0.5	0.12	0.72

Table 12.5: Signature 1: model performances with feature sets of increasing dimension. Note that the number of questioned samples is not equal for H_1 and H_2 in these trials. Accuracy is the complement to the weighted average of RME.

	Signature 2			
	Univariate	Bivariate	Trivariate	Quadrivariate
Feature set	dt1_bar	P_bar, dy1_var	Totaltime, dp3_max, XY_max	Totaltime, dt1_var, TAD_var, XY_max
Accuracy [%]	88.39	91.97	95.43	96.63
Acc. Variance [%]	1.42	1.44	4.09	2.20
RME_{H_1} [%]	12.76	7.38	6.75	4.86
RME_{H_2} [%]	9.68	9.12	0.89	0.86

Table 12.6: Signature 2: model performances with feature sets of increasing dimension. Note that the number of questioned samples is not equal for H_1 and H_2 in these trials. Accuracy is the complement to the weighted average of RME.

accuracy variance drops off slightly. The gain in terms of accuracy is mostly due to a decrement of the misleading evidence versus H_1 ($RME(H_1)$). This means that fewer simulated signatures produced misleading BFs (i.e. $BF > 1$ when H_2 holds). A greater accuracy for models of larger dimension is also observed for signature 2. For feature sets of at least three variables, almost no genuine signatures produced misleading evidence (i.e. no BFs smaller than 1 have been observed when H_1 holds). However, an increasing variability is observed for trivariate feature sets. For signature 3, the expected increment in terms of average accuracy and decrement in terms of variability is confirmed. Moreover, no simulated signature produced misleading evidence.

The model performances reported in tables 12.5 to 12.7 are related to experimental studies where the BF has been calculated as in (1) using the multivariate statistical model in Section 12.2.2. However, there are far more naïve and faster ways to calculate the BF for each fictive case. If variables (features) are assumed to be independent, the BF can be obtained by multiplying the BFs calculated in correspondence of each feature treated singularly. The performances of the two best performing features for each signature are reported in tables 12.8 to 12.10. Features are considered singularly (rows 1 and 2) or jointly (rows 3 and 4), using a multivariate approach (row 3) or a multiplicative approach (row 4). Instead of improving the accuracy, naïvely combining features may even decrease it. The only exception is the multiplicative approach for signature 3. The loss in terms of accuracy can be explained by the shared information content between features. In fact, it is better to combine features that are 'worse' individually but contain 'different' information, rather than combining 'good' but 'similar' features. This can be seen in tables 12.5 to 12.7, where adding a second variable in the multivariate setting actually improves accuracy. Feature selection should be based on features providing complementary information, which is apparent

	Signature 3 Univariate	Bivariate	Trivariate	Quadrivariate
Feature set	Uptime	Uptime, XY_max	Uptime, dt3_var, XY_max	Totaltime, dx1_bar, dy1_max, XY_max
Accuracy [%]	98.75	99.88	99.90	99.99
Acc. Variance [%]	0.44	0.06	0.04	0.00
RME_{H_1} [%]	0.00	0.00	0.00	0.00
RME_{H_2} [%]	1.93	0.19	0.15	0.01

Table 12.7: Signature 3: model performances with feature sets of increasing dimension. Note that the number of questioned samples is not equal for H_1 and H_2 in these trials. Accuracy is the complement to the weighted average of RME.

Signature 1	Accuracy [%]	RME_{H_1} [%]	RME_{H_2} [%]
Univariate | Totaltime	91.01	8.92	0.08
Univariate | Uptime	86.99	12.49	0.53
Multivariate	88.75	10.41	0.84
Multiplication	88.82	11.03	0.16

Table 12.8: Signature 1. Accuracy and rate of misleading evidence obtained with a univariate, multivariate and multiplicative approach for the two best performing singularly evaluated features (i.e., totaltime and uptime).

from tables 12.5 to 12.7. The best performing feature sets generally include a time-related, a graphical and one or several dynamic features. Two different timing features (such as time spent with the pen lifted and time spend with the pen writing) convey to much common information to be good choices for modelling. The decrease in value is an expected results because of the high covariance between the combined features, which impacts the BF. It should be noted that multiplication models are simply dominated by the extreme elements and do not allow for such diagnostics.

The model dimension and the method used to evaluate jointly multiple variables, also affect the magnitude of the Bayes' factor values. Consider a feature set of size equal to 2. When comparing the multivariate with the multiplicative approach, the BFs obtained with the latter tend to be more extreme than those obtained with their multivariate counterparts that are more tempered (this has been observed in 94.5% of cases). This means having higher values when H_1 is supported and, vice versa, lower values when H_2 is supported. Clearly, an extreme BF obtained in this fashion is not indicative of high reliability.

On average, the "multiplied" BF values for signature 1 are 17 and 2.75 million times higher (for BFs supporting genuine and simulated propositions, respectively) than their multivariate counterparts. An interesting way to study these effects further may be to investigate their mutual information content. As such, features providing various information about the time of execution are highly correlated and the multiplicative approach yields BFs of increased magnitude with respect to a multivariate approach where the dependence structure is taken into account (e.g, for signature 1, the total time of execution, Total time, and the time spent with lifted pen, Uptime, see Table 12.8). Similar observations are valid for all signatures. For signature 2, the best performing features resulted to be the average tangential speed (dt1_bar) and the horizontal speed variance (dx1_bar). The multiplied BF values are 2.9 and 3.6 times higher (for BF supporting genuine and simulated propositions, respectively) than their multivariate

	Signature 2		
	Accuracy [%]	RME_{H_1} [%]	RME_{H_2} [%]
Univariate | dt1_bar	87.64	11.79	0.57
Univariate | dx1_var	90.16	7.16	2.69
Multivariate	86.49	11.89	1.63
Multiplication	88.66	10.75	0.60

Table 12.9: Signature 2. Accuracy and rate of misleading evidence obtained with a univariate, multivariate and multiplicative approach for the two best performing singularly evaluated features (i.e., mean of tangential speed and variance of horizontal speed).

	Signature 3		
	Accuracy [%]	RME_{H_1} [%]	RME_{H_2} [%]
Univariate | Uptime	99.83	0.00	0.17
Univariate | dx1_bar	98.94	1.04	0.03
Multivariate	99.68	0.00	0.32
Multiplication	99.95	0.00	0.06

Table 12.10: Signature 3. Accuracy and rate of misleading evidence obtained with a univariate, multivariate and multiplicative approach for the two best performing singularly evaluated features (i.e., uptime and mean of horizontal speed).

counterparts. Finally, the best performing features for signature 3 resulted to be the time spent with lifted pen (Uptime) and the average horizontal speed (dx1_bar). The multiplied BF values are 5 and 3.4 times higher (for BF supporting genuine and simulated propositions, respectively) than their multivariate counterparts. These results confirm the expected overestimation of evidential value that can be made when correlated variables are treated as independent. Overall, the multivariate approach is a more coherent way to quantify the value of correlated signature features [13]. The so-called multiplicative strategy is not necessarily a good choice for signature evidence, as it tends to deliver an unrealistic assessment of the evidential value. In some cases, such as with handwriting evidence, this effect may be small enough to be neglected.

12.3.3 Experiment 3 – Feature Selection and Discriminative Features

Following the previous discussion about joint evaluation of multiple features, the objective of experiment 3 was to search for the best performing feature types and feature sets useful for discriminative purposes in presence of questioned signatures. For well-performing feature sets, the idea was to determine the single feature's contribution, as an indirect measure of discriminative power. Features were classified on the basis of the type of measurement. The contribution of features was measured by calculating the percentage of cases where the single feature is included in one of the 10 top ranked feature sets for every experimental condition (Table 12.11). The size of the control materials varied from a minimum of 10 signatures to a maximum of 160 signatures. The overall performance was measured in terms of accuracy and reproducibility, as detailed in Section 12.2.4. This approach has, however, some limits, as it cannot express the 'importance' of the contribution, nor directly express the complementarity of the features.

The contribution of feature types varied with the reference signature and its complexity. If time-based

	Signature 1	Signature 2	Signature 3
Type	Mixed Style	Stylized	Text-based
Complexity	Medium	Low	High
Time Features	100%	72.00%	46.22%
Length Features	0%	24.00%	0.00%
Expansion Features	20.44%	35.11%	98.22%
Directional Features	2.67%	16.00%	0.00%
Pressure Features	87.56%	60.89%	63.11%
Speed Features	16.89%	40.44%	4.89%
Acceleration Features	12.44%	13.78%	24.00%
Jerk Features	5.78%	9.78%	36.44%

Table 12.11: Summary of feature type contributions to the best performing feature sets. Percentages express in how many of the best performing feature sets a specific feature type was included.

features showed a high contribution in either signature 1 and 3, in signature 2 the best performing feature sets tend to privilege dynamic information (such as pressure, speed, acceleration, jerk). Although the set of genuine signatures is not sufficiently big to establish a direct relationship between signature complexity, type and discriminative features, results do suggest patterns. The performed exploratory analysis would suggest that for short, stylized and rapidly executed signatures using dynamic data may be of interest. Signatures 1 and 3, which are longer, slower and more legible have feature sets that rely more heavily on time-related information. Best performing feature sets are peculiar for each signature: no feature set appeared in the top list for more than one signature. Unfortunately, there does not exist an optimal feature set independently on the feature type. This appears clearly in tables 12.12 and 12.13, where the experimental results obtained for the three reference signatures on a same set of features is reported. The feature sets were chosen because they performed well and were common to the three signatures in the top 100 features sets. They were also chosen for generality, with the features (Time, pressure, speed, direction) being available on most dynamic signature hardware. Although the results in table 12.12 may seem to suggest that these feature sets work well, they are not 'optimal'. Further, the rates of misleading evidence and their balance, therefore the bias of the feature set, varies a lot. Vastly superior results can be achieved by signature specific feature selection. As an example, Signature 3 had trivariate feature sets with 99.9% accuracy. These results are however not reproducible when using the feature set on another signature. Optimal performance is only achieved by careful, specific selection of features. However, some feature sets appear as good compromises between applicability and performance. The present observations suggest that no generally transposable feature-set exists. Performed experiments confirm however that some feature types may be privileged for short and fast signatures, while others are better suited to legible and long signatures. Dynamic features, such as pressure and speed may be more informative for short signatures and time-related features for longer signatures. Additionally, for left-handed writers, incorporating direction-related features, in particular trajectory direction may be of interest. Clearly, a more extensive study on a large panel of genuine signatures with different complexities and of different styles should be conducted in order to investigate further the robustness of observations and remarks.

	Accuracy [%]	RME_{H_1} [%]	RME_{H_2} [%]
Signature 1	91.05	8.67	9.23
Signature 2	93.54	9.57	1.13
Signature 3	92.07	0	12.22

Table 12.12: Results for all signatures with the same feature set. Features are {Totaltime, dt1_bar, TVD_bar}

	Accuracy [%]	RME_{H_1} [%]	RME_{H_2} [%]
Signature 1	95.86	7.51	0.60
Signature 2	83.78	25.80	0
Signature 3	97.98	0	3.12

Table 12.13: Results for all signatures with the same feature set. Features are {Totaltime, P_var, dp3_max}

12.4 Conclusion

The present study explored evidence evaluation in dynamic signature examination under operational conditions. Three specifics aspects were addressed: data sparsity, feature combination and feature discriminative power. A Bayesian parametric model was used to calculate Bayes' factors expressing evidential value on signature authenticity. The approach incorporates natural variation and feature rarity using empirical data. Experimental results obtained using global features showed that roughly 15 signatures are sufficient to obtain accurate and reproducible results. However, performances are still good even in presence of lower sample sizes. The joint evaluation of multiple features represents another important issue. A multiplicative approach where variables are assumed independent tends to produced more extreme evidential value statements. Features characterizing dynamic signatures show however a non-negligible correlation, and a multivariate approach is to be preferred. As far as the feature selection, it has been observed that the best performing feature sets are signature specific.

When interpreting the results of this study, one must keep in mind that the number of forgers involved is modest. The same can be said for the case-related signature number. Only three signatures of different styles were examined in detail. Although results cannot be generalized due to the absence of a database of similar styled signatures, the in-depth analysis of the three samples has at least exploratory value. These three signatures illustrate that different signature complexity and type may play a role for feature selection and necessitate contextual information and special care for this step. Short, stylized signatures may perform better with dynamic features, while longer signature types may show better perform with time-related features. For a left-handed individual, directional features such as the velocity and acceleration direction appear to be more discriminative. It cannot be excluded that generalization may be possible for signature classes based on styles or graphical elements. The results suggest that feature selection should be case-specific. Some features types seem to be inherently better suited to different signatures. Further studies should be implemented to study features' discriminative power with respect to signature complexity and type, as well as the signer's handedness.

The feature sets tested in this scenario included exclusively global features. These feature sets contain limited information on the signature's dynamics. The available raw data contains local features, such as the speed and pressure profile of the signature. Summarizing these features into global features produces a loss of information. Another point that is worth to mention is that dimensionality plays an important role in multivariate statistics [624]; the higher the model dimensionality, the more data is needed to assess feature variability and estimate model parameters. Realistically speaking, forensic scientists rarely dispose of large background data, that might therefore be not adequate for models of high dimension. Alternatively, a score-based approach might be implemented. This would allow to shrink the model dimensions and may turn out to be a viable alternative to deal with local features.

An additional consideration should be devoted to the temporal proximity between the questioned and the reference signatures. In the current study, however, the contemporaneity of the evidence material was not taken into account.

Finally, the proposed Bayesian model has shown good accuracy under operational conditions. The model provides a methodologically sound way to assess dynamic signatures through the use of empirical data and statistical techniques. It can also be easily implemented for handling features measured on static or paper-based signatures. Criticism to forensic science includes the lack of validated statistical models and of adequate empirical data to justify conclusions given by examiners. The model meets recent recommendations for communicating evidence and it adheres to the important principle of transparency [141]. The model is versatile and applicable under operational conditions, and it may represent a valuable contribute to practice in handwriting examination. It may help the forensic handwriting examiner in providing - through the Bayes' factor - a quantified support in favor or against a set of hypotheses about the origin of questioned documents. This numerical value guarantees that examiner's conclusions are sound because supported by strong scientific principles.

12.5 Acknowledgments

The authors would like to thank all volunteers that have provided genuine and/or simulated signatures. Their valuable contribution made this study possible. The authors would like to thank Wacom Europe GmbH for providing the soft- and hardware necessary for the study.

13 The influence of time on dynamic signatures: An exploratory data analysis

Unsubmitted manuscript. Currently in correction by the co-authors.

Abstract

Dynamic signatures are a digitalized form of handwritten signatures. Their use has seen a steep increase for important transactions, such as life insurance and telecommunication contracts, sales and banking operations. A dynamic signature can be disputed and a forensic handwriting examiner may be hired to help determine whether it is genuine or not. Such a conflict may emerge years after the questioned signature has been affixed. Since reference material contemporary to the relevant period may not be made available to the expert, it can be questioned whether time influences dynamic signature data, which could affect the expert's results. This study was designed to explore this possible influence. Dynamic signatures of three participants were collected over a duration of 18 months, during 44 acquisition sessions. Based on this sample, the goals of describing variation of dynamic features over short and longer time periods, defining adequate sample collection strategies and sampling time frames, as well as laying down the foundation for using the time information for comparative analysis of dynamic signatures, were pursued. Both the relative stability and the slow 'drift' of signatures over time were illustrated by our results. The findings of this study lead to recommendations for sampling in casework, validate statements previously made by forensic scientist through an empirical investigation on dynamic signatures and strengthen the statistical basis for forensic signature comparison.

Keywords: Forensic Science, Questioned Document Examination (QDE), Forensic Handwriting Examination (FHE), Online Signature, Contemporaneity, Time, Template Ageing, Biometrics

13.1 Research Context

Forensic Handwriting Examiners (FHE) are often tasked with determining the authenticity of signatures. Their expertise is called upon whenever the source of a signature is being doubted and case circumstances are complex. They deal with difficult conditions when performing their signature comparisons. FHEs mostly work on cases that are contested years after the signature was made. Signatures are known to be affected by ageing [325], which may cause issues when comparing samples from different periods. This makes contemporaneity an important factor for valid comparisons. FHEs often struggle to obtain enough trustworthy reference materials (control data) from the presumed source from the time of the questioned signature. As a result, they need to invite the suspect to produce control materials. These newly produced references are often created years after the questioned signature. The changes in handwriting occurring both short term [563] and long term [325] may prove problematic in these situations. For dynamic signatures in particular, there is often little contemporary reference material available. Hence the examiner needs new materials produced by the suspect. Questions about the reliability of conclusions based on samples not contemporary to the case are thus of great interest in forensic science. In this study, the conclusions are based on the case assessment and interpretation model. Therefore the evidence is assessed under two competing propositions and its value expressed as a Bayes Factor (BF), a probabilistic statement of inferential strength.

The presence of changes in handwriting and signatures throughout a person's lifetime have been acknowledged by the scientific communities in neuroscience, forensics and biometrics [72, 165, 244, 392, 461, 616]. Ageing is known to produce a variety of effects directly or indirectly affecting our signatures. Some studies investigate the effects of ageing and writer age on signature dynamics in particular. Gomez-Barrero et al. suggest that when ageing, we tend to produce a higher number of shorter strokes, produce more trembling and move away from log normality [223]. Many other authors have noted a difference in velocity between age groups [123, 203, 233, 261, 375, 606]. While other studies focus on handwriting, interesting results relating to ageing have been reported. Guest suggests that especially velocity and acceleration show large changes when ageing [233]. Faundez-Zanuy et al. [167] corroborate the previous results based on writing velocity. The non-constancy of signature features is also known and relevant to biometrics, where it is referred to as 'template ageing'. Researchers have developed two types of strategies to deal with this effect. As a first approach, some authors propose to use 'time invariant' features that are less affected by ageing [233]. The second approach is called template updating and consists of changing the reference materials to adapt to the new conditions [288, 475, 574, 575].

A 4 stage model for the development and decline of handwriting, relying on the concepts of graphic maturity [396] and movement control [71], has been proposed [244]. These stages are chronologically the 'learning' phase, the 'impressionable' phase, the 'maturity' and finally the 'senility' phase. Ageing is non-linear and therefore there are age ranges and periods, throughout which the handwritten product evolves or declines quickly. Two age groups are susceptible to present greater changes in short time frames: the young children and the elderly [149]. Rosenblum et al. [490] investigated handwriting as a test for motor function decline, establishing linkage between aging, motor function and the handwritten product. Only in the maturity stage, signatures are supposed to remain relatively stable. Forensic science is, luckily, mostly concerned with signatures from adults. These individuals mostly fall within the maturity or senility phases. No exact indications exist on when these phases start. Caligiuri and Mohammed review a number of studies on the influence of aging, mentioning that influences on speed appear subtle between ages 20 and 70, but are more noticeable afterwards [72]. Other studies mention important changes around 65 years of age [73]. Other studies mention that individuals in the 18-50 year age band will show less effect than other age bands [325]. When further investigating the effect of aging and evolution (of the signature), some authors have found the extent of ageing effects to be highly user-dependent [203].

One of the key problems in forensic signature comparison is the limited amount of comparison materials, as well as the trustworthiness and the availability of reference signatures. Contemporaneous materials in particular are often very limited in forensic cases. As such, materials that are further spread out through time have to be used, or the examiner needs to obtain new samples from the presumed source. The previously described subsamples are meant to represent the data a forensic scientist may acquire as reference materials, specifically requested samples [187]. Requested samples are produced post-case, by inviting the presumed source to produce material during (most often) a single session. This limits the available materials to a very short timeframe. Reference sample collection should cover an adequate timeframe, contemporary to the case [157, 187, 244]. The period of interest in the case may additionally depend on the 'phase' the writer of interest is or was in at the time of the events. Population sampling in forensic science should also use age information, as additional variation but also other factors such as illness may appear over time. Lanitis et al. [325] suggest that a period of roughly 2 years is only enough time to study short term variation. According to their research long term variation occurs over a period of at least 5 years [322], which was also corroborated by Walton [606]. Additional studies suggest using a scheme based on age groups, and indirectly relating to the stability of the learned signature motion, improve the performance of their signature verification system [150].

In forensic science, researchers have made recommendations to choose a contemporary reference sample. These recommendations do mention sampling throughout time, but often do not specify what an adequate timeframe to sample from might be [187, 244]. Several studies have however observed similarities in measurements taken on signatures from the same batch and differences between batches, stressing adequate sampling throughout time [157, 563] . Mohammed [392] very recently suggested that signatures within one year to be ideal. However, no empirical data is presented as basis for this recommendation. Little empirical research on contemporaneity, the effect of the temporal distance and ageing on signatures and evidence evaluation in general is available. The impact of contemporaneity and time thus remain vague in most analyses and evaluation methods. This lack of empirical evidence and precision is the main reason for our interest in the influence of aging on signature data. In addition, dynamic data, which is highly movement related was previously inaccessible to the forensic examiner. Galbally et al. noticed that dynamic global features show greater variation in time than static ones [203]. This is hypothesized to stem from a focus on reproducing shapes rather than movement kinetics, such as velocity or pressure. It appears of interest for the forensic handwriting community to extend the knowledge about template age and contemporaneity.

Adequate reference materials are an essential basis for the successful examination of dynamic signatures. Contemporaneity in particular plays a role for dynamic signatures, where little to no reference materials can be obtained from the period of interest. The objective of this study is to determine if there are observable effects of short-term ageing in dynamic signatures and extrapolate on how to collect samples to mitigate these. In case of presence of such effects, it is of interest to determine how the reference data collection process can be adjusted to account for them. The signature data was acquired during an 18-month period. First, a visual and statistical exploration of a set of movement-related features - selected subjectively for the common presence in dynamic signature systems and comprehensive interpretation – is proposed. The impact of different sample collection strategies on the representativeness of the sample is studied. We then use this information to produce recommendations for reliable reference signature collection in forensic casework.

13.2 Materials and Methods

13.2.1 Data

All signatures were acquired in standardized conditions. Participants sat at a desk on an adjustable office chair. Writing position, seat height and tablet position was the participant's choice. They were also allowed to rotate the digitizing tablet on the table, as long as it remained flat on the table. A Wacom DTU 1141 signature tablet was used for the data acquisition. The tablet samples 200 points per second with a spatial resolution of 2540 lpi. Pen pressure is measured axially in 1024 levels, which do not correspond to physical units. Wacom specific drivers and SignatureScope software were used. The signature data is composed of four measurements and three input related columns. The untreated signature data were exported as CSV files. We then imported the signature data into the R Statistical Software package for analysis, comparison and evaluation. The extracted feature data is multivariate (multidimensional). Sixty different global features describing the original signature data were extracted. This data included measurements such as signature length, width and height, but also averages, variances and maxima of speed, pressure, acceleration and other features. In this study, we chose to represent only four of these features, namely the signature duration, the average pressure, the average velocity and the velocity variance. These features have a physical interpretation and are not just abstract mathematical constructs. They are both intelligible and relatively common. By extension, these could also be inferred for static signatures and are of interest. We had to limit the study to this number of features for brevity. Additional materials and figures, omitted for brevity, can be found online in the Supplementary Materials. Genuine signatures were acquired from three writers over multiple sessions, in a timespan of approximately 18 months. The signers started out signing once every day of the week, then every 2-3 days, then every week up to once every 3 weeks. Overall, every signer completed 44 sessions from February 2017 to December 2018. During the first sessions, we only sampled 10 signatures per session. We subsequently decided to raise this number to 20 signatures per session, due to the large intra-session variability observed. At the end of the sampling period, each of the three signers had produced approximately 800 genuine signatures. We started with frequent acquisition sessions, diminishing frequency of sessions as we went along. This was done to accustom people to signing on a digitizer, as the participants reported the surface as 'slippery' and difficult to sign on initially. During these first sessions, there was some evidence of people relearning and adapting the signature movements to the new environment. The first week, signatures movements were often slow and awkward, while signatures made during later sessions were faster and smoother.

13.2.2 Experimental Design

We explore the evolution of the signature's features over time in signatures from a stable age group (22-60 years). In the scope of our study, we define mid-term variation as variation over a period of at least 10 months, and the short-term variation as variation over up to three months. These artificial intervals do not reflect the literature, but are what our more limited datasets permits studying. To this effect, we decided to perform various visualizations, models and statistical analyses of the dynamic features of the signature.

13.2.3 Univariate analysis by boxplots

The signature data, grouped by acquisition session, were first visualized by using box-and-whiskers diagrams (boxplots). A boxplot shows the main data concentration (50% of the values) through the central box, as well as the central tendency (the median) through the bar inside the box, as well as extreme values in the whiskers, and possible outliers as filled black dots. The analysis of said boxplots does not only show the data distribution through a few summary values, but also allows to see the symmetry and skewness of the data. By comparing these boxplots, similarities and differences within and between sessions can be highlighted. We then looked for changes and effects occurring over time. Special attention was given to differences between initial and final sessions, as well as long-term changes. In order to emphasize an evolution in the dataset, we used locally estimated scatterplot smoothing (LOESS), a form of regression analysis, to determine a trend in the acquired data. This regression analysis shows a trend line, data points and the 95% confidence interval. The trendline schematizes the evolution over time more clearly and smoothly than the boxplots themselves. Both analyses were used jointly to describe the signature features and their behavior over time. A descriptive analysis of results will be provided for every signature.

13.2.4 Multivariate analysis by PCA

While very useful and illustrative analyses, the boxplots only show how time affects a specific variable. In order to understand better how signatures, defined by 60 different features, are affected by the passing of time, we propose to use a multivariate ordination technique. In this study, we used Principal Component Analysis (PCA) based on singular value decomposition (SVD) to study the projection into a new feature space and feature correlation. We proceeded by grouping the signatures into several 'clusters' based on the temporal distance to the first acquisition session and then performing a PCA, followed by the visualization of the PCA scatterplot. The PCA scatterplot of the two (uncorrelated) principal component (PC) axes maximizes the sample's represented variance in 2D space. This essentially means we are reducing the complexity of multiple correlated features by projecting a combination of them into two 'artificial' uncorrelated ones, the PCs. We drew Hotelling-t 95% confidence ellipses around the groups to show main data concentrations in the feature space. In a first trial, we used seven groups, of about 100 days (approximately 3 months) each, to investigate short term-variation in the data. Data clustered closely or overlapping indicates little difference due to timing, while well-separated groups would indicate large variation. In a follow-up trial, we defined only two clusters based on our data. These larger groups of about 300 days (approximately 10 months) were assumed to show long-term variation. A short description of the clustering and observed variation is proposed. In addition to the PCA itself, we look at feature correlation through the PCA loadings plots and scree plots. The loadings, i.e. the feature contributions to the principal components, may help highlight uncorrelated, complementary features. The PCA scree plot shows the convergence of the eigenvalue and thus represents the percentage of represented variation.

13.2.5 Representativeness of data sampling by univariate distributions

For every studied feature (i.e. average velocity, velocity variance, average pressure and signature duration), we first represented and analyzed the data distribution of the entire sample of a given signature, labelled as the population data of that signature. Distributions were visualized by using a non-parametric estimation technique known as kernel density estimation (kde; aka Parzen-Rosenblatt window). Kernel density estimations are based on summations of kernel functions (usually Gaussian), which represent a smoothened density. The kde analysis shows the cumulated, smooth density of the feature values. One must note that kernel densities, although non-parametric, depend on choosing a smoothing factor. This factor defines the smoothness of the final distribution. Here we used the default estimation implemented in the 'ggplot2' package [608], for the R statistical software [466]. The visual examination of the population informs us about data structure, breadth of variation and may reveal the presence of multiple modes (high concentrations around a value; multiple 'peak' values). The population distributions, representative of the 18-month period of the study, will be shown and described. Second, we have drawn samples of the population and compared the resulting distributions with the population distribution, to investigate whether the drawn samples can be considered as representative of the population. Samples were drawn according to the four following procedures:

- Subsampling 1: signatures of a single randomly selected session
- Subsampling 2: signatures of several randomly selected sessions, by cumulating the data

- Subsampling 3: signatures of several randomly selected sessions, by drawing randomly 15 signatures evenly spread among the sessions (in order to eliminate the effect of the dataset size)
- Subsampling 4: a number of signatures randomly selected from the entire population, without considering the acquisition sessions

The first three sampling procedures were followed to represent a general scenario than can be encountered by FHEs, in case the reference material is produced upon request, at a single date or at several dates [187]. The fourth sampling procedure refers to situations where the reference material available to FHEs is made of a series of course of business signatures collected on different dates.

Both the population and subsamples are represented by estimated densities. The comparison was based on qualitative comparisons. If the distributions appear similar in shape, mode and density, the subsample can be considered as representative of the population, and therefore adequate and sufficient in size for forensic usage.

13.3 Results & Discussion

13.3.1 Univariate analysis by boxplots

The results will be discussed signature by signature. Some of the graphs have been omitted from the main article and can be found in the Supplementary Materials. The boxplots of the data of signature 1 are illustrated in figure 13.1. It appeared that several changes occurred over the period of 18 months. Both within and between-session variability was illustrated. There are, spread throughout the sampling period, sessions with very low within session variability, but most sessions have large boxes and whiskers. The medians of the different sessions, as well as the boxes, which represent the main data concentration, are situated at different values between sessions. This would imply that there is variation or evolution in the signature. The question to pursue is whether or not this is short term variation. If it were, the result should be stable around a central value, otherwise the feature values change over time and evolve towards a different central value. The variability between sessions means that an examiner who selects and uses reference material coming from a single session is exposed to the risk of a sampling effect. A single session may or may not lie in the main data concentration of the population, and thus it may not well represent the writer's usual variation. If an outlier session is the only available data, the assessment of evidential value will be skewed by the fact that the data is not representative of the writer's actual 'average' variation.

The LOESS results of signature 1 are illustrated in figure 13.2. When looking at the average pressure results (fig. 13.2 a), a progressive increase can be observed in the initial sessions. These initial acquisition sessions were, however, much closer in time to each other than the later sessions. They highlight the changes and variation in tighter timeframes and the gradual changes over time. Average pressure seems to stabilize after about two months of signing on the tablet. This initial change is hypothesized to be an effect of the unusual surface dynamics and writing sensation on the signature pad. The signer must indeed get accustomed to the new surface, as well as relearn his signature movements on the slippery tablet surface. The average speed results (fig. 13.2 b) show no initial increase. The average values appear to gravitate around the same region, until toward the last ten months, when a decrease occurs. The speed variance results (fig. 13.2 c) show some increased values in the first four acquisition sessions but appears stable afterwards. As for the overall signing duration (fig. 13.2 d), the value seems to steadily decrease throughout the sampling period. When further investigating the stroke length (i.e. length sum of all the signature strokes) and the total width of the signature over time, a decrease in both features is observed.

This decrease can be interpreted as an improvement in the fine motor control on the tablet surface, due to training. Our hypothesis is that the signature movements changed over time to increasingly resemble the original signature on paper. As an alternative hypothesis, an adaptation by simplifying the signature movements may have occurred. A visual investigation of the signature shape would suggest that the first hypothesis is more likely, as no apparent simplifications and changes in signature shape have been observed. The decrease in signature duration may be attributed to much the same hypotheses. An additional cause may reside in the fact that the tablet provides slightly delayed visual feedback. At first, the participant was reticent to sign faster than the appearance of the digital ink. The participant was inclined to wait for the system to catch up to their signing. Two of the three participants showed this behavior initially. Learning and automation took over after a few acquisition sessions and accustomation to the system, also explaining part of the decrease in signature time.

As for signature 2, the behavior of the within and between-session information, visible on (Suppl. Mat. Fig I), appears to be much the same as for signature 1. Some changes in feature values have occurred during the sampling period, but signature 2 is more stable than the first one overall. In (Suppl. Mat. Fig II) we can see an increase in the average pen pressure in the initial acquisition sessions, similarly to signature 1. The average speed in (Suppl. Mat. Fig II) shows some oscillation, but no steady changes. Average speed decreases slightly after the first 4 months and then increases after 8 months. The speed variance shown in (Suppl. Mat. Fig II) somewhat follows the pattern of the speed itself. At first, the variance is quite stable but then increases toward the last 10 months. Most likely the increase in speed is due to a training effect or adaptation to the tablet surface. In (Suppl. Mat. Fig II), the total duration of the signature can be seen to oscillate slightly, but mostly stabilize around 800-900 ms. Surprisingly, the signature duration did not decrease while the average speed increases in the last months. When looking at the size of the signature and the stroke length, these values can be explained, however. In the last sessions, the overall stroke length is longer, due to an increase of both width and height. Longer strokes permit longer accelerations and higher velocities, as acceleration and deceleration phases are further separated. Again, no major changes were made to the actual signature shape indicating a different writing rhythm.

Signature 3 has, in comparison to signatures 1 and 2, relatively large within-session variation. Signature 3's boxes, in (Suppl. Mat. Fig III) are very constant in size, indicating a regularity absent in the other two other signatures. The signature 3 results generally show a higher spread (box size) than the other two signatures, but also smaller whiskers and more outliers. In signature 3, three initial sessions differ from the other sessions in all of the shown features, especially signature duration and velocity variance. Signer 3 also had the hardest time to adapt to the new surface and took more time than the other participants. Signature 3 also features some changes over time. The average pen pressure, shown in (Suppl. Mat. Fig IV), increases over time. Additionally, the distribution of average pressure for signature 3 is more dispersed than for the other two signatures. As for the average speed, shown in (Suppl. Mat. Fig IV), we can see a small decrease after the first two months, then after 6 months the values increase again, attaining the levels of the initial acquisition sessions. The speed variance (Suppl. Mat. Fig IV) follows that pattern closely, staying relatively stable over time. As for the signature duration, (Suppl. Mat. Fig IV) shows that after three initial long sessions, this feature decreased steeply in the fourth session. The participant reported having trouble with the visual feedback of the digital ink of the tablet. As there is a small delay between the acquisition of the point and the appearance of the digital ink, the participant felt he needed to pause for the system to catch up. After being accustomed to the tablet, the participant started signing more naturally, resulting in a steep decrease in signature time. During the rest of the acquisition period, the signature duration steadily decreased. We hypothesize that this is due to further training and learning with the signature tablet, but in this case, also to some slight simplifications in the signature movements. For signature 3, some ornamental loops and some segments were linked in later signatures, whereas early signatures resembled the paper version more strongly.

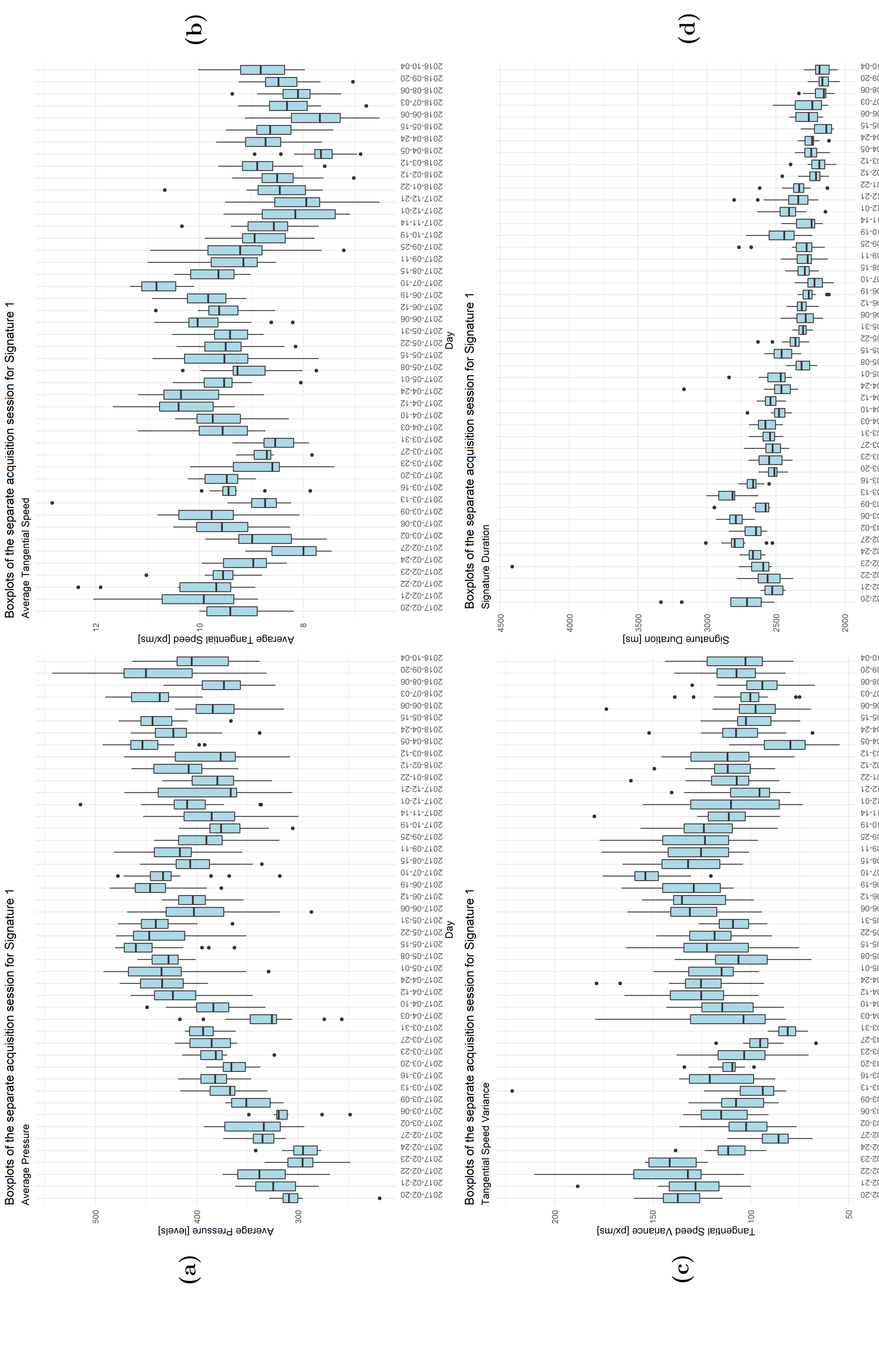


Figure 13.1: Boxplots of Signature 1, depicting the 18-month acquisition period. The graph shows average pressure (in levels, units defined by the system). On the x-axis, we have the date of the acquisition session in a YYYY-MM-DD format. (a) Avg Pressure (b) Avg Speed (c) Speed Variance (d) Signature duration

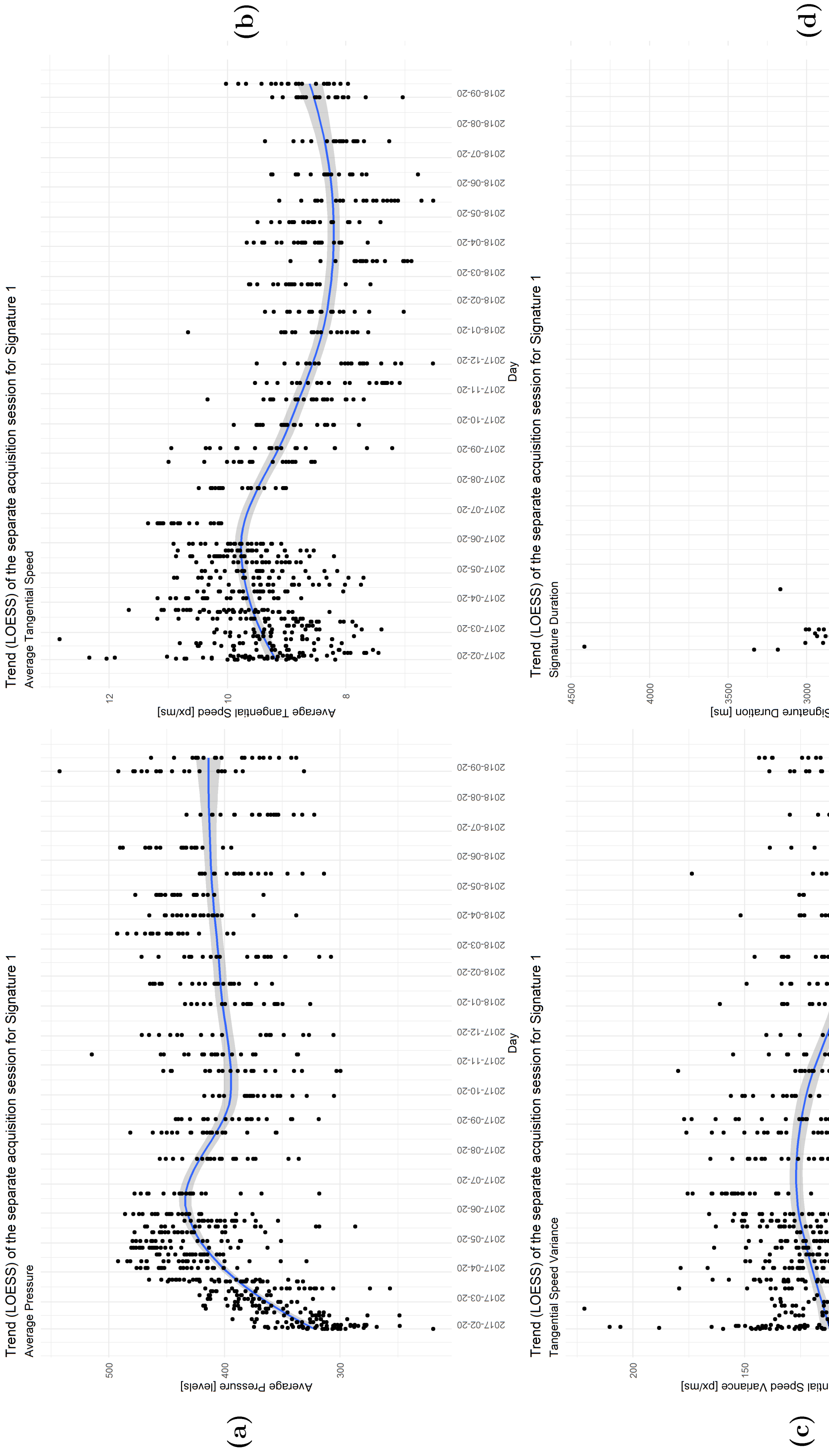


Figure 13.2: LOESS trend lines for the Signature 1's features. The blue line is the LOESS regression, the points are the available data and the grey interval is the 95% confidence interval. (a) Avg Pressure (b) Avg Speed (c) Speed Variance (d) Signature duration

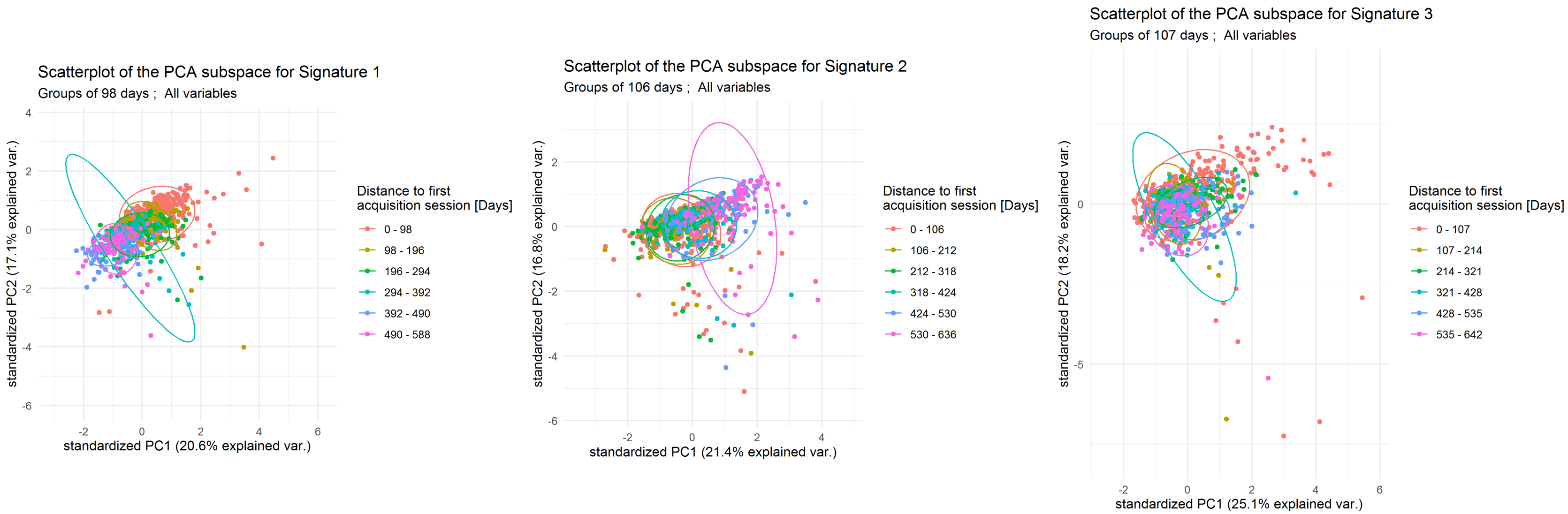
Overall, changes and variations were observed in the different dynamic features we have investigated. Not only do the central tendencies, such as the medians, vary and evolve throughout the period, but the data distribution also varies. As such, some sessions present very large variation, whereas other present very small variation. Even session with similar medians do not show the same variation, and conversely sessions with similar variation do not show the same median and data concentration (box part of the plots). The first sessions generally showed somewhat erratic and unpredictable behavior, which can be related to accustomation to the new writing conditions. Once this short period is over, feature changes are very personal as previously observed by Galbally et al. [203]. The only similar effect between participants was a decrease in signature duration, either due to the training on the new writing surface or to slight adaptation of the signature. In two of the three participants, there is a temporary decrease in average speed. The third participant did not exhibit this temporary decrease, showing instead a decrease in speed in the later sessions. Most of these changes are hypothesized to reflect the need for more movement control on the unusual, more slippery surface. The signer needs to perform a slight 'adaptation' of the original signature movements in some cases in order to preserve movement fluidity. We suppose that some parts of the movement need to be relearned.

As practical implications for FHE casework, the effects of accustomation to the unusual writing surface may be important in the advent of dynamic signatures, as people are not yet used to signing on these devices. The first signatures produced may well be 'outliers' for most people and present differing features, until some training is done. A possible recommendation may be to make clients at a Point of Sale read and write on the devices to make sure he is accustomed to its use before signing. On another note, due to within and between-session variation, FHEs may not solely rely on reference data from a single point in time, as most authors already pointed out [187, 244, 392, 396, 563]. In signature comparison, reference data representing a minuscule time period such as a day may lead to significant over or underestimation of the value of the evidence.

13.3.2 Multivariate analysis by PCA

The projection of the data when using all of the available variables shows no clear separation between groups. Of course, even the PCA is a dimensionality reduction technique and the visualization is only 2D. There are differences in the dispersion of the points in the PCA subspace, especially for participants 1 and 2. In the PCA scatterplot for signature 1, visible on figure 13.3 (a), we can see that the groups overlap, but that the points are gradually shifted laterally with template age. A similar trend for signature 2 can be seen in figure 13.3 (b). For signature 3 (figure 13.3 (c)), we observed a higher amount of clustering around the same area, no matter when signatures were acquired. As a possible explanation, we considered that participants 1 and 2 were younger (late twenties) than participant 3 (late fifties). If we consider the general handwriting maturity model proposed in [244], participants 1 and 2 may still vary, not having yet attained full maturity. The weak point in the theory would be that participant 3 is nearing the limit of the 'stable' age range and should show increased changes. However, since there is evidence that ageing is extremely related to the individuals [203], our observations may simply be the result of individual differences. As our sample includes only three writers, these hypotheses are left open for more extensive future studies.

When reducing the number of groups to only 2 in the PCA visualization, to observe more long-term changes and clear-cut groupings, we can see the separation is still partial. In figures 13.4 (a) and (b), we can observe the PCA projection for signatures 1 and 2. We can see a separation between the leftmost and rightmost group, with some overlapping occurring. The overlapping is lower for signature 1 than for signature 2, implying there are changes that are more marked over time for signature 1. For signature 3 (figure 13.4 (c)), the two groups are mostly overlapping and show limited differences. These observations



(a) (b) (c)

Figure 13.3: Scatterplot of the PCA subspaces. The principal components 1 and 2 maximize variation across the feature space. Data have been divided into 6 equally spaced time groups, as shown by color. Hotelling-t 95% Confidence ellipses have been drawn around the group for illustrating the main data concentration.

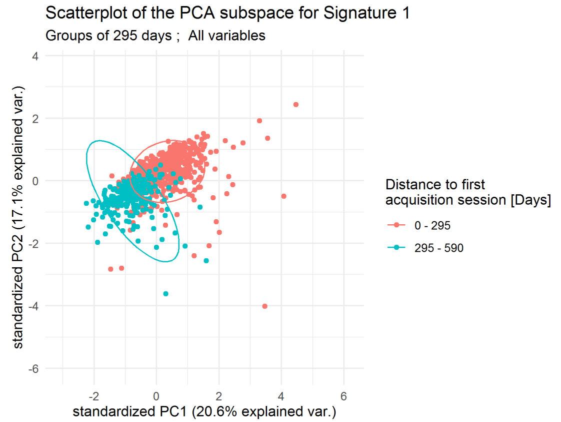
suggest that the variation of the signature over time is very personal and strongly depends on the user signing on the surface. We hypothesized beforehand that the age of the participant may play a role in long-term evolution and variation of the signature. The PCA results again suggest that participant 3 shows less (or slower) changes in his signatures. Further study should look at the influence of age, signing habits and frequency, as well as personal impact on the variation throughout a time period.

The analysis of the Scree plots on the left-hand column in figure 13.5 (a-c) supports a strong correlation between signature features. As a matter of fact, two principal components explain about 35% of the total variance in the dataset. As can be seen in the Scree plots, all of the PCs contain a large amount of variation, which means that we cannot easily eliminate variables to reduce dimensionality of the data. The high correlation also implies that the PCA projection into 2D space does not represent the dataset very well, because the percentage of explained variance is low. The loadings values in figure 13.5 (d-f), also point out the strong correlation between features. We can see that different features contribute the most variation for all of the three signatures. Most similarities are observed between signatures 1 and 2, where the signature length features contribute a lot of variation. Correlation between features is expressed by the angle between the arrows. At 90°, the features are uncorrelated, at 180° they are negatively correlated and at 0° they are correlated. Our data set shows that the features exhibit grouping and sharp angles between them, which is evidence for high feature correlation. This correlation should be taken into account for evidence evaluation, as it may cause overestimation of evidence if assuming independence between features.

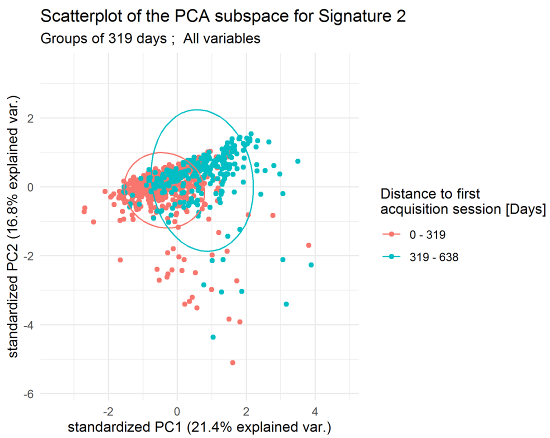
For our short-term analysis, groups of 3 months showed large overlap, while some separation occurred with 12-month groupings. Lanitis et al. [322] claim that long-term changes occur slowly over periods of at least 5 years, which implies that our sample is too small to detect these changes. According to these authors, our 18-month sample is only large enough to study short term variation [325]. Our results appear to show some separation nevertheless. One hypothesis is that there is a faster evolution due to the adaptation to the tablet, which would be analogous to learning a signature in the first place. Another is simply that long-term variation depends on the age of the participant and his graphical maturity. Finally, we have observed variation in time and noted its importance in forensic signature comparisons. Predicting these long-term changes or their pattern may be a potential research topic for future studies, improving both our knowledge on time invariant features and robustness to template ageing effects. This theoretical ageing model may help in forensic cases with aged documents. All of these observations show the need for individualized feature selection, as well as multivariate combination of feature information. Information for the different signatures is often contained in features specific to one signature. Statistical methods and adequate models need to be developed to deal with the intricate links between the data, which is why multivariate statistics are useful in this context.

13.3.3 Representativeness of data sampling by univariate distributions

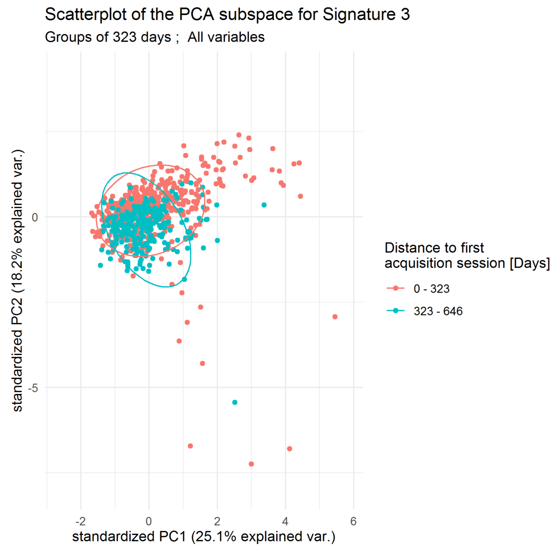
The population densities for signature 1 (Suppl. Mat. – Fig V), signature 2 (Suppl. Mat. – Fig VI), and signature 3 (Suppl. Mat. – Fig VII), all present relatively 'normal' shapes, although most are slightly skewed. We can generally see a clear mode (most frequent value, highest density, 'peak') for all distributions, except for the average pressure for signature 3 (Suppl. Mat. – Fig VII), which is bimodal. This bimodality may be due to experimentation of the corresponding participant to adapt his signature to the unusual surface. We have previously described that this participant changed some details in his signature movements during the data acquisition. Bimodality may cause problems when comparing the signature features with insufficient control samples. In addition, modeling this kind of feature distribution correctly would require complex models, such as Gaussian Mixture Models (GMM). The revealed information may guide and validate an ulterior probabilistic model choice. This strengthens



(a)



(b)



(c)

Figure 13.4: Scatterplot of the PCA subspace. The principal components 1 and 2 maximize variation across the feature space. Data have been divided into 2 equally spaced time groups, as shown by color. Hotelling-t 95% confidence ellipses have been drawn around the group for illustrating the main data concentration. (a) Signature 1 (b) Signature 2 (c) Signature 3

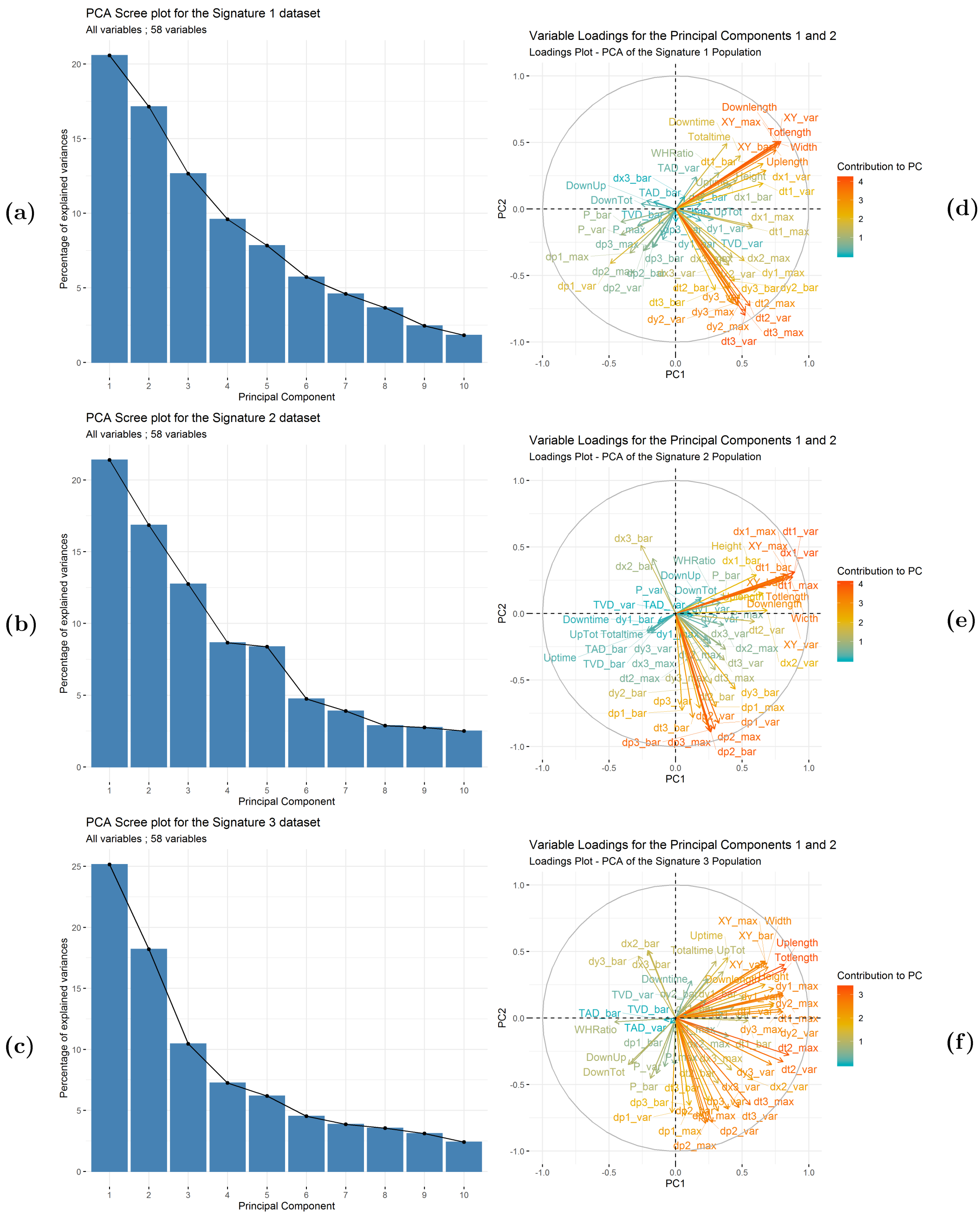


Figure 13.5: PCA Analytics for the SVD PCAs. The lefthand column features the Screeplots, showing explained variance for each additional component (PC). The righthand column shows the loadings plots, which show the strongest contributions to the PCs, but also feature correlation. (a)-(c) Scree plots for signatures 1-3, (d)-(f), Variable loadings for signatures 1-3

the foundations of dynamic signature examination through the study of previously unknown variation over time. It also provides an incentive to produce more research into probability models that can be used to provide evaluative opinions on signature evidence and support examiners.

13.3.3.1 Subsampling 1

When looking at the sessions for signature 1 (fig. 13.6), signature 2 (Suppl. Mat. – Fig VIII) and signature 3 (Suppl. Mat. – Fig IX), we can see that the densities of the population and the individual sessions are very different in both shape, mode and density. The individual sessions have irregular shapes, some sessions are quite concentrated around a few values and show high densities, but several peaks, while others are diffuse. All of the data is mostly clustered around a central value per session, which does not necessarily coincide with the population value. Overall, the population distribution is also lower in terms of density, which is caused by the prevalence of many more and different feature values. However, some similarities between individual sessions and the population are present. Signature 1 shows remarkable reproducibility in the signature duration, even between sessions (fig. 13.6). Signature 2 shows the same but for the average velocity and velocity variance (Suppl. Mat. – Fig VIII). Signature 3 shows rather high reproducibility in the average pen pressure feature (Suppl. Mat. – Fig IX). These observations support that time-invariant features [207] exist, but apparently they are intimately linked to the participant and not generalizable.

When executing comparisons based on single acquisition sessions, many undesirable effects may occur. First, due to the high data concentration, density may be very high. Second, the session inter-variability is high, so choosing a single session at random may misrepresent the population distribution. Major differences include their modes, their variance or even the distribution shape. The problem appears when thinking about the usage the forensic scientist will make of the data, notably providing support for two competing propositions. In most cases this means providing support for the signature having been made by a person A or by someone else. Conclusions based on a single session may be significantly different from conclusions based on the population. An individual session has a high chance to fall within the population mode, density, variance and shape and therefore be representative. There are however also sessions that are further away and can be considered as ‘outliers’. When using non-representative session, the value of the evidence will not be reproducible for a time period. It appears crucial to use data that is more akin to using the ‘population’, to better reflect the signature throughout a period of time, rather than reflecting the signature in a specific point in time. Sampling data from different sessions appears crucial for reproducibility, and thus validity.

13.3.3.2 Subsampling 2

For this section, we chose to only illustrate values for signature 1. In fig. 13.7, we can see that the more acquisition sessions are used to generate the densities, the closer the density’s shape and mode approximate the overall population. With as few as two sessions, the densities start to resemble the population density in both height, shape and mode. Starting from 5 sessions, almost no notable difference is observed in the density. The results of our experiment suggest that a density of two to five distinct acquisition sessions approaches the population’s density. While five sessions are better, two to three sessions already significantly reduces the problems faced with individual sessions.

These results further illustrate the need for an adequate sampling throughout a relevant timeframe. Natural short term variation is more strongly present between acquisition sessions than within acquisition sessions. In order to compensate for this variation and to produce reliable inference, data from multiple

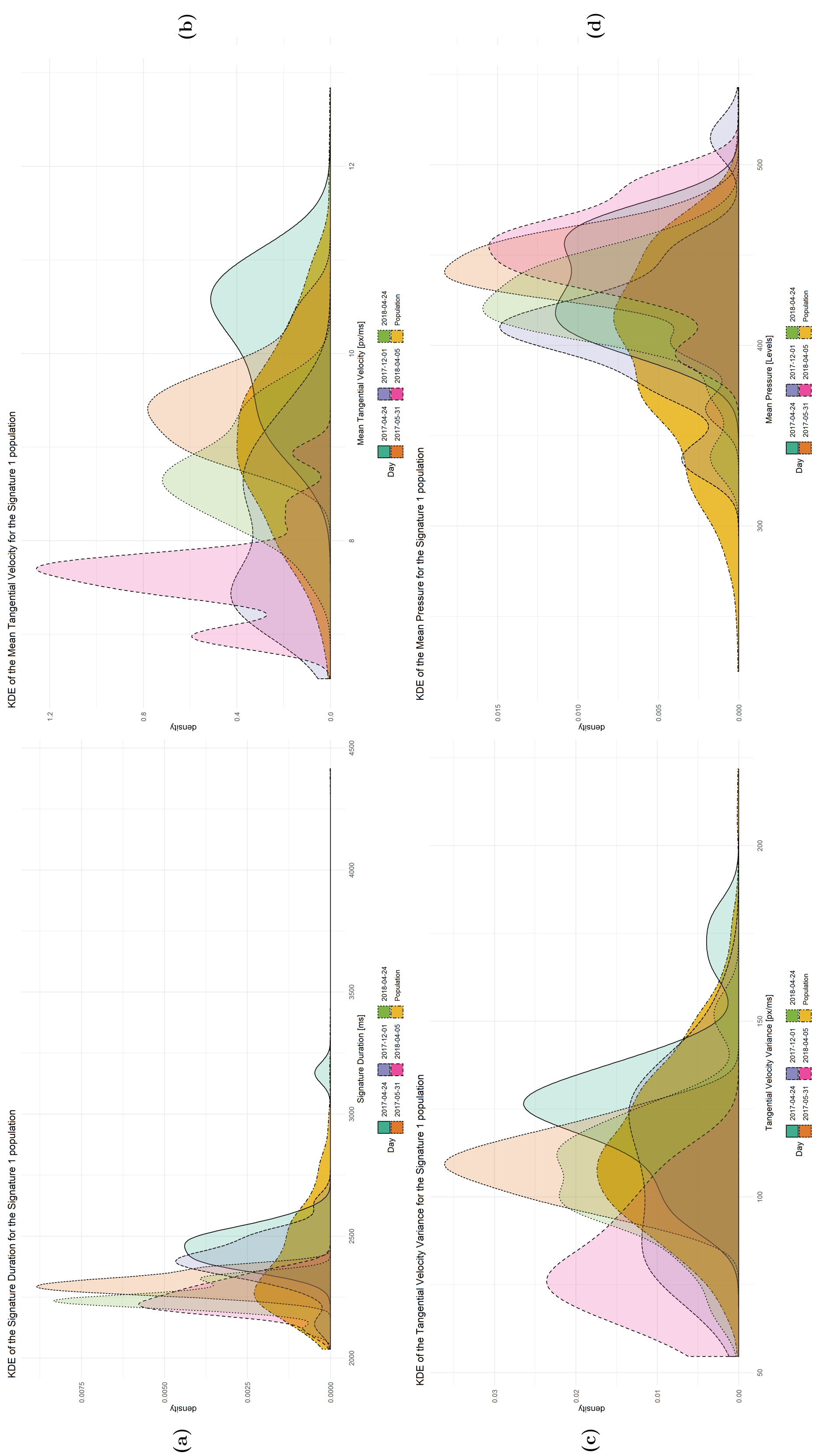


Figure 13.6: Densities of the individual session subsamples for signature 1. Five sessions were randomly selected and plotted against the population density.

sessions should be used. Our experiment shows that a sufficient quantity of reference materials from a certain timeframe is necessary to capture overall variability, corroborating results from previous studies [157, 563]. Our study would suggest that at least two distinct sessions are sufficient to approximate the population.

13.3.3.3 Subsampling 3

The third sampling method is similar to the second one, only eliminating the growing sample size. We limited the number of signature total to 15 (drawn from the number of sessions, which was varied). The results for signature 2 can be seen in fig. 13.8. The results are very similar to the previous results. Two to four sessions appear sufficient to account for the influence of time in sampling. Reproducibility however depends on the size of the dataset. With only 15 signatures, the position of the mode is more variable and a bit further from the population mode because of the relatively small sample size. Furthermore, the densities are often too high and overestimating the frequency of those particular values. In some additional trials not represented in this article, we increased the number of signatures to twenty, and the reproducibility increased significantly. An examiner should have a final sample size of at least 20 signatures over all of the sessions. Given our results, we recommend to organize at least 2 sessions on different days and collecting 20 signatures per session as a good rule of thumb to capture variation.

13.3.3.4 Subsampling 4

Our final method consisted of randomly drawing signatures from the entire population. This method simulates the presence of collected samples [187], anterior to the case. The visualizations presented in this research show a random sample from the signature 3 population. In fig. 13.9, we showed the density of a sample of 3 to 10 signatures. Our results show the rather obvious fact that the more signatures are available, the more representative the sample. When looking at the number of signatures, the density start looking similar from 8 to 10 signatures. Based on these results, we recommend using at least 8 to 10 signatures in casework with collected samples (given that no two were produced on the same day).

In this section, we briefly examined different methods of sampling that simulate reference collection in forensic science. Our main aim was to study the importance of time in the sampling process. A previous study [563] has shown between session variation to be greater than within session variation, creating potential biases in the interpretation of research results and casework. Our question was naturally to determine what sample would sufficiently represent the selected time period. We proceeded by visually and qualitatively comparing kernel density estimations, showing the distribution and variation of the features. First, our qualitative analysis of the feature distributions of both the population data and the single sessions corroborate the previously cited authors' results [563]. In additional trials, we tried to determine the importance of the covered time period and the sample size, while reflecting the usual methods of acquiring reference materials in forensic science. We determined that both sample size and contemporaneity have an impact on sample representativeness. We concluded that the time parameter seems less important than the sample size itself. For collected reference materials, randomly sampled from a large time period, as few as 10 signatures may be sufficient for statistical analysis. Requested materials, produced post-case come from specific timeframes and may be more susceptible to the influence of time. We determined that a minimum number of two different sessions, with preferably 20 signatures per session, would be necessary to approximate the relevant population's distribution. Further, we can conclude that signatures from an 18-month period are representative enough to use in comparison procedures. These results are coherent with recommendations for collecting reference materials in forensic casework [187].

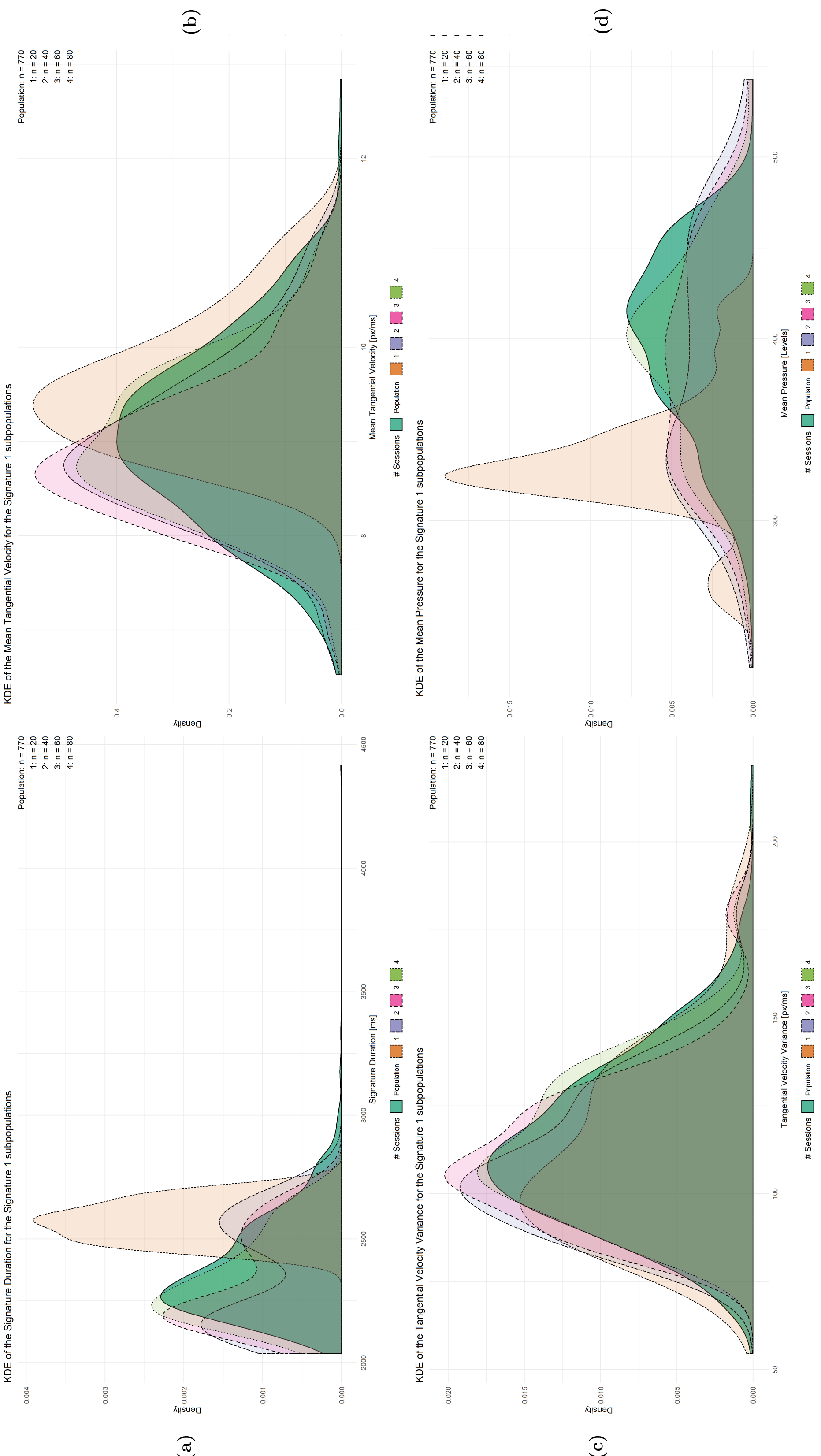


Figure 13.7: Densities of subsamples generated by cumulating 1 to 4 acquisition sessions for signature 1. Densities are kde of randomly drawn sessions. We incrementally increased the number of sessions included in the densities. The complete population sample is plotted as a reference point.

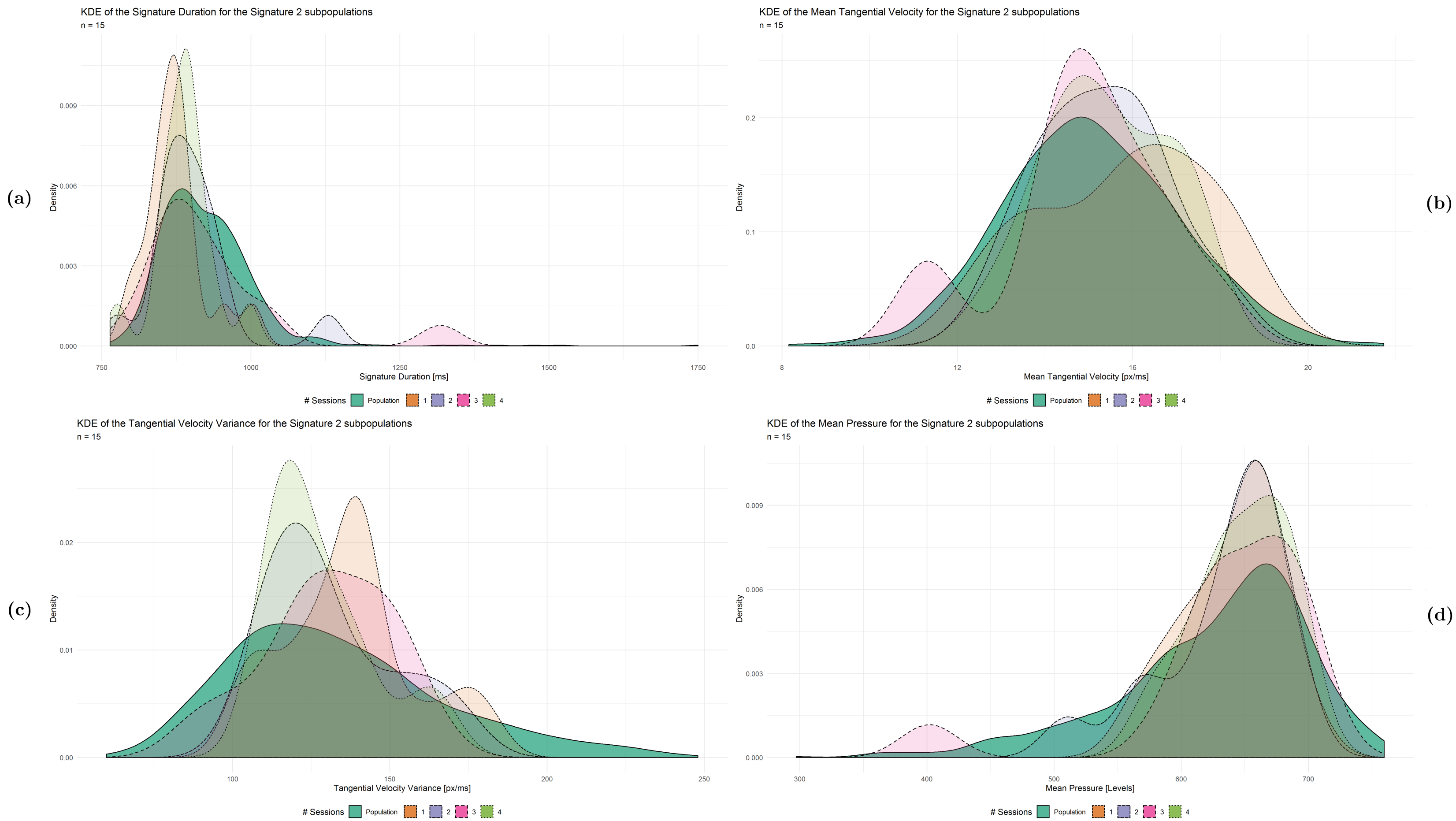


Figure 13.8: Densities of subsamples generated by drawing 15 signatures from 1 to 4 acquisition sessions for signature 2. Densities are kde of randomly drawn sessions. We incrementally increased the number of sessions included in the densities. The complete population sample is plotted as a reference point.

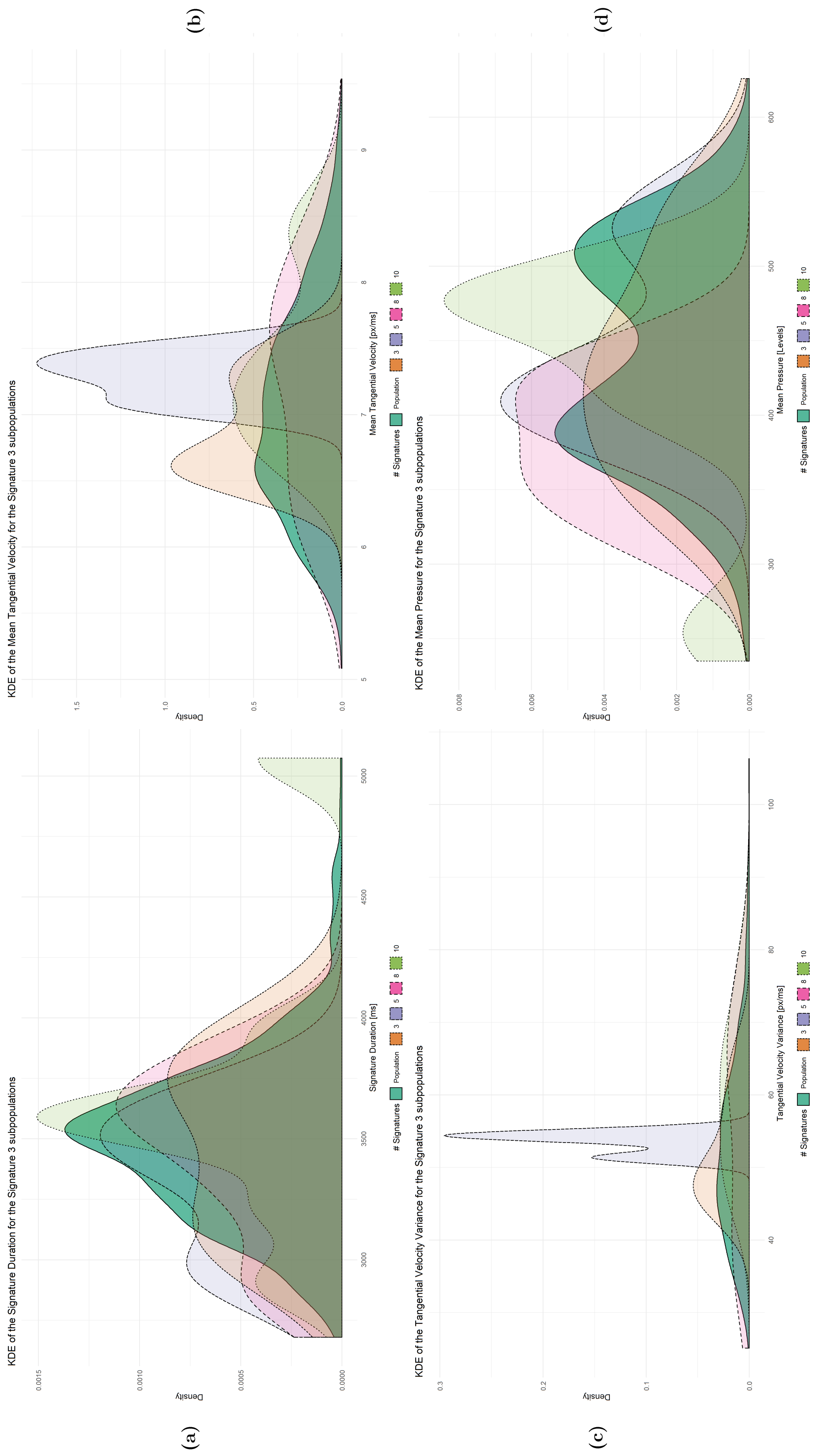


Figure 13.9: Randomly sampled signature distributions. Low number of sample signatures.

13.4 Conclusion

This study has allowed us to investigate the impact of time on features of dynamic signatures. We specifically studied signatures from three participants, who signed regularly over a period of 18 months on a dynamic signature tablet. We started with a descriptive analysis of the data and its inherent variation. We have found that dynamic features present very diverse variation during this period, both in terms of short- and long-term variation of the signatures. The important difference both within and between data acquisition sessions was noted. Variation in time appears to be a very personal, participant-related effect. Some features measured do seem to vary not only as a form of ageing and variation, but also as an adaptation to the new writing conditions. The slippery surface and unusual visual feedback may lead to changes in writing behavior, which normalizes over time with training.

Variation appears quite limited during the studied timeframe. As expected, we were not able to observe 'true' long-term variation during an 18-month period. We, however, did observe the effects of template ageing and a separation when visualizing the multivariate data. We showed that short-term variation is very minute, while the variation over years is noticeable, although personal differences exist. Based on our results, we concluded that signatures up to 18 months prior or posterior to the case are contemporaneous and adequate for comparison.

Our investigation into reference sample size and representativeness helped us point out the potential bias caused by using inadequate samples. This may be important for forensic examiners that have access to only limited materials and need to acquire additional samples from the putative source. An adequate sampling throughout time, such as collected materials, is preferable to requested materials. Additionally, fewer signatures are necessary when sampling freely throughout time. This process is more reproducible and reliable than sampling from specific time periods, such as specific days. When requested materials need to be produced post-case, we showed that at least two sessions on different days should be organized to avoid using an outlier session. Our study has shown that both sample size and contemporaneity have their importance in sample collection and should be considered carefully in forensic casework. Failure to do so may result in difficulties and bias in case assessment and interpretation methodologies. When assigning the value of the evidence, especially using statistical procedures, the data adequacy is vital to produce valid and coherent results. The impact of both sample collection and time in an evidence evaluation setting may be an interesting subject for further studies.

Although all of our conclusions are limited to a 'stable' and graphically mature population, we feel that our results highlight the importance of studying and considering time in signature examination. Template ageing, contemporaneity and quantity of reference materials deserve more attention than is actually accorded to them in forensic science. Further long-term studies on larger populations of signers should be conducted in order to investigate the reasons for variation and certain aging effects. This idea may be extensible to other age groups, as well as being applicable when the writer enters another phase in graphic maturity or motor decline because of aging. The field of research into time and temporal effects on samples, populations, age groups and motor control are open fields of study.

13.5 Acknowledgments

We would like to thank all volunteers that either gave their signatures or took time to produce simulated signatures. We especially thank the three volunteers that participated in the 18-month study. We would also like to thank Wacom GmbH Europe, who has provided us with both the hard- and software required for our study.

14 The impact of time on probabilistic case assessment and interpretation of dynamic signatures

Unsubmitted manuscript. Currently in correction by the co-authors.

Abstract

Dynamic signatures are handwritten signatures recorded on electronic devices. These dynamic signatures are easy to set up and cheap, which makes them an interesting alternative to paper-based signatures for businesses. An increasing number of important transactions, such as medical consent forms, insurance contracts or even sales, are nowadays being signed with dynamic signatures. Naturally, these cases may lead to legal conflict and court procedures. Parties often contest being the source of the signature on these important documents. In such instances, the court calls upon the expertise of forensic scientists to provide evidence as to the signature's source. Forensic scientists perform comparative analysis between the questioned and reference signatures. In many cases, reference material from within the same week or even month of the case is not available. The examiner must then make do with the available material, while still assuring a valid and justifiable assessment.

In this study, we investigate the influence of time on dynamic signature data. We look at the effects of time on evidential values when assessing authenticity. We use a specifically collected data set of genuine signatures, collected over a period of 18 months, as well as a specific simulated signature dataset. We create mock cases with this known source data to test the performance of a probabilistic model when the available reference data is temporally distant from the case. This probabilistic model provides Bayes Factors (BF) as a quantitative assessment of the evidence's value. Our aim was to study the impact of time on the evaluation of signatures with this probabilistic methodology. We were able to show that both contemporaneity and the size of the reference set clearly have their importance in forensic casework. Furthermore, we found that the methodology used in this study is more sensitive to the quantity of available reference signatures than to temporal distance. Time and temporal distance have, however, an important role in reference signature selection. The evaluation of signatures with reference material sampled from multiple points in time has greater reproducibility and validity than evaluation with punctually sampled sets. Overall, we found that the probabilistic model was quite robust to the variations in time and temporal distance. Temporal distances as long as 18 months are not problematic in

forensic casework, for writers who are in a ‘stable’ neuro-motor age range. These signatures can be considered adequate and contemporaneous for these writers.
Keywords: Forensic Science, Time, Template Ageing, Stability, Validation, Sample collection, Forensic statistics, Case assessment and interpretation

14.1 Context

Signatures are not constant in time, like DNA for example. The human motor system in general and signatures in particular are known to be affected by aging [72, 165, 244, 325, 392, 461, 616]. Therefore, they do not respect the ‘permanence’ principle, one of the five conditions of ‘ideal’ biometrics, as defined by Jain et al. [290]. In forensic science, the ageing and change inherent to signatures is especially important. Forensic scientists often have to deal with aged questioned signatures. Cases sometimes involve historical documents, are only contested years after the actual contract conclusion, or even because of long legal procedures and delays until an expert is hired. In these cases, obtaining contemporary reference signatures from the presumed source may be difficult. The forensic examiner has to manage with the available, often temporally distant reference material. Time-related changes are known to have caused misinterpretations in research [157, 563]. This also raises the questions of the reliability and validity of conclusions based on non-contemporary reference material.

One cannot discuss the effects of the passing of time without also mentioning ageing. The writers age and evolve, which affects their neuro-motor system, their muscles, their vision and other factors. All of these changes have an impact on signing behavior. In addition to time-related changes, signatures are also subject to ‘natural’ variation [244, 396], which is inherent to the execution of the signature. Both of these causes lead to large variation over time and require careful collection of reference material for comparison. Ageing is a non-linear process, meaning that there are age bands during which the handwritten product changes more rapidly and significantly. These changes have been categorized as short-term [563] and long-term [325, 606] . Lanitis et al. [325] suggest that a short-term variation occurs over periods less than 5 years, while long-term variation is noticed only over longer intervals [322]. The ageing process has been classified into 4 stages [244], based on graphic maturity [396] and movement control [72, 223, 490]. Two age groups show great variability in short time frames: the young children and the elderly [149] . The adult population shows little (or simply slower) ‘ageing’ effects. Caligiuri and Mohammed mention the period between 20 and 70 years as stable [72], while other studies mention important changes after 65 years [73] or even after 50 years [325]. Further, changes due to time are slow and progressive for signatures. Furthermore, the cardinal changes affect the signature dynamics, such as velocity and acceleration [123, 167, 168, 203, 233, 261, 375, 606]. However, some authors have reported that the changes caused by ageing as well as their extent are first and foremost user-dependent [203].

The large variation in signatures presents a great challenge when assessing evidence for signature authenticity. In a first approach, forensic scientists have proposed to mitigate the effects of variation by choosing an adequate comparative sample. These proposals mention sampling throughout time, but are vague about contemporaneity and the adequate timeframe [187, 244, 392]. Authors in the field of biometrics have studied ageing on a few biometric identifiers [135, 149, 165, 206, 220, 325, 474]. They have come up with a similar strategy to counteract what is commonly referred to as ’template ageing’. The authors propose to perform ‘template updating’ by replacing aged reference material with a more contemporary one [288, 475, 574, 575]. This inherently assumes that the examiner is able to obtain new signatures, that they are ‘trustworthy’ and that the period of interest is the present. In forensic casework, the period of interest lies in the past. A more relevant strategy lies in using 'time invariant' features, which are less affected by ageing [233]. These features may, however, be less discriminating than other less constant features. Overall, the examiner has to handle a trade-off between performance and applicability.

In the field of biometrics, ageing effects have often been studied through verification-task performance in automated systems. Verification performance was reported to increase when age is added as a soft biometric [104] , corroborating the utility of the age ranges. Another research team has found that in their 15-month study, the individuals themselves have a strong effect and that ageing is a very personal process [203]. Ageing does however not affect all feature types to the same extent. Two research teams have both found that local features are less affected than global features [203, 574, 575]. Aged templates produced not only worse, but also more scattered match scores. Given these findings, we expect that in forensic science, we should be subject to the same effects and scattering in performance.

Voices in the forensic and legal communities have been calling for a new paradigm for the assessment of forensic evidence. Probabilistic frameworks and the systematic use of empirical data [25, 145, 410, 430, 444, 489, 614] are currently changing forensic science. These frameworks have also become relevant to signature examination [108, 194, 201, 392, 430]. Forensic scientists have started to embrace the concept of Likelihood Ratio (LR) and the Bayes Factor (BF) as the expression of evidential strength.

The objective of this study is to determine the effects of ageing in forensic signature comparison. Dynamic signatures are a rich source of information, but were previously unavailable for study by forensic scientists. Furthermore, in forensic signature evaluation little work on probabilistic methodologies exists. We propose to model the variation through a probabilistic model and analyze accuracy and evidential value. Therefore, we explore the effects of time and ageing using mock cases and computer simulations. We study the impact of temporal distance between the reference sample and the questioned signature in both genuine and simulated signature situations. We also test and validate a probabilistic dynamic signature evaluation methodology, which offers evidential values based on empirical data under the form of Bayes Factors. Finally, we demonstrate the validity of forensic signature comparison using 18-month old templates within a probabilistic framework.

In section 14.2.1, we describe the collected dataset and the sampling through time in order to cover the 18-month period. We also briefly explain the probabilistic model and parameters, as well as the concept of Bayes Factor. In section 14.2.2 we give a detailed explanation on the mock case generation and the sampling of the reference signatures used to create specific conditions. We expose the four different sets of experiments we undertook. In section 14.3, we describe our experimental results and illustrate the model's performance in the varied conditions. Finally, in section 14.4 we provide generalizations and a conclusion on the impact of ageing on the forensic comparison of signatures.

14.2 Materials and Methods

14.2.1 Data & Model

Before defining the methods used in this study, we would like to define both temporal distance and ageing. Temporal distance is used to signify the relative duration between the production dates of two signatures. It is, more simply said, the duration between the productions of two signatures. In this study, we measured this distance in days. Ageing is the process of changing over time. It concerns a specific, numerically identical, entity [318]. In our case this entity is the writer, who is changing his neurological system, his movement patterns, his effector muscles, due to time. The concept of ageing itself is only meaningful for signatures made by the same writer. Genuine signatures show these effects over both short and long periods of time.

There are few extensive datasets, effectively sampled through time and fitting the requirements for a study of the influence of time and the effects of ageing. We acquired our own, specific dataset in order to ensure having adequate and sufficient data. In order to provide good conditions for reproducibility over the 18-month period, we standardized the signature acquisition conditions. These 'physical' conditions were the same for both signers and forgers. A Wacom DTU 1141 signature tablet was used for the data acquisition. The tablet samples 200 points per second with a spatial resolution of 2540 lpi. Pen pressure is measured axially in 1024 levels, which do not correspond to physical units. Wacom specific drivers and their 'Signature Scope' software were used for recording the data. Participants sat down at a desk on an adjustable office chair. They were asked to adjust the chair for their comfort. They were also allowed to rotate the digitizing tablet on the desk, for writing comfort. However, digitizer inclination and height could not be changed during the acquisition. The stylus used is the standard Wacom pen, which has a hard-plastic nib. It was also tethered to the tablet, creating some slight initial discomfort for some participants.

The genuine signature acquisition was conducted over a period of approximately 18 months. A total of 44 distinct acquisition sessions per participant were organized. Three participants volunteered for this longitudinal study. The data acquisition also had several phases and rhythm changes, with some adaptation along the way. In a first phase, we started with frequent acquisition sessions; at first once per day, then once every 2-3 days. The participants reported the surface as 'slippery' and difficult to sign on initially. They needed frequent sessions to train and get used to the experimental setup, especially the digitizing device. In the first sessions, we sampled 10 signatures per session. We subsequently decided to raise this number to 20 signatures per session, due to the large variability observed in the first sessions. In a second phase, we organized sessions with diminishing frequency of sessions (once per week, once every other week), until we stabilized around one acquisition session every three weeks. Signature acquisitions lasted from February 2017 to December 2018. Each of the three signers produced about 800 signatures during this period.

The simulated signatures were made by volunteer forgers in a forgery contest. Forgers had time to practice on both paper and the digitizer before forging. They were required to produce 10 simulated signatures per chosen signature 'target', but could also reject and restart forgeries. The contest was organized to provide a financial incentive to produce good forgeries. 57 people participated and simulated from one to three of the signatures. All forgers were University staff or forensic science students. The signature models provided to the forgers were five paper-based signatures produced at the beginning of the study. Simulators were free to change the writing conditions for comfort, like the genuine signers. No specific instructions on how to forge the signatures were given. Participants were free to do freehand simulations or trace the signature. They were, however, aware that dynamic data would be used to examine the forgeries. The raw signature data was exported as CSV files and imported into the R Statistical Software package for analysis, comparison and evaluation. The signature data is composed of four measurements and three pen-input and button-related columns. The data is multivariate (multidimensional) and formed by a chronological series of points, describing the signature and its execution movement. Sixty different global features describing the original signature data were extracted. These features include signature length, width and height, but also averages, variances and maxima of speed, pressure, acceleration. For sake of stability and simplicity of the probabilistic model, a feature selection step was carried out before the trials. During feature selection, we tested for discriminative sets of three features. Feature selection is very important, as they define the extent of variation and discriminative power of the method, while reducing the computations and complexity. It is a complex endeavor, which surpasses the scope of this article. Our requirements included low misleading evidence rates, as well as high reproducibility. In

Feature set	Applied to	Feature 1	Feature 2	Feature 3
1	Signature 1	Signature Duration	Maximum of the third temporal differential of pen pressure	Maximum distance to centroid
2	Signature 2	Signature Duration	Variance of the horizontal acceleration	Variance of the tangential jerk
3	Signature 2	Length of in-air pen movement	Pen on surface duration	Average pen pressure
4	Signature 3	In-air movement duration	Variance of tangential jerk	Maximum distance to centroid

Table 14.1: Feature sets used in this study

some cases, we used alternative feature sets to check if the effects were generalizable. These features are summarized in table 14.1.

The Bayes Factors (BFs) were evaluated through a probabilistic model, proposed by Linden et al. [338]. This probabilistic model is a conjugate Bayesian model, with a Normal-Wishart prior distribution and a normal likelihood, as specified in equations 14.1, 14.2, and 14.3. In this paper, two propositions, H1 and H2 have been used. Under H1, we assume the questioned signature y and the reference signatures x to come from the presumed source, while under H2, we assume them to come from a different source. When the questioned signature is genuine, the BF should be above 1, and below 1 if otherwise. A BF of exactly 1 means that the evidence is equally likely under both propositions and provides no relevant information. We transformed the Bayes Factor (Eqn. 14.4) into weight of evidence (WoE), as shown in equation 14.5. This facilitates the visualization of the results, as it reduces spread. This transformation also maps the BFs to a more intuitive scale, which is centered on zero as neutral value. This means that positive weight of evidence implies tipping the scales towards authenticity, while negative WoE would imply tipping the scales towards simulation. Ideally, genuine signatures produce positive weight of evidence, simulated signatures produce negative WoE and non-discriminating, irrelevant evidence amounts to zero.

$$\mathcal{Z}_{i,j} \sim \mathcal{N}(\theta_i, W) \tag{14.1}$$

$$\theta_i \sim \mathcal{N}(\mu, \kappa W) \tag{14.2}$$

$$W \sim \mathrm{Wish}(U, \nu_w) \tag{14.3}$$

$$\mathrm{BF} = \frac{f(y|x, H_1)}{f(y|x, H_2)} \tag{14.4}$$

$$\mathrm{WoE} = \log_{10}\mathrm{BF} \tag{14.5}$$

14.2.2 Experimental Design

Four different experiments were carried out to evaluate the impact of temporal distance on the assessment of signature evidence. Details for every experiment can be found in the following paragraphs.

14.2.2.1 Reference material from a single session

In the first experiment, we have investigated the effect of taking, as reference signatures, a series of signatures that all come from a single session. This refers to the situation a FHE can encounter if only requested reference signatures produced at a single date are available ("requested" reference signatures are opposed to "collected" reference signatures, which are course of business and not specifically produced for the needs of the expertise; such classification refers to the common terminology used by FHE [187]). This experiment was conducted on genuine signatures only. All the signatures from a single session were used as questioned signatures. Reference materials were composed of all the signatures from a single session. One BF value was obtained by comparing each questioned signature and all the reference signatures of that session. We therefore obtained a series of BF values, which was represented by means of a boxplot. The same operation was repeated with reference material coming from other sessions, but always with the same set of questioned signatures. Finally, this procedure produces a series of BFs for every questioned and reference set couple, that are shown on a same boxplot in the results section (14.3). The boxes and quantiles are represented in a common scale to allow their comparison. During the latter, we focused on the spread of BFs, the range of the values and whether they are misleading or not.

Note that while the experiment was also repeated with sets of questioned signatures coming from other sessions, only the results obtained with questioned signatures coming from a single session will be illustrated in this contribution.

14.2.2.2 Reference material from several sessions

In the second experiment, we studied the impact of using reference signatures coming from multiple sessions. This refers to the situation a FHE can encounter if the presumed writer of the signature can be called upon several times to produce "requested" reference samples (see strategies 1 and 2 below) and the situation where a FHE receives a given number of "collected" reference signatures (see strategy 3 below). We have taken all the questioned signatures from a single session. Three strategies were followed for sampling the reference material:

1. the reference material was first composed of all the reference signatures of a given number of randomly taken sessions (cumulative strategy) – this strategy was followed to investigate the effect of time and quantity of reference material;
2. the reference material was then composed of a resampling of 20 signatures drawn among the randomly taken sessions described above (resampling strategy) – this strategy was followed to investigate the effect of multiple sessions and avoid the effect of the sampling size;
3. the reference material was finally composed of a fixed number of randomly taken signatures, without any temporal restrictions (random strategy) – this strategy was followed to investigate the effect of quantity of reference material.

To compare these conditions, the spread and median values for the WoE in these different settings were evaluated, by comparing boxplots of the series of BF values obtained between settings. Further, a reproducibility study on using two sessions is proposed. For this study, several sets of two sessions were randomly drawn and compared to a same set of questioned signatures. The resulting BFs were compared in the same fashion described above.

14.2.2.3 Reference material from sampled intervals

In the third experiment, the impact of non-contemporaneity of the reference material is assessed by using reference signatures coming from sessions more or less distant in time from the session of the questioned signatures. The major difference to the first two experiments is that the reference material is no longer selected according to sessions, but rather selected based on a temporal distance interval. The sampling method used in this experiment can be compared to the situation a FHE encounters when collected reference samples from different periods of time are made available for comparison purposes. As above, questioned signatures were taken from a randomly chosen session. All remaining genuine signatures were then split into groups, based on 10 even time intervals in distance (from the date of the questioned signatures). As reference material, 20 signatures were randomly drawn from each group, in turn. The comparison between each questioned signature and the reference signatures of a given group provided a single BF value. The BF values measured for all the said questioned signatures and a given set of reference signatures were plotted in the form of a boxplot. Based on all groups of reference signatures, a series of boxplots were obtained and represented on a same plot. We especially compared the 'extent' of the BF and its spread according to the increase of temporal distance.

14.2.2.4 The effect of time on the evaluation of simulated signatures

Finally, we were interested in the effect of contemporaneity on the evaluation of simulated signatures. The impact of the point of origin and time coverage of the reference signatures was studied. The same experimental protocol as in the second experiment was followed (section 14.2.2.2), based on the same three sampling strategies for the reference material. The major difference lies in using, as questioned signatures, 100 randomly drawn simulated signatures.

14.3 Results & Discussion

14.3.1 Reference material from a single session

Signature 1 was of medium complexity [17, 188] and of a mixed style [389]. The results for signature 1 can be seen in figure 14.1. The questioned signatures were taken from a session that occurred 24 days after the first session. No misleading BFs (negative in this case) were observed. Most of the results have a median WoE value around 1 (BF of 10), with a small dispersion. Some outliers, showing strong weight of evidence have been observed, but generally it is rather rare for the WoE to be larger than 2. Some sessions show greater dispersion in BF values, most notably the first acquisition sessions. This is most probably due to the lower number of signatures available as references on these particular sessions. A low number of reference signatures leads to less reproducible results and therefore higher dispersion. Another region of interest are the later sessions, where both dispersion and median values are higher than for most sessions. We hypothesized this may be due to the greater temporal distances between the sessions during this period, creating higher variance in the signature features. Other reasons may be the 'training' effects previously observed in signature data (cf. chapter 13). Higher median values, i.e. overall higher BF values, would mean higher similarity between the signatures of these last sessions with the questioned ones. This similarity may be a fortuitous effect of the choice of questioned data. It appeared that results obtained with other sets of questioned signatures (which are not illustrated), no general pattern was observed regarding the sessions showing the higher median BF values, which corroborates the hypothesis of fortuitous similarity.

Signature 2 is of low complexity, especially in terms of signature length, shape and execution. It is illegible and stylized. The results for signature 2 can be seen in figure 14.2. We represented a graph with questioned signatures from an acquisition session toward the middle of the acquisition process. In this graph, we can see that the BFs are also quite low, in general. BF values were not misleading in the shown example, although some misleading evidence was obtained for some other questioned signature sets (not illustrated). The WoE values were around 0.4. Previously examined data also show high variation in features of signature 2, explaining the limited WoE values. For this signature and questioned session, no WoE above 1 was observed. In the signature 2 specific feature sets (sets 2 & 3), two dynamic features were included. Although the WoE values oscillated, they did not vary surprisingly much, as compared to signature 1, whose feature set featured only a single dynamic feature. Researchers have proposed that dynamic global features show less stability in time than static features [233]. We have not been able to see any effect of this instability on our probabilistic system. In comparison to signature 1, WoE values

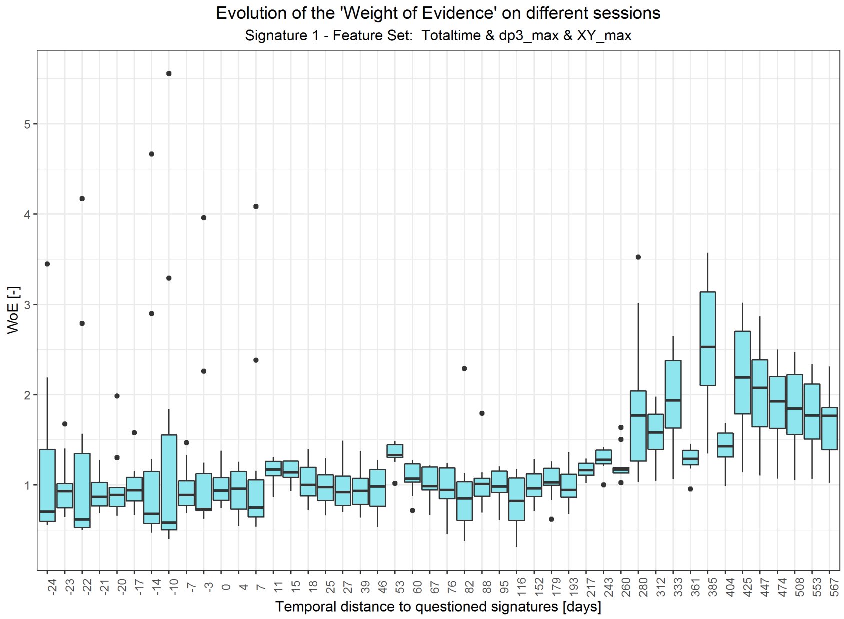


Figure 14.1: Comparison of a set of 'questioned' signatures from an acquisition session of writer 1 to the other acquisition sessions. Temporal distance in days is expressed on the x-axis (the value 0 corresponds to the set of questioned signatures, negative distances involve anterior sessions, while positive distances are posterior). The weight of evidence is expressed on the y-axis.

appear coherent, especially considering the complexity difference.

Signature 3 is a high-complexity, text-based signature. The results for signature 3 can be seen in figure 14.3. The WoE values are quite stable and most lie around 1.5, which corresponds to a BF of 32. No misleading evidence was observed in this example (nor in other trials, not illustrated). The maximum WoE observed lay around 4.5, which corresponds to a BF of 32'000. Dispersion of the WoE values is larger than for the other signatures, although there are some sessions with very small variances. The observed values appear coherent with the two other signatures. The complexity of signature 3 is a lot higher than the other two signatures. The increased length of the signature, as well as the complexity of the shape of the signature lead us to expect higher BF values.

While the results described above are variable between sessions, there did not appear to be any significant tendency that can be attributed to the 'ageing' of the signature. In other words, small time periods [325], such as 18 months, do not heavily affect results obtained with our probabilistic methodology. Two propositions account for the limited impact of time in the setting of this first experiment. First, the robustness of the BF to time effects may be due to the lack of long-term variation in the examined data. As the data set was not collected over the recommended 5-year period (at least), real long-term variation may be absent altogether. Short-term variation may produce changes too small to truly be detected in an evaluative model, especially in an age range where signatures are relatively stable. Second, the alternative explanation for the limited impact of temporal distance is the use of a probabilistic model. The two types of variation, in time and natural, are modeled through a probability distribution. The signature's variation in these cases may be greater than usual, but the model is still well equipped to 'smooth over' the minor variations due to time. A surprising observation includes the evaluation of the questioned data with itself as a reference sample. One may expect that this data would produce very high BFs compared to the other sessions. This is not the case, which may also be attributed to the use of background data. This data is used in conjunction with the reference material to obtain the BF. This may attenuate the effects of ageing and variation. We hypothesize that our method, much like DTW [203, 575], is relatively robust to ageing effects. Nevertheless, we have seen a strong variation between

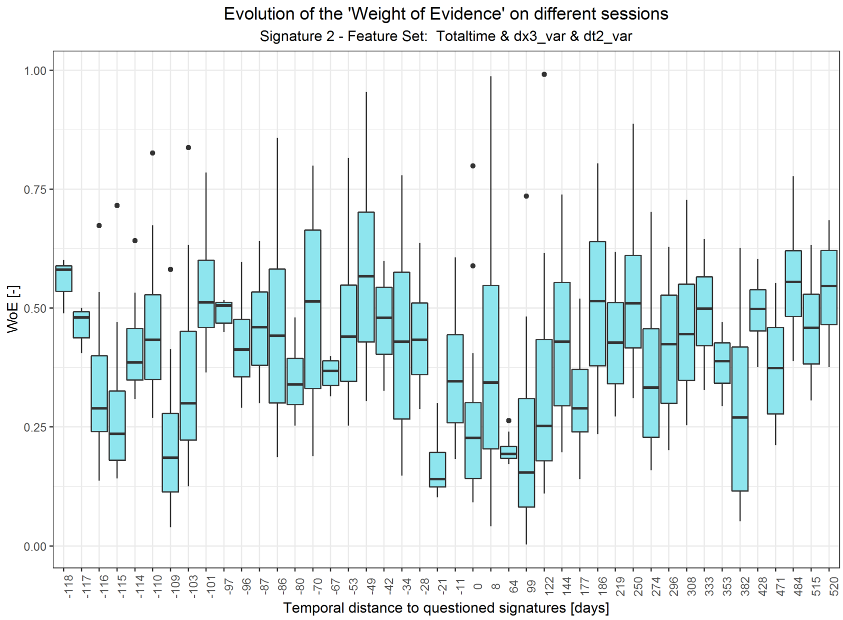


Figure 14.2: Comparison of a set of 'questioned' signatures from an acquisition session of writer 2 to the other acquisition sessions. Temporal distance in days is expressed on the x-axis (the value 0 corresponds to the set of questioned signatures, negative distances involve anterior sessions, while positive distances are posterior). The weight of evidence is expressed on the y-axis.

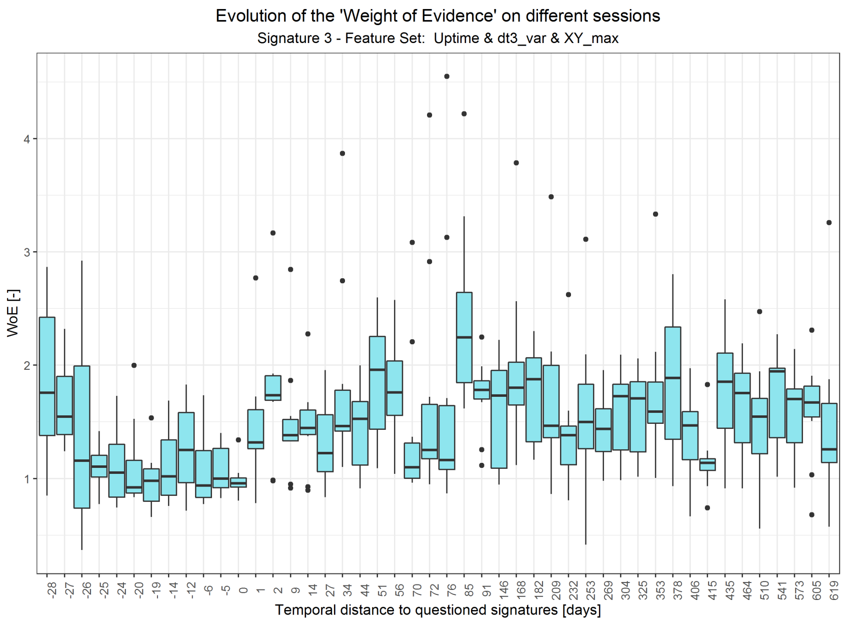


Figure 14.3: Comparison of a set of 'questioned' signatures from an acquisition session of writer 3 to the other acquisition sessions. Temporal distance in days is expressed on the x-axis (the value 0 corresponds to the set of questioned signatures, negative distances involve anterior sessions, while positive distances are posterior). The weight of evidence is expressed on the y-axis.

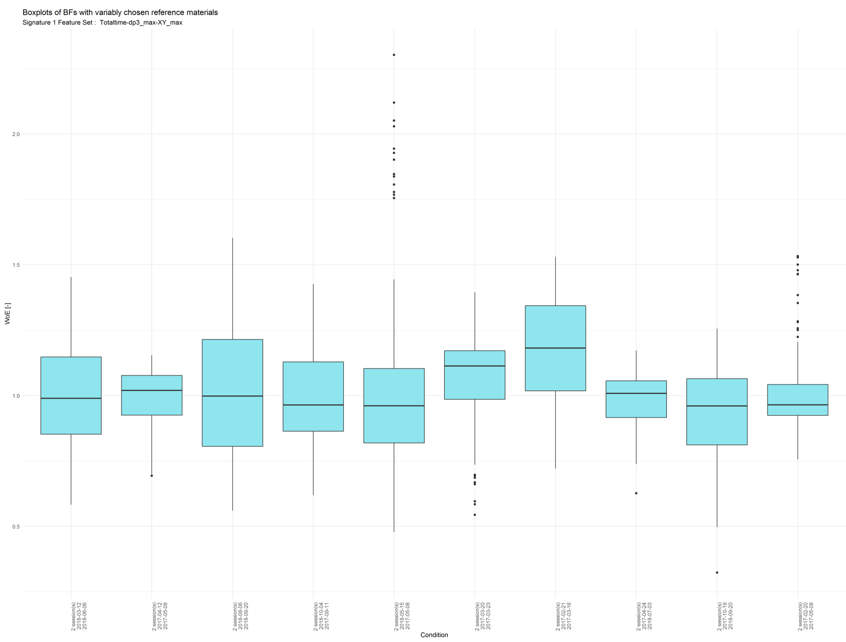


Figure 14.4: Comparison of WoE when using two distinct sessions as reference materials. Several sets of two sessions were randomly drawn from our session pool.

sessions in our dataset, based on both spread and median BF. Examiners should take care when voicing conclusions based on reference material originating from a single data acquisition session. Results may be methodologically ‘correct’ and accurate, but not especially precise and reproducible. The sample of size 1 (in respect to time) is statistically untenable and may lead to over or under evaluation of evidence. Our recommendation for examiners is to, whenever possible, request reference signatures made on multiple days, should the present methodology be applied.

14.3.2 Reference material from several sessions

First and foremost, when using combinations of two sessions drawn at random, results become more reproducible than in the previous experiment (section 14.3.1). In figure 14.4, we can see that the previously very disperse WoE are now all more or less centered on 1. In figure 14.1, these WoE were more spread out, depending on the reference material. Using reference signatures from two different sessions drastically improves reproducibility in both median value and spread.

The results for signature 1 are represented in figure 14.5. The results of the three sampling strategies are shown, from left to right. No misleading evidence was produced. Variation in WoE values appears to be greater when the amount of available reference material is low. This can be seen in the median values and spread, which vary little for the conditions, except for the small samples. As such, all conditions with fewer signatures show higher dispersion (i.e. wider box). Overall time seems to have little impact on the WoE. The results show little difference between the three sampling strategies. There appears to be no specific effect of either time or data quantity on the WoE.

For signature 2 (figure 14.6), many outliers were produced. Some of these outliers are misleading evidence, some are even quite strongly misleading, with a WoE of around -2. The medians are all correctly supporting the genuine signature hypothesis, although WoE values are weak (close to 0). In the signature data, we have observed strong variation in this signature. For the stylised signature with low complexity, the potential for misleading evidence may be higher than for the two other signatures studied here. When changing feature sets (not illustrated), results showed less misleading evidence. The results show little

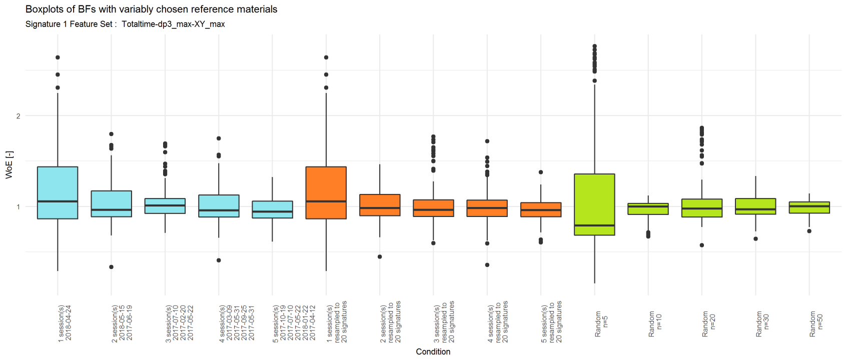


Figure 14.5: Comparison of WoE for Signature 1. The reference material is sampled following the cumulative strategy (five boxplots on the left, blue), the resampling strategy (five central boxplots, orange) and the random strategy (five boxplots on the right, green). The reference material type is indicated on the x axis. The weight of evidence is shown on the y axis. Sampling method 1 (cumulative) is blue, method 2 (resampled) is orange and method 3 (random) is green.

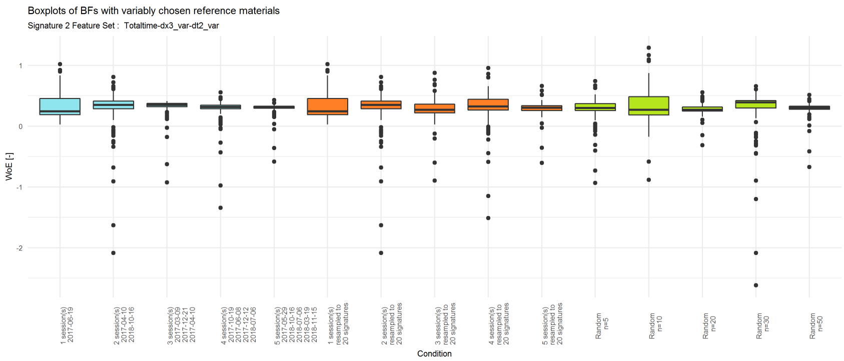


Figure 14.6: Comparison of WoE for Signature 2. The reference material is sampled following the cumulative strategy (five boxplots on the left, blue), the resampling strategy (five central boxplots, orange) and the random strategy (five boxplots on the right, green). The reference material type is indicated on the x axis. The weight of evidence is shown on the y axis. Sampling method 1 (cumulative) is blue, method 2 (resampled) is orange and method 3 (random) is green.

difference between the three sampling strategies. There appears to be no specific effect of either time or data quantity on the WoE.

For signature 3 (figure 14.7), only in the cumulated 4 session condition did we observe a single misleading value. Otherwise, WoE was centered around 1, with many outliers lying toward stronger weight of evidence. The dispersion of the BF values is very low. The only surprising results are those obtained under the condition “resampled 2 sessions” and the condition “randomly sampled 5 signatures”. In these cases, the BF variance appears to be low (small box size), however there are many outliers (points), which could explain the unusually narrow interquartile range. Time and ageing appear to have had almost no impact on this signature. This confirms trends and ageing we have observed in the data itself for this participant (cf. chapter 13). The results show little difference between the three sampling strategies. There appears to be no specific effect of either time or data quantity on the WoE.

In summary, we have observed that increasing the number of sessions from which the reference material is taken is beneficial to reproducibility. The probabilistic model seemed not to be influenced by neither the quantity of data nor the number of sessions. This appears to hold when the questioned signatures are genuine. It would appear that for examiners, simply obtaining signatures written by the suspect on two days makes their inference more reliable.

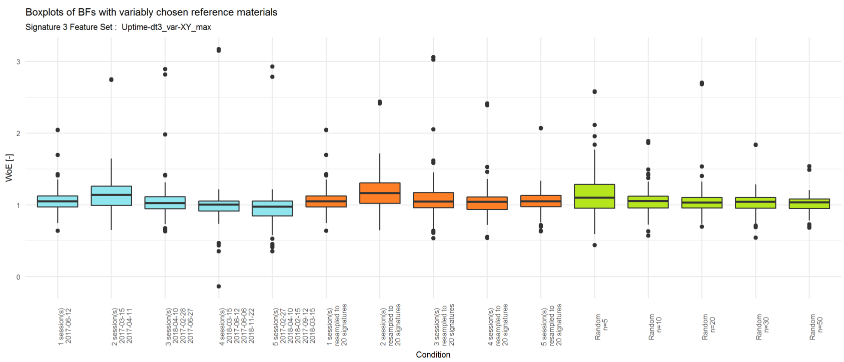


Figure 14.7: Comparison of WoE for Signature 3. The reference material is sampled following the cumulative strategy (five boxplots on the left, blue), the resampling strategy (five central boxplots, orange) and the random strategy (five boxplots on the right, green). The reference material type is indicated on the x axis. The weight of evidence is shown on the y axis. Sampling method 1 (cumulative) is blue, method 2 (resampled) is orange and method 3 (random) is green.

14.3.3 Reference material from sampled intervals

Whatever the signature, the time intervals of the groups of reference signatures were all about 30 or 50 days long. Results were very similar for every signature, showing little impact of the temporal distance on the WoE. The results for signature 1 are shown on figure 14.8. The results for signatures 2 and 3 can be found in figures 14.9 and 14.10 respectively. Several additional trials, with different questioned materials and interval sizes were made. These have had no notable effects and are hence not illustrated in the present document. WoE values depended on the questioned signature dataset. Generally, little or no misleading evidence was obtained, except for signature 2 (not illustrated). As noted in previous experiments, signature 2 was more prone to produce outliers and misleading evidence. The increasing temporal distance between questioned and reference material often did not affect the WoE in terms of value. In some cases, evidential value rose with distanced sessions, while in others it lowered. These effects may however simply be due to natural variation rather than time related s. The dispersion of the WoE values also did not change with the temporal distance. The presence of outliers appears to be related to the signature, rather than temporal distance. In general, temporal distance on the scale studied in our experiment (12 to 18 months) had little impact on the WoE.

Signature 2 is small, short and has low complexity. Found et al. propose that low complexity severely limits the examiners' conclusions [187]. Our results are in agreement with this hypothesis. The second reason for the low BF values is linked to the dynamic features used in the trivariate feature set. We hypothesize that dynamic features are subject to larger variation (over time) than static ones, as proposed by [203, 207, 233]. The high variability leads to spread out (flat; high σ) probability distributions. These flat distributions show less difference in density between them, which in turn leads to lower WoE values. Signature 3 is the most complex and the longest of the three signatures studied. The stability as well as the high overall WoE obtained with this signature surpass both other signature results. These findings are thought to stem either from the high complexity of the signature, or from the low variation of this particular signature.

Overall, the third experiment showed that temporal distance does not seem to affect the probabilistic signature evaluation methodology in a major way. Weight of Evidence, as well as dispersion of the produced values does not vary much when exchanging the reference material. These observations do however have some limitations. First, the maximum temporal distance we could achieve over our 18-month spanning sample is still relatively short, according to the literature. More long-term studies and data sets should be made available to further investigate the impact. In forensic science, cases may

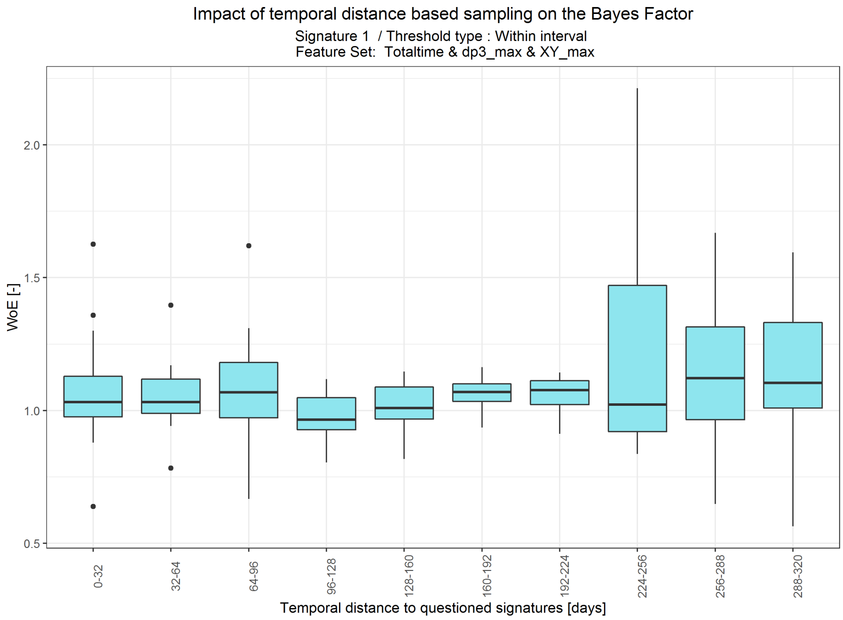


Figure 14.8: Comparison of the WoE produced with interval-based reference selection. WoE was generated on a common set of questioned signatures. 20 reference signatures were randomly drawn from all genuine signature 1 samples falling within the interval. The intervals are represented on the x-axis, the weight of evidence on the y-axis.

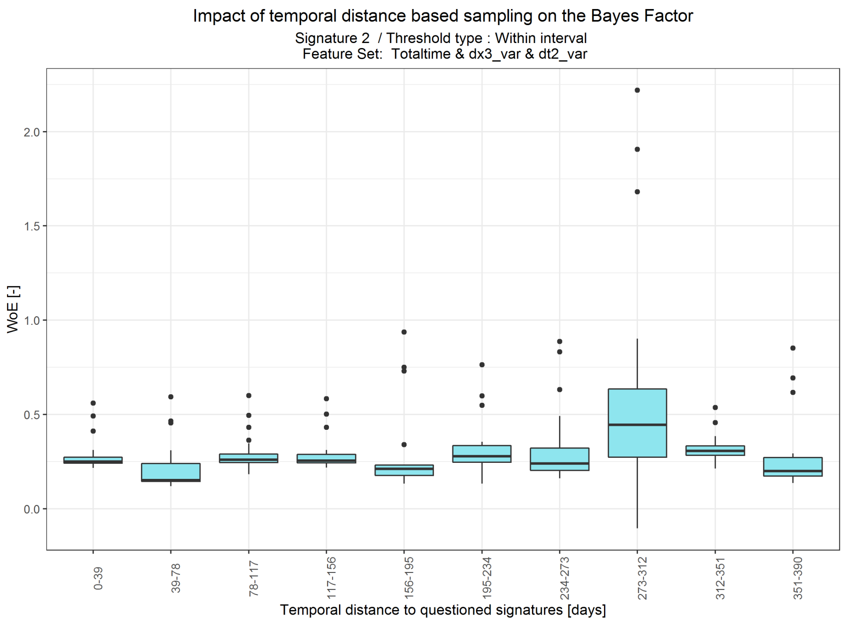


Figure 14.9: Comparison of the WoE produced with interval-based reference selection. WoE was generated on a common set of questioned signatures. 20 reference signatures were randomly drawn from all genuine signature 2 samples falling within the interval. The intervals are represented on the x-axis, the weight of evidence on the y-axis.

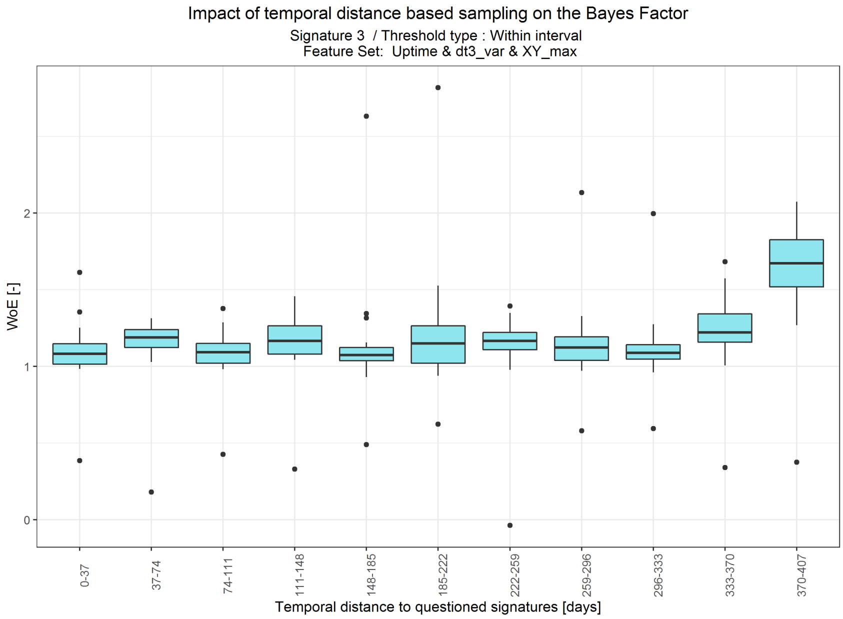


Figure 14.10: Comparison of the WoE produced with interval-based reference selection. WoE was generated on a common set of questioned signatures. 20 reference signatures were randomly drawn from all genuine signature 3 samples falling within the interval. The intervals are represented on the x-axis, the weight of evidence on the y-axis.

also lie much further in the past than 18 months. Second, the age group of the participants matters. All participants were in a stable age range for handwriting. If participants were older, stronger and more diverse effects may be observed. Elderly people may be involved in casework and special care should be taken, as they are susceptible to show stronger changes. In summary, this experiment further supports that data within 18 months is sufficiently representative for signature examination in cases with graphically mature and stable (healthy) people.

14.3.4 The effect of time on the evaluation of simulated signatures

The results obtained with simulations of signature 1 are represented in figure 14.11. It appears that combining multiple sessions (see the first five boxplots) leads to stronger and more dispersed WoE values. In comparison, the values of the WoE produced with random samples are lower and less dispersed. When looking at the random samples, it is clear that the WoE gradually increases in strength (absolute value) when more data is available. In the experiment on genuine signatures (cf. figure 14.5), this effect was not observed. The higher dispersion of the WoE is related to the size of the dataset, rather than the influence of time. Signature 2 shows similar results (cf. figure 14.12). Combining several sessions leads to the same kinds of effects on the WoE, but they are more arbitrary than for signature 1. Depending on the sampled sessions, more or less misleading evidence and dispersion is observed. Random sampling provides a more stable trend towards higher correct WoE. When repeating the experiment, signature 2 shows more variable results than the other signatures. We suppose that this is due to the feature sets, which include two dynamic features, or to the high natural variation of the participant (cf. chapter 13). The combination of low signature complexity and high variability is a known source of errors and misleading evidence. Our hypothesis is that global dynamic features may not be good choices where time invariance is required, as proposed by [233]. When changing the feature set to include only a single dynamic feature, WoE values become more reproducible and less misleading.

In figure 14.13, very similar results can be observed for signature 3. The effects of random sampling

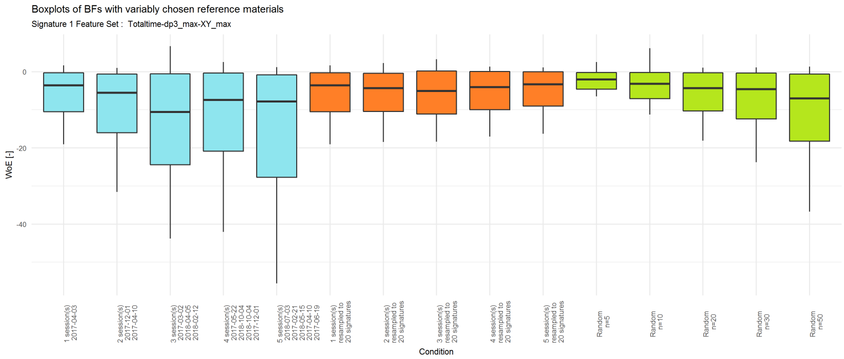


Figure 14.11: Comparison of a sample of simulated questioned signatures with various reference material selections. In the first five conditions, 10 signatures sampled from 1 to 5 sessions were used in the BF computation. In the next 5 conditions, we used randomly sampled signatures from writer 1's genuine signature population. On the x axis, we represented the conditions for the reference material, while on the y axis we represented the weight of evidence.

rather than session-based sampling (i.e. cumulative and resampling strategies) are clearly visible. Session-based sampling had a tendency to produce stronger (higher in absolute value) WoE rather than random sampling. Choosing specific sessions, rather than sampling from a continuous time period, may cause over- or underestimation of the evidential value. It follows naturally that sampling should ideally be done over a continuous timeframe, rather than from separate sessions.

In our methodology, sampling from multiple sessions does not have much impact on the WoE value. It does however impact reproducibility of the results, with more sessions leading to more robust conclusions. The effects of the reference sample size on the WoE, however, are very important when dealing with simulated signatures. The increase of the reference material size produces a steady increase in WoE. This effect was not perceived in the genuine signatures, but is marked for simulated signatures. Our hypothesis for this difference involves the nature of the compared signatures, as well as the quality of the simulated signatures. Simulated signatures are (ideally) very close to their genuine 'target'. Their feature distributions should also be close and partly overlapping (for good forgers), leading to low WoE values. However, if forgers imitate some features badly, they fall completely outside the variation of the signature model (little overlapping in the distributions), leading to high WoE. Furthermore, the reference material set informs the examiner about the presumed source's feature variation. The more material available, the more precise our knowledge about these parameters becomes. This often translates to a narrowing of the feature distributions, leading to higher WoE.

On another note, the size effect far outweighs the influence of time. The 5 central boxplots (reflecting time) show almost no variation in the interquartile range, while the 5 boxplots on the right (reflecting sample size) show a gradual increase with the sample size. The more signatures are in the reference sample, the stronger the produced BFs become. Time appears to causes only slight variations in comparison to the sample size.

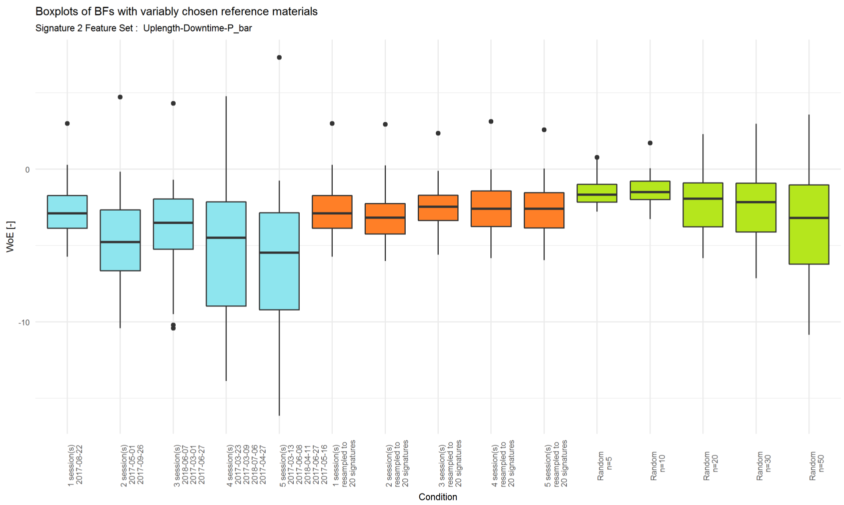


Figure 14.12: Comparison of a sample of simulated questioned signatures with various reference material selections. In the first five conditions, 10 signatures sampled from 1 to 5 sessions were used in the BF computation. In the next 5 conditions, we used randomly sampled signatures from writer 2's genuine signature population. On the x axis, we represented the conditions for the reference material, while on the y axis we represented the weight of evidence.

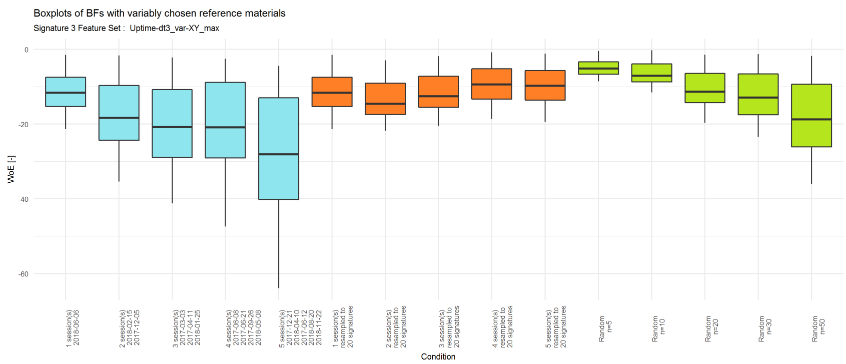


Figure 14.13: Comparison of a sample of simulated questioned signatures with various reference material selections. In the first five conditions, 10 signatures sampled from 1 to 5 sessions were used in the BF computation. In the next 5 conditions, we used randomly sampled signatures from writer 3's genuine signature population. On the x axis, we represented the conditions for the reference material, while on the y axis we represented the weight of evidence.

14.4 Conclusion

We have investigated the impact of time, ageing and temporal distance between questioned and reference signatures throughout this study. We were interested not only in the temporal distance itself, but also in the way reference signatures are selected, whether they are taken from a single session, different sessions, or from time intervals.

First and foremost, the applied methodology worked well on the dataset. A large majority of the calculated BFs correctly supported the ‘true’ hypothesis. The reader needs to keep in mind that the method only uses relatively simple measurements made on the signature, such as durations, lengths and angles. More precision may be achieved by considering shape and local feature information. The methodology is little affected by short temporal distances, such as 18 months. The choice of reference material proves to be an important factor. Choosing reference signatures from single sessions produces greater variance in WoE. Using several sessions produces more reproducible and valid results. Further, we have seen that the expected effect of the temporal distance is absent. We have not observed any consistent and reproducible decreases in the WoE for temporally distant signatures. The sampling method of the reference signatures and the ‘fortuitous’ similarity to the questioned signature produced greater variation than temporal distance. Finally, we have seen that the amount of reference material provided seems to have a stronger effect than time. It appears more important to provide a sufficient number of signatures from the presumed source, than sampling through an extensive time period. Nevertheless, sampling from at least two sessions appears important for reproducibility and validity of the results. As for our study, signatures up to 18 months older are contemporaneous enough to the questioned material not to affect the signature analysis and evaluation.

Although a first in the field of forensic signature analysis, our study has several limitations. The sampled time period of 18 months is reportedly a short time frame [325]. Results justify the use of less recent data for comparisons in a case, but cannot be used to generalize to long temporal distances. Long distances between reference material may react differently. We also studied participants in a ‘stable’ age range. These participants were supposed to be less subject to short-term changes in their signatures. If the participants were to be in other age groups, stronger ageing effects might be present. Additionally, all conclusions of this study are based on three signatures only. A large-scale study is recommended to fully validate and generalize our conclusions. While all of our conclusions are limited to a ‘stable’ and graphically mature population, we feel that our results have shown the benefits of using a probabilistic methodology in forensic casework. They also corroborate statements about contemporaneity and quantity of reference material made by researchers [244, 392].

Given these previous results, we would argue that dealing with time, temporal distance and ageing through a probabilistic approach is promising. The model provides a framework for the variation of the signatures and compensates a lack of knowledge with a small sampling through time. As a result, little misleading evidence was observed overall, which is encouraging. The probabilistic methodology also provides a means of expressing the weight of evidence, rather than reporting a deterministic answer to the question of the source. This weight of evidence dissociates the role of the expert, which resides in informing the court about the evidence, from the role of the court. The decision of who the actual source is and if this person is guilty is left up to the court, which freely appreciates the provided evidence.

14.5 Acknowledgments

The authors would like to thank all volunteers that gave their signatures and put in effort to simulate. We would also like to thank Wacom GmbH Europe, who have provided us with both the hard- and software required for our study.

Bibliography

[1] P. Aepli, O. Ribaux, and E. Summerfield. *Decision Making in Policing: Operations and Management*. EPFL Press, 2011. ISBN: 9782940222544. URL: https://books.google.ch/books?id=8lGDy7DgbL8C.

[2] S. M. Ahmad, L. Y. Ling, R. M. Anwar, M. A. Faudzi, and A. Shakil. "Analysis of the effects and relationship of perceived handwritten signature's size, graphical complexity, and legibility with dynamic parameters for forged and genuine samples". In: *J. Forensic Sci.* 58.3 (2013), pp. 724–31. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.12075. URL: http://www.ncbi.nlm.nih.gov/pubmed/23527753.

[3] S. M. S. Ahmad, A. Shakil, A. R. Ahmad, M. Agil, M. Balbed, and R. M. Anwar. "SIGMA - A Malaysian signatures' database". In: *2008 IEEE/ACS International Conference on Computer Systems and Applications*. 2008, pp. 919–920. ISBN: 2161-5330. DOI: 10.1109/AICCSA.2008.4493644.

[4] C. Aitken and A. Nordgaard. "The Roles of Participants' Differing Background Information in the Evaluation of Evidence". In: *J. Forensic Sci.* 63.2 (2018), pp. 648–649. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.13712. URL: https://www.ncbi.nlm.nih.gov/pubmed/29194595.

[5] C. Aitken, P. Roberts, and G. Jackson. *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings*. English. Ed. by C. Aitken. Electronic Book. Royal Statistical Society, Mar. 2010.

[6] C. G. G. Aitken and D. Lucy. "Evaluation of trace evidence in the form of multivariate data". In: *Appl. Statist.* 53.1 (2004), pp. 109–122.

[7] C. G. G. Aitken and D. Lucy. "Evaluation of trace evidence in the form of multivariate data Corrigendum". In: *Appl. Statist.* 53.4 (2004), pp. 665–666.

[8] C. G. G. Aitken, F. Taroni, and A. Biedermann. "Statistical interpretation of evidence: Bayesian analysis". In: *Forensic Chemistry 1st Ed.* Ed. by Max Houck. Academic Press, 2015.

[9] C.G.G. Aitken and D.A. Stoney. *The Use Of Statistics In Forensic Science*. Ellis Horwood series in forensic science. Taylor & Francis, 1991. ISBN: 9780203017326. URL: https://books.google.ch/books?id=wK9c4KttXj0C.

[10] Colin Aitken, Charles EH Berger, John S Buckleton, Christophe Champod, James Curran, AP Dawid, Ian W Evett, Peter Gill, Joaquin Gonzalez-Rodriguez, Graham Jackson, et al. "Expressing evaluative opinions: a position statement". In: *Science and justice* 51.1 (2011), pp. 1–2.

[11] Colin Aitken, Franco Taroni, and Silvia Bozza. *Statistics and the Evaluation of Evidence for Forensic Scientists 3e*. John Wiley & Sons, Inc, 2021. ISBN: 9781119245438. DOI: https://doi.org/10.1002/9781119245438. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119245438.

[12] Colin G. G. Aitken. "Bayesian Hierarchical Random Effects Models in Forensic Science". In: *Front. Genet.* 9.126 (2018). ISSN: 1664-8021. DOI: 10.3389/fgene.2018.00126. URL: https://www.frontiersin.org/article/10.3389/fgene.2018.00126.

[13] Colin G. G. Aitken and Franco Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists, Second Edition*. The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons, Ltd, 2004.

[14] D. Alamargot and M. F. Morin. "Does handwriting on a tablet screen affect students' graphomotor execution? A comparison between Grades Two and Nine". In: *Human Movement Science* 44 (2015), pp. 32–41. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2015.08.011. URL: http://www.ncbi.nlm.nih.gov/pubmed/26298215.

[15] I. Alberink, A. de Jongh, and C. Rodriguez. "Fingermark evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios". In: *J. Forensic Sci.* 59.1 (2014), pp. 70–81. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.12105. URL: http://www.ncbi.nlm.nih.gov/pubmed/24180303.

[16] Linda C. Alewijnse, C. Elisa van den Heuvel, and Reinoud D. Stoel. "Analysis of Signature Complexity". In: *J. Forensic Document Examination* 21 (2011), pp. 37–49.

[17] Linda C. Alewijnse, C. Elisa Van den Heuvel, Reinoud D. Stoel, and Katrin Franke. "Analysis of Signature Complexity". In: *IGS 2009*. 2009.

[18] F. Alonso-Fernandez, J. Fierrez-Aguilar, and J. Ortega-Garcia. "Sensor Interoperability and Fusion in Signature Verification: A Case Study Using Tablet PC". In: *Intl. Workshop on Biometric Recognition Systems*. Vol. 3781. Oct. 2005, pp. 180–187.

[19] United States of America Government. *PUBLIC LAW 106–229—JUNE 30, 200*. Legal Rule or Regulation. 2000.

[20] American National Standards Institute. *ANSI INCITS 395-2005 Information Technology - Biometric Data Interchange Formats - Signature/Sign Data*. Standard. 2005.

[21] M. Antal and L. Z. Szabó. "On-line verification of finger drawn signatures". In: *2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. 2015, pp. 419–424. DOI: 10.1109/SACI.2016.7507413.

[22] Margit Antal, Laszlo Szabo, and Tünde Tordai. "Online Signature Verification on MOBISIG Finger-Drawn Signature Corpus". In: *Mobile Information Systems* 2018 (2018), pp. 1–15. DOI: 10.1155/2018/3127042.

[23] D.R. Ashbaugh. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. CRC Press, 1999. ISBN: 9781420048810. URL: https://books.google.ch/books?id=SdSOCgAAQBAJ.

[24] Faruk Aşıcıoğlu and Nurten Turan. "Handwriting changes under the effect of alcohol". In: *Forensic Sci. Int.* 132.3 (2003), pp. 201–210. ISSN: 03790738. DOI: 10.1016/s0379-0738(03)00020-3.

[25] Association of Forensic Science Providers. "Standards for the formulation of evaluative forensic science expert opinion". In: *Sci. Justice* 49.3 (2009), pp. 161–164. ISSN: 13550306. DOI: 10.1016/j.scijus.2009.07.004.

[26] Australian Government. *The Electronic Transactions Act 1999*. Legal Rule or Regulation. 1999.

[27] Alessandro Balestrucci, D. Impedovo, and G. Pirlo. "Processing of Handwritten Online Signatures: An Overview and Future Trends". In: *Handwriting: Recognition, Development and Analysis*. Ed. by B.L.D. Bezerra, Cleber Zanchettin, Alejandro Héctor Toselli, and G. Pirlo. Vol. 1. Computer Science, Technology and Applications. Hauppauge NY: Nova Science Publishers, Inc., 2017. Chap. 15, pp. 363–386. ISBN: 978-1536119374.

[28] Agnes S. Bali, Gary Edmond, Kaye N. Ballantyne, Richard I. Kemp, and Kristy A. Martire. "Communicating forensic science opinion: An examination of expert reporting practices". In: *Science & Justice* 60.3 (2020), pp. 216–224. ISSN: 1355-0306. DOI: https://doi.org/10.1016/j.scijus.2019.12.005. URL: http://www.sciencedirect.com/science/article/pii/S1355030619302102.

[29] D.L. Banks, K. Kafadar, D.H. Kaye, and M. Tackett. *Handbook of Forensic Statistics*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2020. ISBN: 9781000096064. URL: https://books.google.ca/books?id=f%5C_cFEAAAQBAJ.

[30] Anastasia Beresneva, Anna Epishkina, Sergey Babkin, Alexey Kurnev, and Vladimir Lermontov. "Handwritten Signature Verification: The State of the Art". In: 636 (2018), pp. 234–238. ISSN: 2194-5357 2194-5365. DOI: 10.1007/978-3-319-63940-6_33.

[31] C. E. Berger, J. Buckleton, C. Champod, I. W. Evett, and G. Jackson. "Evidence evaluation: a response to the court of appeal judgment in R v T". In: *Sci. Justice* 51.2 (2011), pp. 43–9. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2011.03.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/21605824.

[32] David E. Bernstein and Jeffrey D. Jackson. "The Daubert Trilogy in the States". In: *Jurimetrics* 44.3 (2004), pp. 351–366. ISSN: 08971277, 21544344. URL: www.jstor.org/stable/29762857.

[33] Byron Leite Dantas Bezerra, Cleber Zanchettin, Alejandro Héctor Toselli, and Giuseppe Pirlo. *Handwriting : recognition, development and analysis*. https://www.novapublishers.com/: Nova Publishers, Inc., 2017. ISBN: 9781536119572 1536119571.

[34] Kiran Bibi, Saeeda Naz, and Arshia Rehman. "Biometric signature authentication using machine learning techniques: Current trends, challenges and opportunities". In: *Multimedia Tools and Applications* 79.1 (2020), pp. 289–340. ISSN: 1573-7721. DOI: 10.1007/s11042-019-08022-0. URL: https://doi.org/10.1007/s11042-019-08022-0.

[35] A. Biedermann, S. Bozza, P. Garbolino, and F. Taroni. "Decision-theoretic analysis of forensic sampling criteria using bayesian decision networks". In: *Forensic Sci. Int.* 223.1-3 (2012), pp. 217–27. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2012.09.003. URL: https://www.ncbi.nlm.nih.gov/pubmed/23031501.

[36] A. Biedermann, S. Bozza, and F. Taroni. "The decisionalization of individualization". In: *Forensic Sci. Int.* 266 (2016), pp. 29–38. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2016.04.029. URL: https://www.ncbi.nlm.nih.gov/pubmed/27196399.

[37] A. Biedermann, S. Bozza, F. Taroni, and C. Aitken. "Reframing the debate: A question of probability, not of likelihood ratio". In: *Sci. Justice* 56.5 (2016), pp. 392–396. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2016.05.008. URL: https://www.ncbi.nlm.nih.gov/pubmed/27702458.

[38] A. Biedermann, P. Garbolino, and F. Taroni. "The subjectivist interpretation of probability and the problem of individualisation in forensic science". In: *Sci. Justice* 53.2 (2013), pp. 192–200. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2013.01.003. URL: http://www.ncbi.nlm.nih.gov/pubmed/23601728.

[39] A. Biedermann, T. Hicks, F. Taroni, C. Champod, and C. Aitken. "On the use of the likelihood ratio for forensic evaluation: response to Fenton et al". In: *Sci. Justice* 54.4 (2014), pp. 316–8. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2014.04.001. URL: https://www.ncbi.nlm.nih.gov/pubmed/25002051.

[40] A. Biedermann, R. Voisard, and F. Taroni. "Learning about Bayesian networks for forensic interpretation: an example based on the 'the problem of multiple propositions'". In: *Sci. Justice* 52.3 (2012), pp. 191–8. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2012.05.004. URL: http://www.ncbi.nlm.nih.gov/pubmed/22841144.

[41] Alex Biedermann. "Your uncertainty, your probability, your decision". In: *Front. Genet.* 4 (2013). ISSN: 1664-8021. DOI: 10.3389/fgene.2013.00148.

[42] C. Bird, B. Found, K. Ballantyne, and D. Rogers. "Forensic handwriting examiners' opinions on the process of production of disguised and simulated signatures". In: *Forensic Sci. Int.* 195.1-3 (2010), pp. 103–7. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2009.12.001. URL: http://www.ncbi.nlm.nih.gov/pubmed/20060668.

[43] C. Bird, B. Found, and D. Rogers. "Forensic document examiners' skill in distinguishing between natural and disguised handwriting behaviors". In: *J. Forensic Sci.* 55.5 (2010), pp. 1291–5. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/j.1556-4029.2010.01456.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/20533988.

[44] Pascal Bissig. "Signature Verification on Finger Operated Touchscreen Devices". Thesis. 2011.

[45] R. Blanco-Gonzalo, Oscar Miguel-Hurtado, Aitor Mendaza-Ormaza, and Raul Sanchez-Reillo. *Handwritten signature recognition in mobile scenarios: Performance evaluation.* Conference Paper. Newton, MA, USA, 2012. DOI: 10.1109/CCST.2012.6393554.

[46] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and J. Liu-Jimenez. "Performance evaluation of handwritten signature recognition in mobile environments". In: *IET Biom.* 3.3 (2014), pp. 139–146. ISSN: 2047-4946. DOI: 10.1049/iet-bmt.2013.0044.

[47] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and J. Liu-Jimenez. "Usability analysis of dynamic signature verification in mobile environments". In: *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG).* 2013, pp. 1–9.

[48] V.L. Blankers, C.E. van den Heuvel, K. Franke, and L. Vuurpijl. "The ICDAR 2009 Signature Verification Competition". In: *10th International Conference on Document Analysis and Recognition.* 2009, pp. 1403–1407.

[49] Vivian L. Blankers, C. Elisa van den Heuvel, Katrin Y. Franke, and Louis G. Vuurpijl. "ICDAR 2009 Signature Verification Competition". In: (2009), pp. 1403–1407. DOI: 10.1109/icdar.2009.216.

[50] Annabel Bolck, Haifang Ni, and Martin Lopatka. "Evaluating score- and feature-based likelihood ratio models for multivariate continuous data: applied to forensic MDMA comparison". In: *Law, Probability and Risk* 14.3 (2015), pp. 243–266. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgv009.

[51] Annabel Bolck and Amalia Stamouli. "Likelihood Ratios for categorical evidence; Comparison of LR models applied to gunshot residue data". In: *Law, Probability and Risk* 16.2-3 (2017), pp. 71–90. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgx005.

[52] M. H. Bonnet and R. R. Rosa. "Sleep and performance in young adults and older normals and insomniacs during acute sleep loss and recovery". In: *Biol. Psychol.* 25.2 (1987), pp. 153–72. ISSN: 0301-0511 (Print) 0301-0511.

[53] E. Borella, C. Chicherio, A. M. Re, V. Sensini, and C. Cornoldi. "Increased intraindividual variability is a marker of ADHD but also of dyslexia: a study on handwriting". In: *Brain Cogn.* 77.1 (2011), pp. 33–9. ISSN: 1090-2147 (Electronic) 0278-2626 (Linking). DOI: 10.1016/j.bandc.2011.06.005. URL: http://www.ncbi.nlm.nih.gov/pubmed/21798651.

[54] G.E.P. Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis.* Wiley Classics Library. Wiley, 2011. ISBN: 9781118031445. URL: https://books.google.ch/books?id=T8Askeyk1k4C.

[55] George E. P. Box, William Hunter, and Stuart Hunter. *Statistics for Experimenters.* 2nd ed. John Wiley & Sons, 2005.

[56] Silvia Bozza, Franco Taroni, Raymond Marquis, and Matthieu Schmittbuhl. "Probabilistic evaluation of handwriting evidence likelihood ratio for authorship". In: *Applied Statistics.* 57.3 (2008), pp. 329–341.

[57] Jean-Jules Brault and Réjean Plamondon. "A Complexity Measure of Handwritten Curves: Modeling of Dynamic Signature Forgery". In: *IEEE Transaction on systems, man, and cybernetics* 23.2 (1993), pp. 400–413.

[58] GLENN W. BRIER. "VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY". In: *Mon. Weather Rev.* 78.1 (1950), pp. 1–3. DOI: 10.1175/1520-0493(1950)078<0001:Vofeit>2.0.Co;2. URL: https://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281950%29078%3C0001%3AVOFEIT%3E2.0.CO%3B2.

[59] M. Brockly, S. Elliott, J. Burdine, M. Frost, M. Riedle, and R. Guest. "An investigation into biometric signature capture device performance and user acceptance". In: *2014 International Carnahan Conference on Security Technology (ICCST)*. 2014, pp. 1–5. ISBN: 2153-0742. DOI: 10.1109/CCST.2014.6986970.

[60] Michael Brockly, Richard Guest, Stephen Elliott, and James Scott. *Dynamic signature verification and the human biometric sensor interaction model.* Conference Paper. Barcelona, Spain, 2011. DOI: 10.1109/CCST.2011.6095937.

[61] Ton Broeders. "Philosophy of Forensic Identification". In: *Encyclopedia of Criminology and Criminal Justice*. Ed. by G.J.N. Bruinsma and D.L. Weisburd. New York: Springer Science and Business Media, 2014, pp. 3513–3526. DOI: 10.1007/978-1-4614-5690-2_164.

[62] Niko Brümmer. "Measuring, refining and calibrating speaker and language information extracted from speech". Thesis. 2010.

[63] Niko Brümmer. *The BOSARIS Toolkit User Guide Theory, Algorithms and Code for Binary Classifier Score Processing.* Online Multimedia. 2011.

[64] J. S. Buckleton, C. M. Triggs, and C. Champod. "An extended likelihood ratio framework for interpreting evidence". In: *Sci. Justice* 46.2 (2006), pp. 69–78. ISSN: 13550306. DOI: 10.1016/s1355-0306(06)71577-5.

[65] Horst Bunke, Janos Csirik, Zoltan Gingl, and Erika Griechisch. *Online Signature Verification Method Based on the Acceleration Signals of Handwriting Samples.* Conference Paper. Pucón, Chile, 2011.

[66] John M. Butler. "Chapter 16 - Laboratory Reports: Communicating Results and Conclusions". In: *Advanced Topics in Forensic DNA Typing: Interpretation.* Ed. by John M. Butler. San Diego: Academic Press, 2015, pp. 445–495. ISBN: 978-0-12-405213-0. DOI: https://doi.org/10.1016/B978-0-12-405213-0.00016-6. URL: http://www.sciencedirect.com/science/article/pii/B9780124052130000166.

[67] John M. Butler. "Chapter 9 - Statistical Interpretation Overview". In: *Advanced Topics in Forensic DNA Typing: Interpretation.* Ed. by John M. Butler. San Diego: Academic Press, 2015, pp. 213–237. ISBN: 978-0-12-405213-0. DOI: https://doi.org/10.1016/B978-0-12-405213-0.00009-9. URL: http://www.sciencedirect.com/science/article/pii/B9780124052130000099.

[68] L. Cadola, Pierre Margot, and R. Marquis. “Are simple signatures so easy to simulate”. In: *Journal of the American Society of Questioned Document Examiners* 16.2 (2013), pp. 3–11.

[69] M. P. Caligiuri, C. Kim, and K. M. Landy. “Kinematics of signature writing in healthy aging”. In: *J. Forensic Sci.* 59.4 (2014), pp. 1020–4. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.12437. URL: http://www.ncbi.nlm.nih.gov/pubmed/24673648.

[70] M. P. Caligiuri, H. L. Teulings, C. E. Dean, A. B. Niculescu, and J. Lohr. “Handwriting movement analyses for monitoring drug-induced motor side effects in schizophrenia patients treated with risperidone”. In: *Human Movement Science* 28.5 (2009), pp. 633–42. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2009.07.007. URL: http://www.ncbi.nlm.nih.gov/pubmed/19692133.

[71] M. P. Caligiuri, H. L. Teulings, J. V. Filoteo, D. Song, and J. B. Lohr. “Quantitative measurement of handwriting in the assessment of drug-induced parkinsonism”. In: *Hum Mov Sci* 25.4-5 (2006), pp. 510–22. ISSN: 0167-9457 (Print) 0167-9457 (Linking). DOI: 10.1016/j.humov.2006.02.004. URL: https://www.ncbi.nlm.nih.gov/pubmed/16647772.

[72] M.P. Caligiuri and L.A. Mohammed. *The Neuroscience of Handwriting: Applications for Forensic Document Examination*. Taylor & Francis, 2012. ISBN: 9781439871409. URL: https://books.google.ch/books?id=Ng5lR9bECtIC.

[73] R. Camicioli, S. Mizrahi, J. Spagnoli, C. Bula, J. F. Demonet, F. Vingerhoets, A. von Gunten, and B. Santos-Eggimann. “Handwriting and pre-frailty in the Lausanne cohort 65+ (Lc65+) study”. In: *Arch. Gerontol. Geriatr.* 61.1 (2015), pp. 8–13. ISSN: 1872-6976 (Electronic) 0167-4943 (Linking). DOI: 10.1016/j.archger.2015.01.006. URL: http://www.ncbi.nlm.nih.gov/pubmed/25910643.

[74] C. Carmona-Duarte, R. de Torres-Peralta, M. Diaz, M. A. Ferrer, and M. Martin-Rincon. “Myoelectronic signal-based methodology for the analysis of handwritten signatures”. In: *Hum Mov Sci* 55 (2017), pp. 18–30. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2017.07.002. URL: https://www.ncbi.nlm.nih.gov/pubmed/28750258.

[75] Cristina Carmona-Duarte, Miguel A. Ferrer, Antonio Parziale, and Angelo Marcelli. “Temporal evolution in synthetic handwriting”. In: *Pattern Recognit.* 68 (2017), pp. 233–244. ISSN: 00313203. DOI: 10.1016/j.patcog.2017.03.019.

[76] C. Champod. “Fingerprint examination: towards more transparency”. In: *Law, Probability and Risk* 7.2 (2007), pp. 111–118. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgm023.

[77] C. Champod. “Identification and Individualization”. In: *Wiley Encyclopedia of Forensic Science*. John Wiley & Sons, 2009. DOI: 10.1002/9780470061589.fsa140. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470061589.fsa140.

[78] C. Champod. “Overview and Meaning of Identification/Individualization”. In: *Encyclopedia of Forensic Sciences, Second Edition*. Ed. by Jay Siegel and Pekka J. Saukko. Vol. 3. Waltham: Academic Press, 2013, pp. 303–309.

[79] C. Champod and I. Evett. “Evidence Interpretation: A Logical Approach”. In: *Wiley Encyclopedia of Forensic Science*. Ed. by Allan Jamieson and Andre Moenssens. John Wiley & Sons, 2009. ISBN: 9780470018262. DOI: 10.1002/9780470061589.fsa122. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470061589.fsa122.

[80] C. Champod, I. W. Evett, and G. Jackson. “Establishing the most appropriate databases for addressing source level propositions”. In: *Sci. Justice* 44.3 (2004), pp. 153–164. ISSN: 13550306. DOI: 10.1016/s1355-0306(04)71708-6.

[81] Christophe Champod. “Fingerprint identification: advances since the 2009 National Research Council report”. In: *Philosophical Transactions: Biological Sciences* 370.1674 (2015), pp. 1–10. ISSN: 1471-2970 (Electronic) 0962-8436 (Linking). DOI: 10.1098/rstb.2014.0259. URL: www.jstor.org/stable/24505151.

[82] Christophe Champod, D. Baldwin, Franco Taroni, and J. S. Buckleton. “Firearm and tool marks identification: the Bayesian approach”. In: *AFTE Journal* 35 (2003), pp. 307–316.

[83] Christophe Champod and Didier Meuwly. “The inference of identity in forensic speaker recognition”. In: *Speech Communication* 31.2 (2000), pp. 193–203. ISSN: 0167-6393. DOI: https://doi.org/10.1016/S0167-6393(99)00078-3. URL: http://www.sciencedirect.com/science/article/pii/S0167639399000783.

[84] S.H.C. Chan. “Paper-based vs computer-based writing assessment: divergent, equivalent or complementary?” In: *Assessing Writing* 36 (2018), pp. 1–2. ISSN: 10752935. DOI: 10.1016/j.asw.2018.04.001.

[85] X. H. Chen, C. Champod, X. Yang, S. P. Shi, Y. W. Luo, N. Wang, Y. C. Wang, and Q. M. Lu. “Assessment of signature handwriting evidence via score-based likelihood ratio based on comparative measurement of relevant dynamic features”. In: *Forensic Sci. Int.* 282 (2018), pp. 101–110. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2017.11.022. URL: https://www.ncbi.nlm.nih.gov/pubmed/29182954.

[86] Siddhartha Chib and Ivan Jeliazkov. “Marginal Likelihood From the Metropolis–Hastings Output”. In: *J. Am. Stat. Assoc.* 96.453 (2001), pp. 270–281. ISSN: 0162-1459. DOI: 10.1198/016214501750332848. URL: https://doi.org/10.1198/016214501750332848.

[87] S. A. Cole. “Forensics without uniqueness, conclusions without individualization: the new epistemology of forensic identification”. In: *Law, Probability and Risk* 8.3 (2009), pp. 233–255. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgp016.

[88] S. A. Cole. “Individualization is dead, long live individualization! Reforms of reporting practices for fingerprint analysis in the United States”. In: *Law, Probability and Risk* 13.2 (2014), pp. 117–150. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgt014.

[89] Simon A. Cole. "A surfeit of science: The "CSI effect" and the media appropriation of the public understanding of science". In: *Public Understanding of Science* 24.2 (2015), pp. 130–146. DOI: 10.1177/0963662513481294. URL: http://journals.sagepub.com/doi/abs/10.1177/0963662513481294.

[90] Simon A. Cole and Rachel Dioso-Villa. "Investigating the CSI Effect Effect: Media and Litigation Crisis in Criminal Law". In: *Stan. L. Rev.* 61 (2009), pp. 1335–1374.

[91] Collins English Dictionary. *Digital Signature Definition.* Web Page. 2018. URL: https://www.collinsdictionary.com/dictionary/english/digital-signature.

[92] Collins English Dictionary. *Electronic Signature Definition.* Web Page. 2018. URL: https://www.collinsdictionary.com/dictionary/english/electronic-signature.

[93] Computer Hope. *Definition of 'Digitizer'.* Web Page. Apr. 30, 2020. URL: https://www.computerhope.com/jargon/d/digitizer.htm.

[94] Russel Conduit. "The effect of sleep deprivation on the spatial characteristics of handwriting". In: *Journal of Forensic Document Examination* 19.1 (2008), pp. 29–39.

[95] Assemblée fédérale de la Confédération Suisse. *Loi fédérale sur les services de certification dans le domaine de la signature électronique et des autres applications des certificats numériques.* Legal Rule or Regulation. 2017.

[96] Connecting Europe Facility. *What is an electronic signature?* Web Page. 2020. URL: https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=46992760.

[97] James V.P. Conway. *Evidential Documents.* Springfield: Charles C. Thomas, 1959.

[98] R. Cook, I. W. Evett, G. Jackson, P. J. Jones, and J. A. Lambert. "A hierarchy of propositions: deciding which level to address in casework". In: *Sci. Justice* 38.4 (1998), pp. 231–239. ISSN: 1355-0306. DOI: 10.1016/S1355-0306(98)72117-3. URL: http://dx.doi.org/10.1016/S1355-0306(98)72117-3.

[99] R. Cook, I. W. Evett, G. Jackson, P. J. Jones, and J. A. Lambert. "A model for case assessment and interpretation". In: *Sci. Justice* 38.3 (1998), pp. 151–156. ISSN: 13550306. DOI: 10.1016/s1355-0306(98)72099-4.

[100] Court of Appeals of Virginia. *Ira Dale Bennett v Commonwealth of Virginia.* Case. Court of Appeals of Virginia, Salem, 2006.

[101] Krzysztof Cpałka, Marcin Zalasiński, and Leszek Rutkowski. "A new algorithm for identity verification based on the analysis of a handwritten dynamic signature". In: *Appl. Soft Comput.* 43 (2016), pp. 47–56. ISSN: 15684946. DOI: 10.1016/j.asoc.2016.02.017.

[102] Frank Crispino, Olivier Ribaux, Max Houck, and Pierre Margot. "Forensic science – A true science?" In: *Aust. J. Forensic Sci.* 43.2-3 (2011), pp. 157–176. ISSN: 0045-0618 1834-562X. DOI: 10.1080/00450618.2011.555416.

[103] James M. Curran. "The Statistical Interpretation of Forensic Glass Evidence". In: *International Statistical Review / Revue Internationale de Statistique* 71.3 (2003), pp. 497–520. ISSN: 03067734, 17515823. URL: www.jstor.org/stable/1403825.

[104] Marjory Da Costa-Abreu and Michael Fairhurst. "Using age to enhance performance in biometrics". In: *Age Factors in Biometric Processing*. Security. Institution of Engineering and Technology, 2013, pp. 63–89. DOI: doi:10.1049/PBSP010E_ch4. URL: https://digital-library.theiet.org/content/books/10.1049/pbsp010e_ch4.

[105] Krishnanand Das, Christian Viard-Gaudin, Sargur N. Srihari, Harish Srinivasan, and Richard Zanibbi. "Questioned document workflow for handwriting with automated tools". In: 8297 (2012), 82970K-82970K–9. ISSN: 0277786X. DOI: 10.1117/12.912104.

[106] L. J. Davis, C. P. Saunders, A. Hepler, and J. Buscaglia. "Using subsampling to estimate the strength of handwriting evidence via score-based likelihood ratios". In: *Forensic Sci. Int.* 216.1-3 (2012), pp. 146–57. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2011.09.013. URL: http://www.ncbi.nlm.nih.gov/pubmed/22018850.

[107] A.C. Davison, D.V. Hinkley, R. Gill, B.D. Ripley, S. Ross, M. Stein, D. Williams, and B. Silverman. *Bootstrap Methods and Their Application*. Bootstrap Methods and Their Application. Cambridge University Press, 1997. ISBN: 9780521574716. URL: https://books.google.lu/books?id=4aCDbm%5C_t8jUC.

[108] Stephen P. Day. "Handwriting and Signatures, Interpretation of Comparison Results". In: *Wiley Encyclopedia of Forensic Science*. John Wiley & Sons, Ltd, 2009. ISBN: 9780470061589. DOI: 10.1002/9780470061589.fsa138. URL: http://dx.doi.org/10.1002/9780470061589.fsa138.

[109] B. De Finetti. *Probabilismo: saggio critico sulla teoria delle probabilità e sul valore della scienza*. Biblioteca di filosofia / Perrella. Perrella, 1931. URL: https://books.google.ch/books?id=1lw3GwAACAAJ.

[110] M.H. DeGroot, S.E. Fienberg, and CARNEGIE-MELLON UNIV PITTSBURGH PA Dept. of STATISTICS. *Assessing Probability Assessors: Calibration and Refinement*. Defense Technical Information Center, 1981. URL: https://books.google.ch/books?id=IOm7NwAACAAJ.

[111] Morris H. DeGroot and Stephen E. Fienberg. "The Comparison and Evaluation of Forecasters". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 32.1/2 (1983), pp. 12–22. ISSN: 00390526, 14679884. DOI: 10.2307/2987588. URL: www.jstor.org/stable/2987588.

[112] Mark P. Denbeaux and D. Michael Risinger. "Kumho Tire and Expert Reliability: How the Question you ask Gives the Answer You Get". In: *Seton Hall Law Review* 13.15 (2003), pp. 17–75.

[113] D. Dessimoz, J. Richiardi, C. Champod, and A. Drygajlo. "Multimodal biometrics for identity documents (MBioID)". In: *Forensic Sci. Int.* 167.2-3 (2007), pp. 154–9. ISSN: 0379-0738 (Print) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2006.06.037. URL: https://www.ncbi.nlm.nih.gov/pubmed/16890391.

[114] Tahnee N. Dewhurst, Kaye N. Ballantyne, and Bryan Found. “Empirical investigation of biometric, non-visible, intra-signature features in known and simulated signatures”. In: *Aust. J. Forensic Sci.* (2016), pp. 1–17. ISSN: 0045-0618 1834-562X. DOI: 10.1080/00450618.2015.1126637.

[115] Nicolò Di Toma. “Past and present in the research on handwriting: time to improve reliability”. In: *Handwriting Research: Quality and Validation*. 2018. ISBN: 3746720389.

[116] Nicolò Di Toma, Mirka MANTOAN, and Stefania RENDINA. *Modification in Handwriting Size Changes the Velocity of Execution: Literature and Experimental Results*. Conference Paper. Gaeta, Italy, 2017.

[117] M. Diaz, M. A. Ferrer, S. Ramalingam, and R. Guest. “Investigating the Common Authorship of Signatures by Off-Line Automatic Signature Verification Without the Use of Reference Signatures”. In: *IEEE Trans. Inf. Forensics Security* 15 (2020), pp. 487–499. ISSN: 1556-6021. DOI: 10.1109/TIFS.2019.2924195.

[118] M. Diaz, M. A. A. Ferrer, and J. J. Quintana Hernandez. “Anthropomorphic features for On-line Signatures”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* (2018), pp. 1–1. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2018.2869163.

[119] M. Diaz, A. Fischer, M. A. Ferrer, and R. Plamondon. “Dynamic Signature Verification System Based on One Real Signature”. In: *IEEE Trans. Cybern.* 48.1 (2018), pp. 228–239. ISSN: 2168-2267. DOI: 10.1109/TCYB.2016.2630419.

[120] M. Diaz, A. Fischer, R. Plamondon, and M. A. Ferrer. “Towards an automatic on-line signature verifier using only one reference per signer”. In: *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. 2015, pp. 631–635. DOI: 10.1109/ICDAR.2015.7333838.

[121] Moises Diaz, Miguel A. Ferrer, Donato Impedovo, Muhammad Imran Malik, Giuseppe Pirlo, and Réjean Plamondon. “A Perspective Analysis of Handwritten Signature Technology”. In: *ACM Comput. Surv.* 51.6 (2019), pp. 1–39. ISSN: 0360-0300. DOI: 10.1145/3274658.

[122] W. Diffie and M. Hellman. “New directions in cryptography”. In: *IEEE Transactions on Information Theory* 22.6 (1976), pp. 644–654. DOI: 10.1109/TIT.1976.1055638.

[123] R. A. Dixon, D. Kurzman, and I. C. Friesen. “Handwriting performance in younger and older adults: age, familiarity, and practice effects”. In: *Psychol. Aging* 8.3 (1993), pp. 360–70. ISSN: 0882-7974 (Print) 0882-7974.

[124] M. Djioua and R. Plamondon. “A new algorithm and system for the characterization of handwriting strokes with delta-lognormal parameters”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 31.11 (2009), pp. 2060–72. ISSN: 1939-3539 (Electronic) 0098-5589 (Linking). DOI: 10.1109/TPAMI.2008.264. URL: https://www.ncbi.nlm.nih.gov/pubmed/19762931.

[125] M. Djioua and R. Plamondon. "Studying the variability of handwriting patterns using the Kinematic Theory". In: *Human Movement Science* 28.5 (2009), pp. 588–601. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2009.01.005. URL: http://www.ncbi.nlm.nih.gov/pubmed/19328575.

[126] J. G. A. Dolfing, E. H. L. Aarts, and J. J. G. M. van Oosterhout. "On-line signature verification with hidden Markov models". In: *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*. Vol. 2. 1998, 1309–1312 vol.2. ISBN: 1051-4651. DOI: 10.1109/ICPR.1998.711942.

[127] Rafal Doroz, Piotr Porwik, and Tomasz Orczyk. "Dynamic signature verification method based on association of features with similarity measures". In: *Neurocomputing* 171 (2016), pp. 921–931. ISSN: 09252312. DOI: 10.1016/j.neucom.2015.07.026.

[128] N. van Drempt, A. McCluskey, and N. A. Lannin. "A review of factors that influence adult handwriting performance". In: *Aust Occup Ther J* 58.5 (2011), pp. 321–8. ISSN: 1440-1630 (Electronic) 0045-0766 (Linking). DOI: 10.1111/j.1440-1630.2011.00960.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/21957916.

[129] P. Drobintsev, D. Sergeev, and N. Voinov. "Online signature verification system for mobile devices with multilevel pressure sensor". In: *2017 XX IEEE International Conference on Soft Computing and Measurements (SCM)*. 2017, pp. 424–426. DOI: 10.1109/SCM.2017.7970606.

[130] I. Dror. "The ambition to be scientific: human expert performance and objectivity". In: *Sci. Justice* 53.2 (2013), pp. 81–2. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2013.03.002. URL: http://www.ncbi.nlm.nih.gov/pubmed/23601713.

[131] I. E. Dror. "Biases in forensic experts". In: *Science* 360.6386 (2018), p. 243. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.aat8443. URL: https://www.ncbi.nlm.nih.gov/pubmed/29674567.

[132] Itiel Dror. "Cognitive Bias in Forensic Science". In: *2012 Yearbook of Science & Technology*. McGraw-Hill, 2012, pp. 43–45.

[133] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy. "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease". In: *Comput. Methods Programs Biomed.* 117.3 (2014), pp. 405–11. ISSN: 1872-7565 (Electronic) 0169-2607 (Linking). DOI: 10.1016/j.cmpb.2014.08.007. URL: http://www.ncbi.nlm.nih.gov/pubmed/25261003.

[134] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy. "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease". In: *Artif. Intell. Med.* 67 (2016), pp. 39–46. ISSN: 1873-2860 (Electronic) 0933-3657 (Linking). DOI: 10.1016/j.artmed.2016.01.004. URL: https://www.ncbi.nlm.nih.gov/pubmed/26874552.

[135] Andrzej Drygajlo and Weifeng Li. “Client-specific A-stack model for adult face verification across aging”. In: *Signal, Image and Video Processing* 5.4 (2011), pp. 431–441. ISSN: 1863-1703 1863-1711. DOI: 10.1007/s11760-011-0246-4.

[136] Bruno Dumas, Catherine Pugin, Jean Hennebert, Dijana Petrovska-Delacrétaz, Andreas Humm, Florian Evequoz, Rolf Ingold, and Didier Rotz. “MyIDea - Multimodal Biometrics Database, Description of Acquisition Protocols”. In: *Proc. of Third COST 275 Workshop* (2005).

[137] J. S. Durmer and D. F. Dinges. “Neurocognitive consequences of sleep deprivation”. In: *Semin. Neurol.* 25.1 (2005), pp. 117–29. ISSN: 0271-8235 (Print) 0271-8235. DOI: 10.1055/s-2005-867080.

[138] Mehr Yahya Durrani, Salabat Khan, and Shehzad Khalid. “VerSig: a new approach for online signature verification”. In: *Cluster Computing* (2017). ISSN: 1386-7857 1573-7543. DOI: 10.1007/s10586-017-1129-4.

[139] Shimon Edelman and Tamar Flash. “A model of handwriting”. In: *Biol. Cybern.* 57.1 (1987), pp. 25–36. ISSN: 1432-0770. DOI: 10.1007/BF00318713. URL: https://doi.org/10.1007/BF00318713.

[140] G. Edmond, A. Towler, B. Growns, G. Ribeiro, B. Found, D. White, K. Ballantyne, R. A. Searston, M. B. Thompson, J. M. Tangen, R. I. Kemp, and K. Martire. “Thinking forensics: Cognitive science for forensic practitioners”. In: *Sci. Justice* 57.2 (2017), pp. 144–154. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2016.11.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/28284440.

[141] Gary Edmond, Bryan Found, Kristy Martire, Kaye Ballantyne, David Hamer, Rachel Searston, Matthew Thompson, Emma Cunliffe, Richard Kemp, Mehera San Roque, Jason Tangen, Rachel Dioso-Villa, Andrew Ligertwood, David Hibbert, David White, Gianni Ribeiro, Glenn Porter, Alice Towler, and Andrew Roberts. “Model forensic science”. In: *Aust. J. Forensic Sci.* 48.5 (2016), pp. 496–537. ISSN: 0045-0618 1834-562X. DOI: 10.1080/00450618.2015.1128969.

[142] S. J. Elliott, K. O’connor, E. Bartlow, J. J. Robertson, and R. M. Guest. “Expanding the human-biometric sensor interaction model to identity claim scenarios”. In: *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*. 2015, pp. 1–6. DOI: 10.1109/ISBA.2015.7126362.

[143] Stephen Elliott. “Differentiation of Signature Traits vis-à-vis Mobile and Table Based Digitizers”. In: *Etri Journal - ETRI J* 26 (2004), pp. 641–646. DOI: 10.4218/etrij.04.0104.0030.

[144] European Network of Forensic Science Institutes. *Best Practice Manual for the Forensic Examination of Handwriting Version 2*. Report. European Network of Forensic Science Institutes, June 2015. URL: http://enfsi.eu/documents/best-practice-manuals/.

[145] European Network of Forensic Science Institutes. *ENFSI Guideline for evaluative reporting in forensic science*. Report. European Network of Forensic Science Institutes, 2015. URL: `http://enfsi.eu/news/enfsi-guideline-evaluative-reporting-forensic-science/`.

[146] E. Enzinger, G. S. Morrison, and F. Ochoa. "A demonstration of the application of the new paradigm for the evaluation of forensic evidence under conditions reflecting those of a real forensic-voice-comparison case". In: *Sci. Justice* 56.1 (2016), pp. 42–57. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: `10.1016/j.scijus.2015.06.005`. URL: `https://www.ncbi.nlm.nih.gov/pubmed/26746825`.

[147] Brian David Eoff and Tracy Hammond. *Who dotted that 'i'?: context free user differentiation through pressure and tilt pen data*. Conference Paper. Kelowna, British Columbia, Canada, 2009.

[148] Céline Equey, R. Marquis, and Williams D. Mazzella. "Influence of Writing Posture on the Dimensions of Signatures". In: *Journal of the American Society of Questioned Document Examiners* (2007), pp. 53–59.

[149] M. Erbilek and M. Fairhurst. "Analysis of ageing effects in biometric systems: difficulties and limitations". In: *Age factors in biometric processing*. Ed. by M. Fairhurst. Vol. 1. 2013, pp. 279–301.

[150] M. Erbilek and M. Fairhurst. "Framework for managing ageing effects in signature biometrics". In: *IET Biom.* 1.2 (2012). ISSN: 20474938. DOI: `10.1049/iet-bmt.2012.0011`.

[151] Colin B. Estabrooks. "Measuring Relative Pen Pressure to Authenticate Signatures". In: *Journal of the American Society of Questioned Document Examiners* (2000), pp. 56–64.

[152] I. W. Evett. "Bayesian Inference and Forensic Science: Problems and Perspectives". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 36.2/3 (1987), pp. 99–105. ISSN: 00390526, 14679884. DOI: `10.2307/2348502`. URL: `www.jstor.org/stable/2348502`.

[153] I. W. Evett. "Towards a uniform framework for reporting opinions in forensic science casework". In: *Sci. Justice* 38.3 (1998), pp. 198–202. ISSN: 13550306. DOI: `10.1016/s1355-0306(98)72105-7`.

[154] I. W. Evett, C. E. H. Berger, J. S. Buckleton, C. Champod, and G. Jackson. "Finding the way forward for forensic science in the US-A commentary on the PCAST report". In: *Forensic Sci. Int.* 278 (2017), pp. 16–23. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: `10.1016/j.forsciint.2017.06.018`. URL: `https://www.ncbi.nlm.nih.gov/pubmed/28688344`.

[155] I. W. Evett, G. Jackson, and J. A. Lambert. "More on the hierarchy of propositions: exploring the distinction between explanations and propositions". In: *Sci. Justice* 40.1 (2000), pp. 3–10. ISSN: 13550306. DOI: `10.1016/s1355-0306(00)71926-5`.

[156] I. W. Evett, G. Jackson, J. A. Lambert, and S. McCrossan. "The impact of the principles of evidence interpretation on the structure and content of statements". In: *Sci. Justice* 40.4 (2000), pp. 233–239. ISSN: 13550306. DOI: `10.1016/s1355-0306(00)71993-9`.

[157] I. W. Evett and R. N. Totty. “A Study of the Variation in the Dimensions of Genuine Signatures”. In: *J. Forensic Sci. Soc.* 25.3 (1985), pp. 207–215. ISSN: 00157368. DOI: 10.1016/s0015-7368(85)72393-6.

[158] Ian Evett. “The logical foundations of forensic science: towards reliable knowledge”. In: *Philosophical Transactions: Biological Sciences* 370.1674 (2015), pp. 1–10. ISSN: 1471-2970 (Electronic) 0962-8436 (Linking). DOI: 10.1098/rstb.2014.0263. URL: www.jstor.org/stable/24505156.

[159] Logan A. Ewanation, Susan Yamamoto, Jordan Monnink, and Evelyn M. Maeder. “Perceived realism and the CSI-effect”. In: *Cogent Social Sciences* 3.1 (2017), p. 1294446. ISSN: null. DOI: 10.1080/23311886.2017.1294446. URL: https://www.tandfonline.com/doi/abs/10.1080/23311886.2017.1294446.

[160] The Expert Working Group for Human Factors in Handwriting Examination. “Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach”. In: (2020). DOI: https://doi.org/10.6028/NIST.IR.8282r1.

[161] David Faigman. “The Daubert Revolution and the Birth of Modernity: Managing Scientific Evidence in the Age of Science”. In: *University of California at Davis Law Review* 46 (2013).

[162] David L. Faigman. “Expert Evidence in Flatland The Geometry of a World Without Scientific Culture”. In: *Seton Hall L. Rev.* 34 (2003), pp. 255–267.

[163] David L. Faigman and A. J. Baglioni. “Bayes’ theorem in the trial process”. In: *Law and Human Behavior* 12.1 (1988), pp. 1–17. ISSN: 1573-661X. DOI: 10.1007/BF01064271. URL: https://doi.org/10.1007/BF01064271.

[164] M. Fairhurst, M. Erbilek, and C. Li. “ENHANCING THE FORENSIC VALUE OF HANDWRITING USING EMOTION PREDICTION”. In: (2014).

[165] Michael Fairhurst. *Age Factors in Biometric Processing*. Institution of Engineering and Technology, 2013. ISBN: 1849195021. URL: https://digital-library.theiet.org/content/books/sc/pbsp010e.

[166] Yuxun Fang, Wenxiong Kang, Qiuxia Wu, and Lei Tang. “A novel video-based system for in-air signature verification”. In: *Computers & Electrical Engineering* 57 (2017), pp. 1–14. ISSN: 00457906. DOI: 10.1016/j.compeleceng.2016.11.010.

[167] M. Faundez-Zanuy, E. Sesa-Nogueras, and J. Roure-Alcobé. “On the relevance of age in handwritten biometric recognition”. In: *2012 IEEE International Carnahan Conference on Security Technology (ICCST)*. 2012, pp. 105–109. ISBN: 2153-0742. DOI: 10.1109/CCST.2012.6393544.

[168] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobé, A. Esposito, J. Mekyska, and K. López-de-Ipiña. “A preliminary study on aging examining online handwriting”. In: *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*. 2014, pp. 221–224. DOI: 10.1109/CogInfoCom.2014.7020449.

[169] Marcos Faundez-Zanuy, Antonio Satue-Villar, Jiri Mekyska, Viridiana Arreola, Pilar Sanz, Carles Paul, Luis Guirao, Mateu Serra, Laia Rofes, Pere Clavé, Enric Sesa-Nogueras, and Josep Roure. "A Multimodal Approach for Parkinson Disease Analysis". In: *Recent Advances of Neural Networks Models and Applications*. Ed. by S. Bassis. Vol. 37. Switzerland: Springer International Publishing, 2015, pp. 311–318. ISBN: 2190-3018 2190-3026. DOI: 10.1007/978-3-319-18164-6_30.

[170] Kristen Fazio. "The effects of constraint on a signature's static and dynamic features". Thesis. 2011.

[171] M. A. Ferrer, S. Chanda, M. Diaz, C. K. Banerjee, A. Majumdar, C. Carmona-Duarte, P. Acharya, and U. Pal. "Static and Dynamic Synthesis of Bengali and Devanagari Signatures". In: *IEEE Trans. Cybern.* (2017). ISSN: 2168-2275 (Electronic) 2168-2267 (Linking). DOI: 10.1109/TCYB.2017.2751740. URL: https://www.ncbi.nlm.nih.gov/pubmed/28961136.

[172] M. A. Ferrer, M. Diaz, C. Carmona-Duarte, and A. Morales. "A Behavioral Handwriting Model for Static and Dynamic Signature Synthesis". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 39.6 (2017), pp. 1041–1053. ISSN: 1939-3539 (Electronic) 0098-5589 (Linking). DOI: 10.1109/TPAMI.2016.2582167. URL: https://www.ncbi.nlm.nih.gov/pubmed/27333600.

[173] M. A. Ferrer, M. Diaz, C. Carmona-Duarte, and R. Plamondon. "A Biometric Attack Case Based on Signature Synthesis". In: *2018 International Carnahan Conference on Security Technology (ICCST)*. 2018, pp. 1–6. ISBN: 2153-0742. DOI: 10.1109/CCST.2018.8585714.

[174] M. A. Ferrer, M. Diaz, C. Carmona-Duarte, and R. Plamondon. "Generating Off-line and On-line Forgeries from On-line Genuine Signatures". In: *2019 International Carnahan Conference on Security Technology (ICCST)*. 2019, pp. 1–6. ISBN: 2153-0742. DOI: 10.1109/CCST.2019.8888418.

[175] S.E. Fienberg. *The evolving role of statistical assessments as evidence in the courts*. The evolving role of statistical assessments as evidence in the courts. New York, NY, US: Springer-Verlag Publishing, 1989, pp. xvii, 357–xvii, 357. ISBN: 0-387-96914-4 (Hardcover); 3-540-96914-4 (Hardcover). DOI: 10.1007/978-1-4612-3604-7.

[176] Stephen E. Fienberg and Joseph B. Kadane. "The Presentation of Bayesian Statistical Analyses in Legal Proceedings". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 32.1/2 (1983), pp. 88–98. ISSN: 00390526, 14679884. DOI: 10.2307/2987595. URL: www.jstor.org/stable/2987595.

[177] Stephen E. Fienberg, Samuel H. Krislov, and Miron L. Straf. "UNDERSTANDING AND EVALUATING STATISTICAL EVIDENCE IN LITIGATION". In: *Jurimetrics* 36.1 (1995), pp. 1–32. ISSN: 08971277, 21544344. URL: www.jstor.org/stable/29762401.

[178] J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos, D. T. Toledano, J. Gonzalez-Rodriguez, J. A. Siguenza, J. Garrido-Salas, E. Anguiano, G. Gonzalez-De-Rivera, R. Ribalda, M. Faundez-Zanuy, J. A. Ortega, V. Cardeñoso-Payo, A. Viloria, C. E. Vivaracho, Q. I. Moro, J. J. Igarza, J. Sanchez, I. Hernaez, C. Orrite-Uruñuela, F. Martinez-Contreras, and J. J. Gracia-Roche. "BiosecurID: a multimodal

biometric database". In: *Pattern Anal. Appl.* 13.2 (2010), pp. 235–246. ISSN: 1433-7541. DOI: 10.1007/s10044-009-0151-4. URL: https://doi.org/10.1007/s10044-009-0151-4.

[179] Julian Fierrez and Javier Ortega-Garcia. "On-Line Signature Verification". In: *Handbook of Biometrics*. Ed. by Anil K. Jain, Patrick Flynn, and Arun A. Ross. Boston, MA: Springer US, 2008, pp. 189–209. ISBN: 978-0-387-71041-9. DOI: 10.1007/978-0-387-71041-9_10.

[180] Julian Fierrez-Aguilar, Stephen Krawczyk, Javier Ortega-Garcia, and Anil K. Jain. *Fusion of Local and Regional Approaches for On-Line Signature Verification.* Conference Paper. 2005.

[181] Rainhard Findling. "Mobile Signature Recognition". Thesis. 2017.

[182] Michael O. Finkelstein and William B. Fairley. "A Bayesian Approach to Identification Evidence". In: *Harvard Law Review* 83.3 (1970), pp. 489–517. ISSN: 0017811X. DOI: 10.2307/1339656. URL: www.jstor.org/stable/1339656.

[183] Andreas Fischer and Réjean Plamondon. *A Dissimilarity Measure for On-Line Signature Verification Based on the Sigma-Lognormal Model.* Conference Paper. Pointe-à-Pitre, Guadeloupe, 2015.

[184] Andreas Fischer and Rejean Plamondon. "Signature Verification Based on the Kinematic Theory of Rapid Human Movements". In: *IEEE Trans. Human-Mach. Syst.* 47.2 (2017), pp. 169–180. ISSN: 2168-2291 2168-2305. DOI: 10.1109/thms.2016.2634922.

[185] W. Flynn. "Executing a forensic examination of electronically captured signatures". In: *Journal of the American Society of Questioned Document Examiners* 15.1 (2012), pp. 3–10.

[186] William J. Flynn, Kathleen Annunziata Nicolaides, and Linton A. Mohammed. "Chapter 12 - The examination of electronic signatures". In: *Forensic Examination of Signatures.* Ed. by Linton A. Mohammed. Academic Press, 2019, pp. 177–198. ISBN: 978-0-12-813029-2. DOI: https://doi.org/10.1016/B978-0-12-813029-2.00012-8. URL: http://www.sciencedirect.com/science/article/pii/B9780128130292000128.

[187] B. Found and C. Bird. "The Modular Forensic Handwriting Method". In: *Journal of Forensic Document Examination* 26.1 (2016), pp. 7–83.

[188] B. Found and D. Rogers. "The Forensic Investigation of Signature Complexity". In: *Handwriting and Drawing Research: Basic and Applied Issues* (1996), pp. 483–492.

[189] B. Found and D. Rogers. "The probative character of Forensic Handwriting Examiners' identification and elimination opinions on questioned signatures". In: *Forensic Sci. Int.* 178.1 (2008), pp. 54–60. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2008.02.001. URL: http://www.ncbi.nlm.nih.gov/pubmed/18367353.

[190] B. Found and D. K. Rogers. "Investigating forensic document examiners' skill relating to opinions on photocopied signatures". In: *Sci. Justice* 45.4 (2005), pp. 199–206. ISSN: 13550306. DOI: 10.1016/s1355-0306(05)71667-1.

[191] Bryan Found. "Deciphering the human condition: the rise of cognitive forensics". In: *Aust. J. Forensic Sci.* 47.4 (2014), pp. 386–401. ISSN: 0045-0618 1834-562X. DOI: 10.1080/00450618.2014.965204.

[192] Wikimedia Foundation. *Definition of 'Likelihood Ratio'*. Web Page. 2020. URL: https://en.wikipedia.org/wiki/Likelihood_function#Likelihood_ratio_and_relative_likelihood.

[193] Henry F. Fradella, Lauren O'Neill, and Adam Fogarty. "The Impact of Daubert on Forensic Science". In: *Pepp. L. Rev.* 31.2 (2004), pp. 323–362.

[194] J. Franco-Pedroso, D. Ramos, and J. Gonzalez-Rodriguez. "Gaussian Mixture Models of Between-Source Variation for Likelihood Ratio Computation from Multivariate Data". In: *PLoS One* 11.2 (2016), e0149958. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0149958. URL: https://www.ncbi.nlm.nih.gov/pubmed/26901680.

[195] K. Franke, L. R. Schomaker, C. Veenhuis, C. Taubenheim, Isabelle Guyon, Louis Vuurpijl, Merijn Van Erp, and Geertje Zwarts. "WANDA: A Generic Framework applied in Forensic Handwriting Analysis and Writer Identification". In: *Proceedings of the Third International Conference on Hybrid Intelligent Systems*. Melbourne, Australia, 2003.

[196] Katrin Franke. *Capturing Reliable Data for Computer-Based Forensic Handwriting Analysis II: Pen-position Activations*. Conference Paper. 2009. DOI: 10.1109/icdar.2009.255.

[197] Katrin Franke. "The Influence of Physical and Biomechanical Processes on the Ink Trace Methodological foundations for the forensic analysis of signatures". Ph.D. 2005.

[198] Katrin Franke and Mario Köppen. "A computer-based system to support forensic studies on handwritten documents". In: *International Journal of Document Analysis and Recognition (IJDAR)* 3 (2000), pp. 218–231.

[199] Katrin Franke and Sargur N. Srihari. "Computational Forensics: An Overview". In: *Computational Forensics*. Ed. by Sargur N. Srihari and Katrin Franke. Springer Berlin Heidelberg, 2008, pp. 1–10. ISBN: 978-3-540-85303-9.

[200] T.M. Freeman. *Cognitive Human Factors and Call Accuracy in Limited-Information Signature Identification Tasks*. Conference Paper. Washington DC, USA, 2015.

[201] L. Gaborini, A. Biedermann, and F. Taroni. "Towards a Bayesian evaluation of features in questioned handwritten signatures". In: *Sci. Justice* 57.3 (2017), pp. 209–220. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2017.01.004. URL: https://www.ncbi.nlm.nih.gov/pubmed/28454630.

[202] J. Galbally, M. Gomez-Barrero, and A. Ross. "Accuracy evaluation of handwritten signature verification: Rethinking the random-skilled forgeries dichotomy". In: *2017 IEEE International Joint Conference on Biometrics (IJCB)*. 2017, pp. 302–310. ISBN: 2474-9699. DOI: 10.1109/BTAS.2017.8272711.

[203] Javier Galbally, Moises Diaz-Cabrera, Miguel A. Ferrer, Marta Gomez-Barrero, Aythami Morales, and Julian Fierrez. "On-line signature recognition through the combination of real dynamic data and synthetically generated static data". In: *Pattern Recognit.* 48.9 (2015), pp. 2921–2934. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2015.03.019. URL: http://www.sciencedirect.com/science/article/pii/S003132031500120X.

[204] Javier Galbally, Julian Fierrez, Marcos Martinez-Diaz, and Javier Ortega-Garcia. *Evaluation of Brute-force Attack to Dynamic Signature Verification Using Synthetic Samples.* Conference Paper. Barcelona, Spain, 2009. DOI: 10.1109/icdar.2009.39.

[205] Javier Galbally, Julian Fierrez, Javier Ortega-Garcia, and Réjean Plamondon. "Synthetic on-line signature generation. Part II: Experimental validation". In: *Pattern Recognit.* 45.7 (2012), pp. 2622–2632. ISSN: 00313203. DOI: 10.1016/j.patcog.2011.12.007.

[206] Javier Galbally, Rudolf Haraksim, and Laurent Beslay. "A Study of Age and Ageing in Fingerprint Biometrics". In: *IEEE Trans. Inf. Forensics Security* 14.5 (2019), pp. 1351–1365. ISSN: 1556-6013 1556-6021. DOI: 10.1109/tifs.2018.2878160.

[207] Javier Galbally, Marcos Martinez-Diaz, and Julian Fierrez. "Aging in biometrics: an experimental analysis on on-line signature". In: *PLoS One* 8.7 (2013), e69897–e69897. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0069897. URL: https://www.ncbi.nlm.nih.gov/pubmed/23894557%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3720939/.

[208] Javier Galbally, Réjean Plamondon, Julian Fierrez, and Javier Ortega-Garcia. "Synthetic on-line signature generation. Part I: Methodology and algorithms". In: *Pattern Recognit.* 45.7 (2012), pp. 2610–2621. ISSN: 00313203. DOI: 10.1016/j.patcog.2011.12.011.

[209] G.P. van Galen and A.W.A. van Gemmert. "Kinematic and dynamic features of forging another person's handwriting". In: *Journal of Forensic Document Examination* 9.1 (1996), pp. 1–25.

[210] D. T. Gantz and C. P. Saunders. *Quantifying the Effects of Database Size and Sample Quality on Measures of Individualization Validity and Accuracy in Forensics.* Report 2009_DN_BX_K234. U.S. Department of Justice, 2015.

[211] Donald T. Gantz and Mark A. Walch. *FLASH ID Handwriting Derived Biometric Analysis Software.* Conference Paper. NIST, 2013.

[212] Sonia Garcia-Salicetti, Charles Beumier, Gérard Chollet, Bernadette Dorizzi, Jean Leroux les Jardins, Jan Lunter, Yang Ni, and Dijana Petrovska-Delacrétaz. "BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities". In: *Audio- and Video-Based Biometric Person Authentication.* Ed. by Josef Kittler and Mark S. Nixon. Springer Berlin Heidelberg, 2003, pp. 845–853. ISBN: 978-3-540-44887-7.

[213] A. W. Van Gemmert and H. L. Teulings. "Advances in graphonomics: studies on fine motor control, its development and disorders". In: *Hum Mov Sci* 25.4-5 (2006), pp. 447–53. ISSN: 0167-9457 (Print) 0167-9457 (Linking). DOI: 10.1016/j.humov.2006.07.002. URL: https://www.ncbi.nlm.nih.gov/pubmed/17027111.

[214] A.W.A. van Gemmert and G.P. van Galen. "Dynamic features of mimicking another person's writing and signature". In: *Handwriting and Drawing Research: Basic and Applied Issues*. Ed. by M.L. Simner, C.G. Leedham, and A.J.W.M. Thomassen. Amsterdam: IOS Press, 1996, pp. 459–471. ISBN: 9789051992809. URL: https://books.google.ch/books?id=WEpw1o69AW8C.

[215] A. W. Van Gemmert and G. P. Van Galen. "Stress, neuromotor noise, and human performance: a theoretical perspective". In: *J. Exp. Psychol. Hum. Percept. Perform.* 23.5 (1997), pp. 1299–313. ISSN: 0096-1523 (Print) 0096-1523.

[216] A.W.A. van Gemmert. "The effects of mental load and stress on the dynamics of fine motor tasks". Thesis. 1997.

[217] G. Gennari. "Even judges are CSI fans". In: *Forensic Sci. Int.* 292 (2018), e1–e2. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2018.10.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/30404758.

[218] S. Gerth, T. Dolk, A. Klassert, M. Fliesser, M. H. Fischer, G. Nottbusch, and J. Festman. "Adapting to the surface: A comparison of handwriting measures when writing on a tablet computer and on paper". In: *Hum Mov Sci* 48 (2016), pp. 62–73. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2016.04.006. URL: https://www.ncbi.nlm.nih.gov/pubmed/27132154.

[219] D. A. Gillies. "The Subjective Theory of Probability". In: *The British Journal for the Philosophy of Science* 23.2 (1972), pp. 138–157. ISSN: 00070882, 14643537. URL: www.jstor.org/stable/686439.

[220] Romain Giot, Christophe Rosenberger, and Bernadette Dorizzi. "Performance Evaluation of Biometric Template Update". In: *eprint arXiv:1203.1502* (2012), arXiv:1203.1502. URL: https://ui.adsabs.harvard.edu/abs/2012arXiv1203.1502G.

[221] S. N. Gittelson. "Evolving from Inferences to Decisions in the Interpretation of Scientific Evidence". Thesis. 2013.

[222] Peter W. Glynn and Donald L. Iglehart. "Importance Sampling for Stochastic Simulations". In: *Manage. Sci.* 35.11 (1989), pp. 1367–1392. ISSN: 0025-1909. DOI: 10.1287/mnsc.35.11.1367. URL: https://doi.org/10.1287/mnsc.35.11.1367.

[223] Marta Gomez-Barrero, Javier Galbally, Julian Fierrez, Javier Ortega-Garcia, and Rejean Plamondon. *Variations of handwritten signatures with time: A sigma-lognormal analysis*. Conference Paper. 2013. DOI: 10.1109/icb.2013.6612974.

[224] J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, D. Ramos-Castro, and J. Ortega-Garcia. "Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems". In: *Forensic Sci. Int.* 155.2-3 (2005), pp. 126–40. ISSN: 0379-0738 (Print) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2004.11.007. URL: http://www.ncbi.nlm.nih.gov/pubmed/16226150.

[225] IJ Good. *Weight of Evidence: A Brief Survey.* Conference Paper. Valencia, Spain, 1985.

[226] R. S. Goonetilleke, E. R. Hoffmann, and A. Luximon. "Effects of pen design on drawing and writing performance". In: *Appl. Ergon.* 40.2 (2009), pp. 292–301. ISSN: 1872-9126 (Electronic) 0003-6870 (Linking). DOI: 10.1016/j.apergo.2008.04.015. URL: https://www.ncbi.nlm.nih.gov/pubmed/18501332.

[227] E. Griechisch, M. I. Malk, and M. Liwicki. "Online Signature Analysis Based on Accelerometric and Gyroscopic Pens and Legendre Series". In: *2013 12th International Conference on Document Analysis and Recognition.* 2013, pp. 374–378. ISBN: 2379-2140. DOI: 10.1109/ICDAR.2013.82.

[228] Erika Griechisch. "Online Signature Verification and Handwriting Classification". Thesis. 2018.

[229] C.M. Grinstead and J.L. Snell. *Grinstead and Snell's Introduction to Probability.* University Press of Florida, 2009. ISBN: 9781616100469. URL: https://books.google.ch/books?id=cE6FRAAACAAJ.

[230] Quentin F. Gronau, Alexandra Sarafoglou, Dora Matzke, Alexander Ly, Udo Boehm, Maarten Marsman, David S. Leslie, Jonathan J. Forster, Eric-Jan Wagenmakers, and Helen Steingroever. "A tutorial on bridge sampling". In: *J. Math. Psychol.* 81 (2017), pp. 80–97. ISSN: 0022-2496. DOI: https://doi.org/10.1016/j.jmp.2017.09.005. URL: http://www.sciencedirect.com/science/article/pii/S0022249617300640.

[231] N. D. Guerin, F. d. B. Vidal, and B. Macchiavello. "Text-dependent User Verification of Handwritten Words and Signatures on Mobile Devices". In: *The Computer Journal* 59.9 (2016), pp. 1415–1425. ISSN: 1460-2067. DOI: 10.1093/comjnl/bxv113.

[232] R. Guest, M. Fairhurst, and Tristan A. Linnel. *Towards an Inferred Data Accuracy Assessment of Forensic Document Examination Methodologies for Signatures.* Conference Paper. Dijon, France, 2009.

[233] Richard Guest. "Age dependency in handwritten dynamic signature verification systems". In: *Pattern Recognit. Lett.* 27.10 (2006), pp. 1098–1104. ISSN: 01678655. DOI: 10.1016/j.patrec.2005.12.008.

[234] Richard Guest, Michael Fairhurst, and Tristan Linnell. *Towards an Inferred Data Accuracy Assessment of Forensic Document Examination Methodologies for Signatures.* Conference Paper. Dijon, France, 2008.

[235] J. Guilbert, D. Alamargot, and M. F. Morin. "Handwriting on a tablet screen: Role of visual and proprioceptive feedback in the control of movement by children and adults". In: *Hum Mov Sci* 65 (2019). ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2018.09.001. URL: https://www.ncbi.nlm.nih.gov/pubmed/30219272.

[236] N. Haase, F. Renkewitz, and C. Betsch. “The measurement of subjective probability: evaluating the sensitivity and accuracy of various scales”. In: *Risk Anal.* 33.10 (2013), pp. 1812–28. ISSN: 1539-6924 (Electronic) 0272-4332 (Linking). DOI: 10.1111/risa.12025. URL: https://www.ncbi.nlm.nih.gov/pubmed/23465071.

[237] Craig S. Halvorson, Javier Galbally, Julian Fierrez, Marcos Martinez-Diaz, Javier Ortega-Garcia, Šárka O. Southern, B. V. K. Vijaya Kumar, Salil Prabhakar, and Arun A. Ross. “Synthetic generation of handwritten signatures based on spectral analysis”. In: 7306 (2009), pp. 730629-730629–10. ISSN: 0277786X. DOI: 10.1117/12.817928.

[238] J. M. Hampton, P. G. Moore, and H. Thomas. “Subjective Probability and Its Measurement”. In: *Journal of the Royal Statistical Society. Series A (General)* 136.1 (1973), pp. 21–42. ISSN: 00359238. DOI: 10.2307/2344419. URL: www.jstor.org/stable/2344419.

[239] Yuko Hara and John H. Morrison. “Synaptic Correlates of Aging and Cognitive Decline”. In: (2014), pp. 301–342. DOI: 10.1016/b978-0-12-418675-0.00010-9.

[240] R. Haraksim, D. Ramos, D. Meuwly, and C. E. Berger. “Measuring coherence of computer-assisted likelihood ratio methods”. In: *Forensic Sci. Int.* 249 (2015), pp. 123–32. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2015.01.033. URL: http://www.ncbi.nlm.nih.gov/pubmed/25698513.

[241] Rudolf Haraksim, Daniel Ramos, and Didier Meuwly. “Validation of likelihood ratio methods for forensic evidence evaluation handling multimodal score distributions”. In: *IET Biom.* 6.2 (2017), pp. 61–69. ISSN: 2047-4938 2047-4946. DOI: 10.1049/iet-bmt.2015.0059.

[242] Heidi Harralson. *Developments in handwriting and signature identification in the digital age.* London: Routledge, Taylor & Francis Group, 2015. ISBN: 9781317522881 1317522885. URL: http://public.eblib.com/choice/publicfullrecord.aspx?p=1798354.

[243] Heidi Harralson. “Forensic document examination of electronically captured signatures”. In: *Digital Evidence and Electronic Signature Law Review* 9 (2012), pp. 67–73. DOI: 10.14296/deeslr.v9i0.1991.

[244] Heidi Harralson and Larry Miller. *Huber and Headrick’s Handwriting Identification: Facts and Fundamentals.* Boca Raton: CRC Press, 2018, p. 443. ISBN: 9781498751308. DOI: https://doi.org/10.4324/9781315152462.

[245] Heidi Harralson, Hans-Leo Teulings, and Larry Miller. *Temporal and spatial differences between online and offline signatures.* Conference Paper. Cancun, Mexico, 2011.

[246] Heidi H. Harralson, Elizabeth Waites, and Emily J. Will. *A Survey of Forensic Handwriting Examination Research in Response to the NAS Report.* Conference Paper. Pointe-à-Pitre, Guadeloupe, 2015.

[247] Wilson R. Harrison. *Suspect Documents.* Chicago: Nelson-Hall Publishers, 1981.

[248] Martina Hauser. “Les effets de la maladie de Parkinson sur l’ecriture manuscrite”. Thesis. 2014.

[249] Martina Hauser, Raymond Marquis, and Liv Cadola. “Les effets de la maladie de Parkinson sur l’écriture manuscrite”. In: *Revue Internationale de Criminologie et de Police Technique et Scientifique* (2016).

[250] Thomas Hecker. *Forensic Signalyzer - Collecting and analysing data of dynamic features of handwriting*. Conference Paper. San Diego, California, 2017.

[251] J. Heckeroth and C. D. Boywitt. “Examining Authenticity: An Initial Exploration of the Suitability of Handwritten Electronic Signatures”. In: *Forensic Sci. Int.* 275 (2017), pp. 144–154. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2017.02.019. URL: https://www.ncbi.nlm.nih.gov/pubmed/28364722.

[252] F. Helm, J. Munzert, and N. F. Troje. “Kinematic patterns underlying disguised movements: Spatial and temporal dissimilarity compared to genuine movement patterns”. In: *Hum Mov Sci* 54 (2017), pp. 308–319. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2017.05.010. URL: https://www.ncbi.nlm.nih.gov/pubmed/28628849.

[253] O. Henniger and U. Waldmann. *ISOIEC 19794 – Austauschformate für biometrische Daten*. Conference Paper. Darmstadt, Germany, 2005.

[254] A. B. Hepler, C. P. Saunders, L. J. Davis, and J. Buscaglia. “Score-based likelihood ratios for handwriting evidence”. In: *Forensic Sci. Int.* 219.1-3 (2012), pp. 129–40. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2011.12.009. URL: http://www.ncbi.nlm.nih.gov/pubmed/22297142.

[255] J. Hermsdorfer, C. Marquardt, A. S. Schneider, W. Furholzer, and B. Baur. “Significance of finger forces and kinematics during handwriting in writer’s cramp”. In: *Human Movement Science* 30.4 (2011), pp. 807–17. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2010.04.004. URL: http://www.ncbi.nlm.nih.gov/pubmed/20594605.

[256] C. E. van Den Heuvel, G. P. van Galen, H. L. Teulings, and A. W. van Gemmert. “Axial pen force increases with processing demands in handwriting”. In: *Acta Psychol (Amst)* 100.1-2 (1998), pp. 145–59. ISSN: 0001-6918 (Print) 0001-6918.

[257] T. Hicks, A. Biedermann, J. A. de Koeijer, F. Taroni, C. Champod, and I. W. Evett. “Reply to Morrison et al. (2016) Refining the relevant population in forensic voice comparison - A response to Hicks et al (2015) The importance of distinguishing information from evidence/observations when formulating propositions”. In: *Sci. Justice* 57.5 (2017), pp. 401–402. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2017.04.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/28889871.

[258] T. Hicks, A. Biedermann, J. A. de Koeijer, F. Taroni, C. Champod, and I. W. Evett. “The importance of distinguishing information from evidence/observations when formulating propositions”. In: *Sci. Justice* 55.6 (2015), pp. 520–5. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2015.06.008. URL: https://www.ncbi.nlm.nih.gov/pubmed/26654089.

[259] Ordway Hilton. *The Scientific Examination of Questioned Documents (revised edition)*. Forensic and Police Science. Boca Raton, London, New York, Washington D.C.: CRC Press, 1993.

[260] Christian Hook, Juergen Kempf, and Georg Scharfenberg. *New pen device for biometrical 3D pressure analysis of handwritten characters, words and signatures*. Conference Paper. Berkley, California, 2003. DOI: 10.1145/982507.982515.

[261] N. Houmani, S. Garcia-Salicetti, and B. Dorizzi. "On assessing the robustness of pen coordinates, pen pressure and pen inclination to time variability with personal entropy". In: *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*. 2009, pp. 1–6. DOI: 10.1109/BTAS.2009.5339074.

[262] N. Houmani, A. Mayoue, S. Garcia-Salicetti, B. Dorizzi, M. I. Khalil, M. N. Moustafa, H. Abbas, D. Muramatsu, B. Yanikoglu, A. Kholmatov, M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia, J. Roure Alcobé, J. Fabregas, M. Faundez-Zanuy, J. M. Pascual-Gaspar, V. Cardeñoso-Payo, and C. Vivaracho-Pascual. "BioSecure signature evaluation campaign (BSEC'2009): Evaluating online signature algorithms depending on the quality of signatures". In: *Pattern Recognit.* 45.3 (2012), pp. 993–1003. ISSN: 00313203. DOI: 10.1016/j.patcog.2011.08.008.

[263] Nesma Houmani and Sonia Garcia-Salicetti. "Quality Measures for Online Handwritten Signatures". In: 292 (2014), pp. 255–283. ISSN: 1876-1100 1876-1119. DOI: 10.1007/978-3-642-54080-6_10.

[264] Nesma Houmani, Sonia Garcia-Salicetti, and Bernadette Dorizzi. "On measuring forgery quality in online signatures". In: *Pattern Recognit.* 45.3 (2012), pp. 1004–1018. ISSN: 00313203. DOI: 10.1016/j.patcog.2011.08.019.

[265] Nesma Houmani, Sonia Garcia-Salicetti, Bernadette Dorizzi, and Mounim El-Yacoubi. "On-line Signature Verification on a Mobile Platform". In: *Mobile Computing, Applications, and Services*. Ed. by Martin Gris and Guang Yang. Springer Berlin Heidelberg, 2010, pp. 396–400. ISBN: 978-3-642-29336-8.

[266] L. M. Howes. "The communication of forensic science in the criminal justice system: A review of theory and proposed directions for research". In: *Sci. Justice* 55.2 (2015), pp. 145–54. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2014.11.002. URL: http://www.ncbi.nlm.nih.gov/pubmed/25754001.

[267] L. M. Howes, K. P. Kirkbride, S. F. Kelty, R. Julian, and N. Kemp. "Forensic scientists' conclusions: how readable are they for non-scientist report-users?" In: *Forensic Sci. Int.* 231.1-3 (2013), pp. 102–12. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2013.04.026. URL: https://www.ncbi.nlm.nih.gov/pubmed/23890623.

[268] RA Huber and AM Headrick. *Handwriting Identification: Facts and Fundamentals*. Boca Raton: CRC Press, 1999.

[269] Expert Working Group for Human Factors in Handwriting Examination. *Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach*. Report. U.S. Department of Commerce National Institute of Standards and Technology, 2020.

[270] ID Wholesaler LLC. *What is a signature capture pad?* Web Page. 2020. URL: `https://www.idwholesaler.com/learning-center/what-is-a-signature-capture-pad/`.

[271] D. Impedovo, G. Pirlo, and R. Plamondon. "Handwritten Signature Verification: New Advancements and Open Issues". In: (2012), pp. 367–372. DOI: `10.1109/icfhr.2012.211`.

[272] Donato Impedovo and Giuseppe Pirlo. "Automatic signature verification in the mobile cloud scenario: survey and way ahead". In: *IEEE Trans. Emerg. Topics Comput.* (2018), pp. 1–1. ISSN: 2168-6750 2376-4562. DOI: `10.1109/tetc.2018.2865345`.

[273] Donato Impedovo and Giuseppe Pirlo. "Automatic Signature Verification: The State of the Art". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 38.5 (2008), pp. 609–635. ISSN: 1094-6977. DOI: `10.1109/tsmcc.2008.923866`.

[274] Donato Impedovo, Giuseppe Pirlo, and Fabrizio Rizzi. *Characteristics of Constrained Handwritten Signatures: An Experimental Investigation*. Conference Paper. Pointe-la-Pitre, Guadeloupe, 2015.

[275] Topaz Systems Incorporated. *Topaz Systems Incorporated Case Studies*. Web Page. 2020. URL: `https://www.topazsystems.com/casestudies.html`.

[276] K. Inman and N. Rudin. *Principles and Practice of Criminalistics: The Profession of Forensic Science*. CRC Press, 2000. ISBN: 9781420036930. URL: `https://books.google.ch/books?id=6OTqqqGooccC`.

[277] International Organization for Standardization. *ISO 30107-3:2017 Standard - Information technology — Biometric presentation attack detection — Part 3: Testing and reporting*. Standard. 2017. URL: `https://www.iso.org/standard/67381.html`.

[278] International Organization for Standardization. *ISO/IEC 19794-11:2013 Information technology — Biometric data interchange formats — Part 11: Signature/sign processed dynamic data*. Standard. 2013. URL: `https://www.iso.org/standard/51824.html`.

[279] International Organization for Standardization. *ISO/IEC 19794-7:2014 Information technology – Biometric data interchange formats – Part 7: Signature/sign time series data*. Standard. 2014. URL: `https://www.iso.org/standard/55938.html`.

[280] International Organization for Standardization. *ISO/IEC TR 19795-3:2007 Information technology — Biometric performance testing and reporting — Part 3: Modality-specific testing*. Standard. 2007. URL: `https://www.iso.org/standard/41449.html`.

[281] G. Jackson, S. Jones, G. Booth, C. Champod, and I. W. Evett. "The nature of forensic science opinion—a possible framework to guide thinking and practice in investigation and in court proceedings". In: *Sci. Justice* 46.1 (2006), pp. 33–44. ISSN: 13550306. DOI: `10.1016/s1355-0306(06)71565-9`.

[282] Graham Jackson. *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses.* 2010.

[283] Graham Jackson, Colin Aitken, and Paul Roberts. *Case Assessment and Interpretation of Expert Evidence.* Vol. 4. Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses. Royal Statistical Society, 2015, p. 145.

[284] Graham Jackson, David Kaye, Cedric Neumann, Anjali Ranadive, and Valerie Reyna. *Communicating the Results of Forensic Science Examinations.* Report. National Institute for Science and Technology (NIST), 2015.

[285] M. Jacquet and C. Champod. "Automated face recognition in forensic science: Review and perspectives". In: *Forensic Sci. Int.* 307 (2019), p. 110124. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2019.110124. URL: https://www.ncbi.nlm.nih.gov/pubmed/31927397.

[286] A. Jain, P. Flynn, and A.A. Ross. *Handbook of Biometrics.* Springer US, 2010. ISBN: 9781441943750. URL: https://books.google.ch/books?id=BsawcQAACAAJ.

[287] A. K. Jain and A. Ross. "Bridging the gap: from biometrics to forensics". In: *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370.1674 (2015). ISSN: 1471-2970 (Electronic) 0962-8436 (Linking). DOI: 10.1098/rstb.2014.0254. URL: https://www.ncbi.nlm.nih.gov/pubmed/26101280.

[288] Anil Jain, Umut Uludag, and Arun Ross. *Biometric Template Selection: A Case Study in Fingerprints.* Vol. 2688. 2003, pp. 1061–1061. DOI: 10.1007/3-540-44887-X_40.

[289] Anil K. Jain, Friederike D. Griess, and Scott D. Connell. "On-line signature verification". In: *Pattern Recognit.* 35.12 (2002), pp. 2963–2972. ISSN: 0031-3203. DOI: https://doi.org/10.1016/S0031-3203(01)00240-0. URL: http://www.sciencedirect.com/science/article/pii/S0031320301002400.

[290] Anil K. Jain, Arun A. Ross, and Karthik Nandakumar. "Introduction". In: *Introduction to Biometrics.* Ed. by Anil K. Jain, Arun A. Ross, and Karthik Nandakumar. Boston, MA: Springer US, 2011, pp. 1–49. ISBN: 978-0-387-77326-1. DOI: 10.1007/978-0-387-77326-1_1. URL: https://doi.org/10.1007/978-0-387-77326-1_1.

[291] I. Jasper, A. Haussler, B. Baur, C. Marquardt, and J. Hermsdorfer. "Circadian variations in the kinematics of handwriting and grip strength". In: *Chronobiol. Int.* 26.3 (2009), pp. 576–94. ISSN: 1525-6073 (Electronic) 0742-0528 (Linking). DOI: 10.1080/07420520902896590. URL: https://www.ncbi.nlm.nih.gov/pubmed/19360497.

[292] ISABELLE JASPER, ANDREAS HÄUßLER, CHRISTIAN MARQUARDT, and JOACHIM HERMSDÖRFER. "Circadian rhythm in handwriting". In: *J. Sleep Res.* 18.2 (2009), pp. 264–271. DOI: doi:10.1111/j.1365-2869.2008.00727.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2869.2008.00727.x.

[293] M. E. Johnson, T. W. Vastrick, M. Boulanger, and E. Schuetzner. "Measuring the Frequency Occurrence of Handwriting and Handprinting Characteristics". In: *J. Forensic Sci.* 62.1 (2017), pp. 142–163. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.13248. URL: https://www.ncbi.nlm.nih.gov/pubmed/27864959.

[294] C. Kahindo, S. Garcia-Salicetti, and N. Houmani. “A Signature Complexity Measure to Select Reference Signatures for Online Signature Verification”. In: *2015 International Conference of the Biometrics Special Interest Group (BIOSIG)*. 2015, pp. 1–8. DOI: 10.1109/BIOSIG.2015.7314595.

[295] D. Kahneman. *Thinking, Fast and Slow*. Penguin Books Limited, 2011. ISBN: 9780141918921. URL: https://books.google.ch/books?id=oV1tXT3HigoC.

[296] M. Kam, G. Fielding, and R. Conn. “Writer Identification by Professional Document Examiners”. In: *J. Forensic Sci.* 42.5 (1997), pp. 778–786. ISSN: 0022-1198. DOI: 10.1520/JFS14207J. URL: https://www.astm.org/DIGITAL_LIBRARY/JOURNALS/FORENSIC/PAGES/JFS14207J.htm.

[297] M. Kam, K. Gummadidala, G. Fielding, and R. Conn. “Signature authentication by forensic document examiners”. In: *J. Forensic Sci.* 46.4 (2001), pp. 884–8. ISSN: 0022-1198 (Print) 0022-1198.

[298] David Kaye. “Identification, Individualization and Uniqueness: What’s the Difference?” In: *Law, Probability and Risk* 8.2 (2009). DOI: 10.1093/lpr/mgp018.

[299] David Kaye. “Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies”. In: *Brooklyn Law Rev.* 75 (2009).

[300] Jan Seaman Kelly and Brian S. Lindblom. *Scientific Examination of Questioned Documents, Second Edition*. Forensic and Police Science Series. CRC Press, 2006. ISBN: 978-0-8493-2044-6. DOI: doi:10.1201/9781420003765.fmatt10.1201/9781420003765.fmatt. URL: http://dx.doi.org/10.1201/9781420003765.fmatt.

[301] Eamonn Keogh and Abdullah Mueen. “Curse of Dimensionality”. In: *Encyclopedia of Machine Learning and Data Mining*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2017, pp. 314–315. ISBN: 978-1-4899-7687-1. DOI: 10.1007/978-1-4899-7687-1_192. URL: https://doi.org/10.1007/978-1-4899-7687-1_192.

[302] Hamed Ketabdar, Jonas Richiardi, and Andrzej Drygajlo. *Global feature selection for on-line signature verification*. Conference Paper. Salerno, Italy, 2005.

[303] Alisher Kholmatov and Berrin Yanikoglu. “Identity authentication using improved online signature verification method”. In: *Pattern Recognit. Lett.* 26.15 (2005), pp. 2400–2408. ISSN: 01678655. DOI: 10.1016/j.patrec.2005.04.017.

[304] Alisher Kholmatov and Berrin Yanikoglu. “SUSIG: an on-line signature database, associated protocols and benchmark results”. In: *Pattern Analysis and Applications* 12.3 (2009), pp. 227–236. ISSN: 1433-755X. DOI: 10.1007/s10044-008-0118-x. URL: https://doi.org/10.1007/s10044-008-0118-x.

[305] Charles R. Kingston. “Applications of Probability Theory in Criminalistics”. In: *J. Am. Stat. Assoc.* 60.309 (1965), pp. 70–80. ISSN: 0162-1459. DOI: 10.1080/01621459.1965.10480776. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1965.10480776.

[306] Charles R. Kingston. “Applications of Probability Theory in Criminalistics—II”. In: *J. Am. Stat. Assoc.* 60.312 (1965), pp. 1028–1034. ISSN: 0162-1459. DOI: 10.1080/01621459.1965.10480847. URL: https://doi.org/10.1080/01621459.1965.10480847.

[307] Charles R. Kingston. “Probability and Legal Proceedings”. In: *Journal of Criminal Law and Criminology* 57.1 (1966), pp. 93–98.

[308] Charles R. Kingston and P. L. Kirk. “The Use of Statistics in Criminalistics”. In: *Journal of Criminal Law and Criminology* 54.4 (1964), pp. 514–521.

[309] Paul L. Kirk. “Criminalistics”. In: *Science* 140.3565 (1963), pp. 367–370. ISSN: 00368075, 10959203. URL: www.jstor.org/stable/1710916.

[310] Paul L. Kirk. “The Ontogeny of Criminalistics”. In: *The Journal of Criminal Law Criminology and Police Science* 54.2 (1963), p. 235.

[311] Paul L. Kirk. “The Ontogeny of Criminalistics”. In: *The Journal of Criminal Law, Criminology, and Police Science* 54.2 (1963), pp. 235–238. ISSN: 00220205. URL: http://www.jstor.org/stable/1141173.

[312] K.M. Koppenhaver. *Forensic Document Examination: Principles and Practice*. Humana Press, 2007. ISBN: 9781597453011. URL: https://books.google.ch/books?id=WBww6-HjCksC.

[313] Katherine Mainolfi Koppenhaver. “Detecting Fraudulent Documents”. In: *Forensic Document Examination: Principles and Practice*. Ed. by Katherine Mainolfi Koppenhaver. Totowa, NJ: Humana Press, 2007, pp. 113–131. ISBN: 978-1-59745-301-1. DOI: 10.1007/978-1-59745-301-1_13. URL: https://doi.org/10.1007/978-1-59745-301-1_13.

[314] Ram P. Krish, Julian Fierrez, Javier Galbally, and Marcos Martinez-Diaz. “Dynamic Signature Verification on Smart Phones”. In: *Highlights on Practical Applications of Agents and Multi-Agent Systems*. Ed. by Juan M. Corchado, Javier Bajo, Jaroslaw Kozlak, Pawel Pawlewski, Jose M. Molina, Vicente Julian, Ricardo Azambuja Silveira, Rainer Unland, and Sylvain Giroux. Springer Berlin Heidelberg, 2013, pp. 213–222. ISBN: 978-3-642-38061-7.

[315] K. Kryszczuk and A. Drygajlo. “Improving biometric verification with class-independent quality information”. In: *IET Signal Proc.* 3.4 (2009), p. 310. ISSN: 17519675. DOI: 10.1049/iet-spr.2008.0174.

[316] Sergey Kulik and Denis Nikonets. “Forensic handwriting examination and human factors: improving the practice through automation and expert training”. In: *2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC)*. IEEE. 2016, pp. 221–226.

[317] Erich Kupferschmid. “Statistical Analysis of Electronic Signatures using Dynamic Time Warping: What is the Value of Simple Signature Elements?” Thesis. 2017.

[318] Quon Yin Kwan. “Inference of Identity of Source”. Thesis. 1977.

[319] Francesco Lacquaniti, Carlo Terzuolo, and Paolo Viviani. "The law relating the kinematic and figural aspects of drawing movements". In: *Acta Psychol. (Amst.)* 54.1 (1983), pp. 115–130. ISSN: 0001-6918. DOI: https://doi.org/10.1016/0001-6918(83)90027-6. URL: http://www.sciencedirect.com/science/article/pii/0001691883900276.

[320] G. Langenburg. "An Online Signature Verification System for Forgery and Disguise Detection". Ph.D. 2012.

[321] G. Langenburg. "Critical Analysis of ACE-V". Thesis. 2012.

[322] A. Lanitis and N. Tsapatsoulis. "Quantitative evaluation of the effects of aging on biometric templates". In: *IET Comput. Vision* 5.6 (2011). ISSN: 17519632. DOI: 10.1049/iet-cvi.2010.0197.

[323] Andreas Lanitis. "A survey of the effects of aging on biometric identity verification". In: *International Journal of Biometrics* 2.1 (2010). ISSN: 1755-8301 1755-831X. DOI: 10.1504/ijbm.2010.030415.

[324] Andreas Lanitis and Nicolas Tsapatsoulis. "On the Quantification of Aging Effects on Biometric Features". In: *Artificial Intelligence Applications and Innovations*. Ed. by Harris Papadopoulos, Andreas S. Andreou, and Max Bramer. Springer Berlin Heidelberg, 2010, pp. 360–367. ISBN: 978-3-642-16239-8.

[325] Andreas Lanitis, Nicolas Tsapatsoulis, and Anastasios Maronidis. "Review of ageing with respect to biometrics and diverse modalities". In: *Age Factors in Biometric Processing*. Ed. by M. Fairhurst. Vol. 1. Security. Institution of Engineering and Technology, 2013, pp. 17–36. DOI: doi:10.1049/PBSP010E_ch2. URL: https://digital-library.theiet.org/content/books/10.1049/pbsp010e_ch2.

[326] Mark L. Latash. *Control of human movement*. Chicago, IL Human Kinetics Publishers, 1993. ISBN: 0873224558.

[327] Claire R. LaVelle. "What Forensic Document Examiners should know about Digital and Electronic Signatures". In: *Journal of Forensic Document Examination* 18.1 (2007).

[328] Franck Leclerc and Réjean Plamondon. "Automatic Signature Verification: The State of the Art - 1989-1993". In: *IJPRAI* 8 (1994), pp. 643–660. DOI: 10.1142/S0218001494000346.

[329] Hansheng Lei and Venu Govindaraju. "A comparative study on the consistency of features in on-line signature verification". In: *Pattern Recognit. Lett.* 26.15 (2005), pp. 2483–2489. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2005.05.005. URL: http://www.sciencedirect.com/science/article/pii/S0167865505001534.

[330] Joerg- M. Lenz. "Dynamic Signature Recognition : A sign of the times". In: *Biom. Technol. Today* 17.2 (2008), pp. 10–11. DOI: https://doi.org/10.1016/S0969-4765(09)70058-X.

[331] Joerg- M. Lenz. "Dynamic Signature Recognition: A sign of the times – Part 2". In: *Biom. Technol. Today* (2009).

[332] Joerg- M. Lenz. "Putting security back into signatures". In: *Biom. Technol. Today* 13.5 (2005), pp. 7–8. ISSN: 09694765. DOI: 10.1016/s0969-4765(05)70327-1.

[333] Jörg Lenz. "Taking dynamic signatures seriously". In: *Biom. Technol. Today* 2011.10 (2011), pp. 9–11. ISSN: 09694765. DOI: 10.1016/s0969-4765(12)70016-4.

[334] J. Lewis. *Forensic Document Examination: Fundamentals and Current Trends*. Elsevier Science, 2014. ISBN: 9780124104198. URL: https://books.google.ch/books?id=P0Z1AgAAQBAJ.

[335] J. Linden, R. Marquis, S. Bozza, and F. Taroni. "Dynamic signatures: A review of dynamic feature variation and forensic methodology". In: *Forensic Sci. Int.* 291 (2018), pp. 216–229. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2018.08.021. URL: https://www.ncbi.nlm.nih.gov/pubmed/30241096.

[336] Jacques Linden. "Les signatures électroniques peuvent-elles être l'objet d'une expertise de signatures". Thesis. 2015.

[337] Jacques Linden, Raymond Marquis, and Williams D. Mazzella. "Forensic Analysis of Digital Dynamic Signatures: New Methods for Data Treatment and Feature Evaluation". In: *J. Forensic Sci.* 62.2 (2017), pp. 382–391. DOI: 10.1111/1556-4029.13288.

[338] Jacques Linden, F. Taroni, R. Marquis, and S. Bozza. "Bayesian multivariate models for case assessment in dynamic signature cases." In: *Forensic science international* 318 (2020), p. 110611. DOI: https://doi.org/10.1016/j.forsciint.2020.110611.

[339] D. V. Lindley. "A Problem in Forensic Science". In: *Biometrika* 64.2 (1977), pp. 207–213. ISSN: 00063444. DOI: 10.2307/2335686. URL: www.jstor.org/stable/2335686.

[340] D. V. Lindley. "Probability and the Law". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 26.3 (1977), pp. 203–220. ISSN: 00390526, 14679884. DOI: 10.2307/2987898. URL: www.jstor.org/stable/2987898.

[341] D.V. Lindley. *Understanding Uncertainty*. Wiley, 2006. ISBN: 9780470055472. URL: https://books.google.ch/books?id=z0ArJ_CDnssC.

[342] D.V. Lindley. *Understanding Uncertainty*. Wiley Series in Probability and Statistics. Wiley, 2013. ISBN: 9781118650127. URL: https://books.google.ch/books?id=RhgfAgAAQBAJ.

[343] Dennis V. Lindley. "The Philosophy of Statistics". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 49.3 (2000), pp. 293–337. ISSN: 00390526, 14679884. URL: http://www.jstor.org/stable/2681060.

[344] Marcus Liwicki. "Automatic Signature Verification : In-depth investigation of novel features and different models". In: *Journal of Forensic Document Examination* 22 (2012), pp. 25–39.

[345] Marcus Liwicki, C. Elisa van den Heuvel, Bryan Found, and Muhammad Imran Malik. *Forensic Signature Verification Competition 4NSigComp2010 - Detection of Simulated and Disguised Signatures*. Conference Paper. Kolkata, India, 2010. DOI: 10.1109/icfhr.2010.116.

[346] Marcus Liwicki, Muhammad Imran Malik, Linda Alewijnse, Elisa van den Heuvel, and Bryan Found. "ICFHR 2012 Competition on Automatic Forensic Signature Verification (4NsigComp 2012)". In: (2012), pp. 823–828. DOI: 10.1109/icfhr.2012.217.

[347] Marcus Liwicki, Muhammad Imran Malik, C. Elisa van den Heuvel, Xiaohong Chen, Charles Berger, Reinoud Stoel, Michael Blumenstein, and Bryan Found. "Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011)". In: (2011), pp. 1480–1484. DOI: 10.1109/icdar.2011.294.

[348] J. B. Lohr, T. May, and M. P. Caligiuri. "Quantitative assessment of motor abnormalities in untreated patients with major depressive disorder". In: *J. Affect. Disord.* 146.1 (2013), pp. 84–90. ISSN: 1573-2517 (Electronic) 0165-0327 (Linking). DOI: 10.1016/j.jad.2012.08.043. URL: http://www.ncbi.nlm.nih.gov/pubmed/22985485.

[349] Xinyi Lu, Yuxun Fang, Wenxiong Kang, Zhiyong Wang, and David Dagan Feng. "SCUT-MMSIG: A Multimodal Online Signature Database". In: *Biometric Recognition.* Ed. by Jie Zhou, Yunhong Wang, Zhenan Sun, Yong Xu, Linlin Shen, Jianjiang Feng, Shiguang Shan, Yu Qiao, Zhenhua Guo, and Shiqi Yu. Springer International Publishing, 2017, pp. 729–738. ISBN: 978-3-319-69923-3.

[350] M. I. Malik, M. Liwicki, and A. Dengel. *Part-Based Automatic System in Comparison to Human Experts for Forensic Signature Verification.* Conference Paper. 2013. DOI: 10.1109/ICDAR.2013.178.

[351] Muhammad Imran Malik. "Automatic Signature Verification Bridging the Gap between Existing Pattern Recognition Methods and Forensic Science". Thesis. 2015.

[352] Muhammad Imran Malik, Sheraz Ahmed, Marcus Liwicki, and Andreas Dengel. "FREAK for Real Time Forensic Signature Verification". In: (2013), pp. 971–975. DOI: 10.1109/icdar.2013.196.

[353] Muhammad Imran Malik, Sheraz Ahmed, Angelo Marcelli, Umapada Pal, Michael Blumenstein, Linda Alewijnse, and Marcus Liwicki. *ICDAR2015 competition on signature verification and writer identification for on- and off-line skilled forgeries (SigWIcomp2015).* Conference Paper. Nancy, France, 2015. DOI: 10.1109/icdar.2015.7333948.

[354] Muhammad Imran Malik, Marcus Liwicki, Linda Alewijnse, Wataru Ohyama, Michael Blumenstein, and Bryan Found. "ICDAR 2013 Competitions on Signature Verification and Writer Identification for On- and Offline Skilled Forgeries (SigWiComp 2013)". In: (2013), pp. 1477–1483. DOI: 10.1109/icdar.2013.220.

[355] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. "Part-Based Automatic System in Comparison to Human Experts for Forensic Signature Verification". In: (2013), pp. 872–876. DOI: 10.1109/icdar.2013.178.

[356] Anne Mangen and Jean-Luc Velay. "Digitizing Literacy: Reflections on the Haptics of Writing". In: M. H. Zadeh. *Advances in Haptics.* Ed. by M. H. Zadeh. 20. Intech Open, 2010. DOI: 10.5772/8710. URL: https://doi.org/10.5772/8710.

[357] Pierre Margot. "Forensic science on trial - What is the law of the land?" In: *Aust. J. Forensic Sci.* 43.2-3 (2011), pp. 89–103. ISSN: 0045-0618 1834-562X. DOI: 10.1080/00450618.2011.555418.

[358] R. Marquis. "Etude de caractères manuscrits : de la caractérisation morphologique à l'individualisation du scripteur". Thesis. 2007.

[359] R. Marquis, A. Biedermann, L. Cadola, C. Champod, L. Gueissaz, G. Massonnet, W. D. Mazzella, F. Taroni, and T. Hicks. "Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings". In: *Sci. Justice* 56.5 (2016), pp. 364–370. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2016.05.009. URL: https://www.ncbi.nlm.nih.gov/pubmed/27702452.

[360] R. Marquis, S. Bozza, M. Schmittbuhl, and F. Taroni. "Handwriting evidence evaluation based on the shape of characters: application of multivariate likelihood ratios". In: *J. Forensic Sci.* 56 Suppl 1 (2011), S238–42. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/j.1556-4029.2010.01602.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/21198625.

[361] R. Marquis, L. Cadola, W. D. Mazzella, and T. Hicks. "What is the error margin of your signature analysis?" In: *Forensic Sci. Int.* 281 (2017), e1–e8. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2017.11.012. URL: https://www.ncbi.nlm.nih.gov/pubmed/29153481.

[362] R. Marquis, M. Schmittbuhl, W. D. Mazzella, and F. Taroni. "Quantification of the shape of handwritten characters: a step to objective discrimination between writers based on the study of the capital character O". In: *Forensic Sci. Int.* 150.1 (2005), pp. 23–32. ISSN: 0379-0738 (Print) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2004.06.028. URL: https://www.ncbi.nlm.nih.gov/pubmed/15837006.

[363] R. Marquis, F. Taroni, S. Bozza, and M. Schmittbuhl. "Quantitative characterization of morphological polymorphism of handwritten characters loops". In: *Forensic Sci. Int.* 164.2-3 (2006), pp. 211–20. ISSN: 0379-0738 (Print) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2006.02.008. URL: http://www.ncbi.nlm.nih.gov/pubmed/16524677.

[364] M. Martinez-Diaz. "Dynamic signature verification for portable devices". Thesis. 2008.

[365] M. Martinez-Diaz, J. Fierrez, and J. Galbally. "The DooDB Graphical Password Database: Data Analysis and Benchmark Results". In: *IEEE Access* 1 (2013), pp. 596–605. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2013.2281773.

[366] M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia. "Towards mobile authentication using dynamic signature verification: Useful features and performance evaluation". In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on.* 2008, pp. 1–5. ISBN: 1051-4651. DOI: 10.1109/ICPR.2008.4761849.

[367] Marcos Martinez-Diaz, Javier Galbally, Ram P. Krish, and Julian Fierrez. "Mobile signature verification: feature robustness and performance comparison". In: *IET Biom.* 3.4 (2014), pp. 267–277. ISSN: 2047-4938 2047-4946. DOI: 10.1049/iet-bmt.2013.0081.

[368] K. A. Martire, G. Edmond, D. J. Navarro, and B. R. Newell. "On the likelihood of "encapsulating all uncertainty"". In: *Sci. Justice* 57.1 (2017), pp. 76–79. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2016.10.004. URL: https://www.ncbi.nlm.nih.gov/pubmed/28063591.

[369] K. A. Martire, R. I. Kemp, M. Sayle, and B. R. Newell. "On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect". In: *Forensic Sci. Int.* 240 (2014), pp. 61–8. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2014.04.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/24814330.

[370] K. A. Martire and I. Watkins. "Perception problems of the verbal scale: A reanalysis and application of a membership function approach". In: *Sci. Justice* 55.4 (2015), pp. 264–73. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2015.01.002. URL: https://www.ncbi.nlm.nih.gov/pubmed/26087874.

[371] Kristy A. Martire, Richard I. Kemp, and Ben R. Newell. "The psychology of interpreting expert evaluative opinions". In: *Aust. J. Forensic Sci.* 45.3 (2013), pp. 305–314. ISSN: 0045-0618. DOI: 10.1080/00450618.2013.784361. URL: https://doi.org/10.1080/00450618.2013.784361.

[372] Stephen Mason and Daniel Seng. *Electronic Evidence: Fourth Edition*. UNIVERSITY OF LONDON, 2017, p. 427. ISBN: 978 1 911507 07 9. DOI: 10.14296/517.9781911507079.

[373] Masaki Matsubara. *Position Detecting Device*. Ed. by The United States Patent and Trademark Office (USPTO). Patent. Wacom Co. Ltd., 2012.

[374] Daniela Mazzolini, Patrizia Pavan, Giuseppe Pirlo, and Gennaro Vessio. "Towards a Decision Support Framework for Forensic Analysis of Dynamic Signatures". In: *Digital Libraries: The Era of Big Data and Data Science*. Ed. by Michelangelo Ceci, Stefano Ferilli, and Antonella Poggi. Springer International Publishing, 2020, pp. 9–14. ISBN: 978-3-030-39905-4.

[375] Roland Mergl, Peter Tigges, Andreas Schröter, Hans-Jürgen Möller, and Ulrich Hegerl. "Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives". In: *J. Neurosci. Methods* 90.2 (1999), pp. 157–169. ISSN: 0165-0270. DOI: https://doi.org/10.1016/S0165-0270(99)00080-1. URL: http://www.sciencedirect.com/science/article/pii/S0165027099000801.

[376] Mara Merlino. *Validity, Reliability, Accuracy, and Bias in Forensic Signature Identification*. Report 2010-DN-BX-K271. U.S. Department of Justice, 2015.

[377] Merriam-Webster Incorporated. *Definition of 'Bayesian'*. Web Page. 2020. URL: https://www.merriam-webster.com/dictionary/Bayesian.

[378] Merriam-Webster Incorporated. *Definition of signature*. Web Page. 2020. URL: https://www.merriam-webster.com/dictionary/signature.

[379] D. Meuwly. "Forensic Individualisation from Biometric Data". In: *Sci. Justice* 46.4 (2006), pp. 205–213. ISSN: 13550306. DOI: 10.1016/s1355-0306(06)71600-8.

[380] D. Meuwly, D. Ramos, and R. Haraksim. "A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation". In: *Forensic Sci. Int.* 276 (2017), pp. 142–153. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2016.03.048. URL: https://www.ncbi.nlm.nih.gov/pubmed/27181589.

[381] D. Meuwly and R. Veldhuis. “Forensic biometrics: From two communities to one discipline”. In: *2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)*. 2012, pp. 1–12. ISBN: 1617-5468.

[382] L. Michel. *Gerichtliche Schriftvergleichung: Eine Einführung in Grundlagen, Methoden und Praxis*. De Gruyter, 2012, p. 309. ISBN: 9783110892413. URL: https://books.google.ch/books?id=bfAgAAAAQBAJ.

[383] Evangelia Micheli-Tzanakou and Konstantinos N. Plataniotis. “Biometrics: Terms and Definitions”. In: *Encyclopedia of Cryptography and Security*. Ed. by Henk C. A. van Tilborg and Sushil Jajodia. Boston, MA: Springer US, 2011, pp. 142–147. ISBN: 978-1-4419-5906-5. DOI: 10.1007/978-1-4419-5906-5_730. URL: https://doi.org/10.1007/978-1-4419-5906-5_730.

[384] J. J. Miller, R. B. Patterson, D. T. Gantz, C. P. Saunders, M. A. Walch, and J. Buscaglia. “A Set of Handwriting Features for Use in Automated Writer Identification”. In: *J. Forensic Sci.* 62.3 (2017), pp. 722–734. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.13345. URL: https://www.ncbi.nlm.nih.gov/pubmed/28054339.

[385] Jennifer L Mnookin. “Scripting expertise: The history of handwriting identification evidence and the judicial construction of reliability”. In: *Virginia Law Review* (2001), pp. 1723–1845. ISSN: 0042-6601.

[386] Jennifer L Mnookin, Max Houck, Keith Inman, David H. Kaye, Jonathan J. Koehler, Glenn Langenburg, D Michael Risinger, Norah Rudin, Jay Siegel, and David A. Stoney. “The Need for a Research Culture in the Forensic Sciences”. In: *UCLA Law Review* 58.1 (2011), pp. 725–779.

[387] Andre Moenssens. “Handwriting Identification Evidence in the Post-Daubert World”. In: *UMKC L. Rev.* 66 (1997), p. 251.

[388] L. Mohammed. “Elucidating Dynamic Features to Discriminate between Signature Disguise and Signature Forgery Behavior”. Thesis. 2011.

[389] L. Mohammed, B. Found, M. Caligiuri, and D. Rogers. “Dynamic characteristics of signatures: effects of writer style on genuine and simulated signatures”. In: *J. Forensic Sci.* 60.1 (2015), pp. 89–94. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.12605. URL: http://www.ncbi.nlm.nih.gov/pubmed/25420668.

[390] L. A. Mohammed, B. Found, M. Caligiuri, and D. Rogers. “The dynamic character of disguise behavior for text-based, mixed, and stylized signatures”. In: *J. Forensic Sci.* 56 Suppl 1 (2011), S136–41. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/j.1556-4029.2010.01584.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/20950317.

[391] Linton A. Mohammed. “Chapter 9 - Validity and reliability”. In: *Forensic Examination of Signatures*. Ed. by Linton A. Mohammed. Academic Press, 2019, pp. 129–139. ISBN: 978-0-12-813029-2. DOI: https://doi.org/10.1016/B978-0-12-813029-2.00009-8. URL: http://www.sciencedirect.com/science/article/pii/B9780128130292000098.

[392] Linton A. Mohammed. *Forensic Examination of Signatures.* Vol. 1. Academic Press, 2019, p. 242. ISBN: 978-0-12-813029-2. DOI: https://doi.org/10.1016/C2016-0-04445-5. URL: http://www.sciencedirect.com/science/article/pii/B9780128130292000098.

[393] Rania A. Mohammed, Rebaz M. Nabi, Sardasht M. Raouf Mahmood, and Rebwar M. Nabi. "State-of-the-Art in Handwritten Signature Verification System". In: (2015), pp. 519–525. DOI: 10.1109/csci.2015.180.

[394] I. Montani. "Exploring transparent approaches to the authentication of signatures on artwork". Thesis. 2015.

[395] I. Montani, R. Marquis, N. Egli Anthonioz, and C. Champod. "Resolving differing expert opinions". In: *Sci. Justice* 59.1 (2019), pp. 1–8. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2018.10.003. URL: https://www.ncbi.nlm.nih.gov/pubmed/30654962.

[396] R. Morris. *Forensic Handwriting Identification: Fundamental Concepts and Principles.* 1st edition. Vol. 1. Academic Press, 2000, p. 238. ISBN: 9780125076401. URL: https://books.google.ch/books?id=G-FgC0kwtLoC.

[397] G. S. Morrison and E. Enzinger. "What should a forensic practitioner's likelihood ratio be?" In: *Sci. Justice* 56.5 (2016), pp. 374–379. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2016.05.007. URL: https://www.ncbi.nlm.nih.gov/pubmed/27702454.

[398] G. S. Morrison, E. Enzinger, and C. Zhang. "Refining the relevant population in forensic voice comparison - A response to Hicks et alii (2015) The importance of distinguishing information from evidence/observations when formulating propositions". In: *Sci. Justice* 56.6 (2016), pp. 492–497. ISSN: 1355-0306 (Print) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2016.07.002. URL: https://www.ncbi.nlm.nih.gov/pubmed/27914557.

[399] G. S. Morrison, D. H. Kaye, D. J. Balding, D. Taylor, P. Dawid, C. G. G. Aitken, S. Gittelson, G. Zadora, B. Robertson, S. Willis, S. Pope, M. Neil, K. A. Martire, A. Hepler, R. D. Gill, A. Jamieson, J. de Zoete, R. B. Ostrum, and A. Caliebe. "A comment on the PCAST report: Skip the "match"/"non-match" stage". In: *Forensic Sci. Int.* 272 (2017), e7–e9. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2016.10.018. URL: https://www.ncbi.nlm.nih.gov/pubmed/27817943.

[400] Geoffrey Stewart Morrison. *Calculation of forensic likelihood ratios: use of monte carlo simulations to compare te output of scorebased approaches with true likelihood-ratio values.* Conference Paper. Leiden University, Netherlands, 2014. URL: http://geoff-morrison.net/#ICFIS2014.

[401] Geoffrey Stewart Morrison. "What should a forensic practitioner's likelihood ratio be? II". In: *Sci. Justice* 57.6 (2017), pp. 472–476. ISSN: 1355-0306. DOI: 10.1016/j.scijus.2017.08.004.

[402] Geoffrey Stewart Morrison and Ewald Enzinger. "Score based procedures for the calculation of forensic likelihood ratios - scores should take account of both similarity and typicality". In: *Sci. Justice* 58.1 (2018). ISSN: 1355-0306. DOI: 10.1016/j.scijus.2017.06.005.

[403] Geoffrey Stewart Morrison, Jonas Lindh, and James M. Curran. "Likelihood ratio calculation for a disputed-utterance analysis with limited available data". In: *Speech Communication* 58 (2014), pp. 81–90. ISSN: 01676393. DOI: 10.1016/j.specom.2013.11.004.

[404] Geoffrey Stewart Morrison and Norman Poh. "Avoiding overstating the strength of forensic evidence: Shrunk likelihood ratios/Bayes factors". In: *Sci. Justice* (2017). ISSN: 13550306. DOI: 10.1016/j.scijus.2017.12.005.

[405] Geoffrey Stewart Morrison and Reinoud D. Stoel. "Forensic strength of evidence statements should preferably be likelihood ratios calculated using relevant data, quantitative measurements, and statistical models – a response to Lennard (2013) Fingerprint identification: how far have we come?" In: *Aust. J. Forensic Sci.* 46.3 (2013), pp. 282–292. ISSN: 0045-0618 1834-562X. DOI: 10.1080/00450618.2013.833648.

[406] Sascha Müller and Olaf Henniger. "Evaluating the Biometric Sample Quality of Handwritten Signatures". In: *Advances in Biometrics*. Ed. by Seong-Whan Lee and Stan Z. Li. Springer Berlin Heidelberg, 2007, pp. 407–414. ISBN: 978-3-540-74549-5.

[407] Mario E. Munich and Pietro Perona. "Camera-based ID verification by signature tracking". In: *Computer Vision — ECCV'98*. Ed. by Hans Burkhardt and Bernd Neumann. Springer Berlin Heidelberg, pp. 782–796. ISBN: 978-3-540-69354-3.

[408] Daigo Muramatsu and Takashi Matsumoto. "Effectiveness of Pen Pressure, Azimuth, and Altitude Features for Online Signature Verification". In: *Advances in Biometrics*. Ed. by Seong-Whan Lee and Stan Z. Li. Vol. 4642. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2007. Chap. 53, pp. 503–512. ISBN: 978-3-540-74548-8. DOI: 10.1007/978-3-540-74549-5_53. URL: http://dx.doi.org/10.1007/978-3-540-74549-5_53.

[409] Kevin Murphy. "Conjugate Bayesian analysis of the Gaussian distribution". In: *University of British Columbia Online* (2007).

[410] National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press, 2009, p. 352. ISBN: 978-0-309-13130-8. DOI: doi:10.17226/12589. URL: https://www.nap.edu/catalog/12589/strengthening-forensic-science-in-the-united-states-a-path-forward.

[411] NATIONAL CONFERENCE OF COMMISSIONERS ON UNIFORM STATE LAWS. *UNIFORM ELECTRONIC TRANSACTIONS ACT*. Legal Rule or Regulation. 1999.

[412] NATIONAL COMMISSION ON FORENSIC SCIENCE. *Views of the Commission Ensuring That Forensic Analysis Is Based Upon Task-Relevant Information*. Report. Department of Justice and NIST, 2015.

[413] Andreas Nautsch, Didier Meuwly, Daniel Ramos, Jonas Lindh, and Christoph Busch. "Making Likelihood Ratios Digestible for Cross-Application Performance Assessment". In: *IEEE Signal Process. Lett.* 24.10 (2017), pp. 1552–1556. ISSN: 1070-9908 1558-2361. DOI: 10.1109/lsp.2017.2748899.

[414] C. Neumann and M. Ausdemore. "Defence Against the Modern Arts the Curse of Statistics "Score-based likelihood ratios"". In: *ArXiv* (2019), p. 24.

[415] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, and A. Bromage-Griffiths. "Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae". In: *J. Forensic Sci.* 52.1 (2007), pp. 54–64. ISSN: 0022-1198 (Print) 0022-1198 (Linking). DOI: 10.1111/j.1556-4029.2006.00327.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/17209910.

[416] C. Neumann, I. W. Evett, and J. Skerrett. "Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175.2 (2012), pp. 371–415. ISSN: 0964-1998. DOI: 10.1111/j.1467-985X.2011.01027.x. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2011.01027.x.

[417] C. Neumann and C. P. Saunders. "Commentary on: Alberink I, de Jongh A, Rodriguez C. Fingermark evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios. J Forensic Sci 2014; 59(1):70-81". In: *J. Forensic Sci.* 60.1 (2015), pp. 252–6. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.12634. URL: https://www.ncbi.nlm.nih.gov/pubmed/25585667.

[418] Neuroscript LLC. *MovAlyzeR*. Visualization and Data Analysis. Version 6.1. Neuroscript LLC, 2016. URL: https://neuroscript.net/movalyzer.php.

[419] Neuroscript LLC. *MovAlyzeR by Neuroscript LLC*. Web Page. Jan. 23, 2016. URL: http://www.neuroscript.net/movalyzer.php.

[420] Kathleen Annunziata Nicolaides. "Using acceleration plots in the forensic examination of electronically captured signatures". In: *Journal of the American Society of Questioned Document Examiners* 15.2 (2012), pp. 29–43.

[421] C. O'Reilly and R. Plamondon. "Agonistic and antagonistic interaction in speed/accuracy tradeoff: a delta-lognormal perspective". In: *Hum Mov Sci* 32.5 (2013), pp. 1040–55. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2012.07.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/23182432.

[422] Yasuo Oda and Yoshihisa Sugiyama. *Position Detection Apparatus*. Ed. by United States Patent and Trademark Office. Patent. Wacom Co. Ltd., 2013.

[423] M. Okawa. "A Single-Template Strategy Using Multi-Distance Measures and Weighting for Signature Verification". In: *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. 2018, pp. 046–051. ISBN: 2162-7843. DOI: 10.1109/ISSPIT.2018.8642687.

[424] Raymond Orta M. and Magdalena Ezcurra G. "Neurosciences Applied to Handwriting Examination". In: *Journal of the American Society of Questioned Document Examiners* (2007), pp. 63–69.

[425] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J. L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. W. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J.

Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S. K. Pavani, A. Frangi, L. Akarun, and A. Savran. “The multiscenario multienvironment BioSecure Multimodal Database (BMDB)”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 32.6 (2010), pp. 1097–111. ISSN: 1939-3539 (Electronic) 0098-5589 (Linking). DOI: 10.1109/TPAMI.2009.76. URL: https://www.ncbi.nlm.nih.gov/pubmed/20431134.

[426] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J. Igarza, C. Vivaracho, D. Escudero, and Q. Moro. “MCYT baseline corpus: a bimodal biometric database”. In: *IEE Proceedings - Vision, Image and Signal Processing* 150.6 (2003), pp. 395–401. ISSN: 1350-245X. DOI: 10.1049/ip-vis:20031078.

[427] Albert S. Osborn. *Questioned Documents.* Chicago: Nelson-Hall Co., 1929.

[428] Brent Ostrum. “The Logical Approach to Evidence Evaluation”. In: *Forensic Document Examination in the 21st Century.* Ed. by J.S. Kelly and M. Angel. 1st ed. CRC Press, 2020. DOI: https://doi.org/10.4324/9780367853587.

[429] Brent R. Ostrum and Tobin A. Tanaka. “Another Look at Handwriting Movement”. In: *Journal of the American Society of Questioned Document Examiners* (2006), pp. 57–67.

[430] R. Brent Ostrum. “CSFS Document Section Position on the Logical Approach to Evidence Evaluation and Corresponding Wording of Conclusions”. In: *Can. Soc. Forensic Sci. J.* 52.3 (2019), pp. 129–138. ISSN: 0008-5030. DOI: 10.1080/00085030.2019.1635736. URL: https://doi.org/10.1080/00085030.2019.1635736.

[431] Mark Page, Jane Taylor, and Matt Blenkin. “Forensic Identification Science Evidence Since Daubert: Part I—A Quantitative Analysis of the Exclusion of Forensic Identification Science Evidence”. In: *J. Forensic Sci.* 56.5 (2011), pp. 1180–1184. DOI: doi:10.1111/j.1556-4029.2011.01777.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1556-4029.2011.01777.x.

[432] Mark Page, Jane Taylor, and Matt Blenkin. “Forensic Identification Science Evidence Since Daubert: Part II—Judicial Reasoning in Decisions to Exclude Forensic Identification Evidence on Grounds of Reliability”. In: *J. Forensic Sci.* 56.4 (2011), pp. 913–917. DOI: doi:10.1111/j.1556-4029.2011.01776.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1556-4029.2011.01776.x.

[433] S. Pal, H. Al-Assam, and H. Sellahewa. “On the discrimination power of dynamic features for online signature”. In: *2017 22nd International Conference on Digital Signal Processing (DSP).* 2017, pp. 1–5. ISBN: 2165-3577. DOI: 10.1109/ICDSP.2017.8096038.

[434] R. C. Park. “Signature identification in the light of science and experience”. In: *Hastings Law Journal* 59.5 (2008), pp. 1101–1158.

[435] European Parliament. *DIRECTIVE 1999/93/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 December 1999 on a Community framework for electronic signatures.* Legal Rule or Regulation. 2000.

[436] Marianela Parodi, Linda Alewijnse, Marcus Liwicki, and Juan C. Gómez. "Orthogonal function representation for online signature verification: which features should be looked at?" In: *IET Biom.* 2.4 (2013), pp. 137–150. ISSN: 2047-4938 2047-4946. DOI: 10.1049/iet-bmt.2013.0025.

[437] Marianela Parodi, Juan C. Gomez, and Marcus Liwicki. "Online Signature Verification Based on Legendre Series Representation: Robustness Assessment of Different Feature Combinations". In: (2012), pp. 379–384. DOI: 10.1109/icfhr.2012.251.

[438] Marianela Parodi, Juan Carlos Gomez, and L. C. Alewijnse. *Automatic Online Signature Verification based only on FHE Features: an Oxymoron?* Conference Paper. Crete Island - Greece, 2014.

[439] Marianela Parodi and Juan C. Gómez. "Legendre polynomials based feature extraction for online signature verification. Consistency analysis of feature combinations". In: *Pattern Recognit.* 47.1 (2014), pp. 128–140. ISSN: 00313203. DOI: 10.1016/j.patcog.2013.06.026.

[440] Antonio Parziale, Angelo Anatrella, and Angelo Marcelli. "Stability, Speed and Accuracy for Online Signature Verification". In: *Journal of the American Society of Questioned Document Examiners* 18.2 (2015), pp. 39–49.

[441] Antonio Parziale, Moises Diaz, Miguel A. Ferrer, and Angelo Marcelli. "SM-DTW: Stability Modulated Dynamic Time Warping for signature verification". In: *Pattern Recognit. Lett.* 121 (2019), pp. 113–122. ISSN: 01678655. DOI: 10.1016/j.patrec.2018.07.029.

[442] Antonio Parziale, Adolfo Santoro, Angelo Marcelli, Anna Paola Rizzo, Cristiano Molinari, Andrea Giuseppe Cappuzzo, and Fabio Fontana. *An Interactive Tool for Forensic Handwriting Examination.* Conference Paper. 2014. DOI: 10.1109/icfhr.2014.80.

[443] Antonio Parziale, Rosa Senatore, Angelo Marcelli, Anna Rizzo, Cristiano Molinari, Andrea Cappuzzo, and Fabio Fontana. "Combining FHE features with machine decision making for automatic writer identification". In: *18th Conference of the International Graphonomics Society (IGS).* 2017.

[444] Executive Office of the President - President's Council of Advisors on Science and Technology. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods.* Report. Office of Science and Technology Policy, Sept. 2016. URL: https://www.whitehouse.gov/blog/2016/09/20/pcast-releases-report-forensic-science-criminal-courts.

[445] Judea Pearl. "The limitations of opaque learning machines". In: *Possible minds: twenty-five ways of looking at AI* (2019), pp. 13–19.

[446] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect.* Penguin books, 2018.

[447] J. G. Phillips, R. P. Ogeil, and C. Best. "Motor constancy and the upsizing of handwriting". In: *Human Movement Science* 28.5 (2009), pp. 578–87. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2009.07.004. URL: http://www.ncbi.nlm.nih.gov/pubmed/19665812.

[448] J. G. Phillips, R. P. Ogeil, and F. Muller. "Alcohol consumption and handwriting: a kinematic analysis". In: *Human Movement Science* 28.5 (2009), pp. 619–32. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2009.01.006. URL: http://www.ncbi.nlm.nih.gov/pubmed/19395108.

[449] M. Piekarczyk and M. R. Ogiela. *The Touchless Person Authentication Using Gesture-Types Emulation of Handwritten Signature Templates.* Conference Paper. Krakow, Poland, 2015. DOI: 10.1109/BWCCA.2015.109.

[450] G. Pirlo, V. Cuccovillo, M. Diaz-Cabrera, D. Impedovo, and P. Mignone. "Multidomain Verification of Dynamic Signatures Using Local Stability Analysis". In: *IEEE Trans. Human-Mach. Syst.* 45.6 (2015), pp. 805–810. ISSN: 2168-2291 2168-2305. DOI: 10.1109/thms.2015.2443050.

[451] G. Pirlo, D. Impedovo, R. Plamondon, and C. O'Reilly. *Advances in Digital Handwritten Signature Processing.* WORLD SCIENTIFIC, 2014. URL: http://www.worldscientific.com/doi/abs/10.1142/9789814579636_0001.

[452] G. Pirlo, D. Impedovo, R. Plamondon, and C. O'Reilly. "Stability Analysis of Online Signatures in the Generation Domain". In: *Advances in Digital Handwritten Signature Processing.* WORLD SCIENTIFIC, 2014, pp. 1–12. DOI: doi:10.1142/9789814579636_0001. URL: http://www.worldscientific.com/doi/abs/10.1142/9789814579636_0001.

[453] G. Pirlo, D. Impedovo, E. Stasolla, and C. A. Trullo. "Learning Local Correspondences for Static Signature Verification". In: *AI*IA 2009: Emergent Perspectives in Artificial Intelligence.* Ed. by Roberto Serra and Rita Cucchiara. Springer Berlin Heidelberg, 2009, pp. 385–394. ISBN: 978-3-642-10291-2.

[454] Giuseppe Pirlo, Moises Diaz, Miguel Angel Ferrer, Donato Impedovo, Fabrizio Occhionero, and Urbano Zurlo. *Early Diagnosis of Neurodegenerative Diseases by Handwritten Signature Analysis.* Conference Paper. Cham, 2015.

[455] Giuseppe Pirlo, Fabrizio Rizzi, Annalisa Vacca, and Donato Impedovo. *Interoperability of Biometric Systems: Analysis of Geometric Characteristics of Handwritten Signatures.* Conference Paper. Cham, 2015. DOI: 10.1007/978-3-319-23222-5_30.

[456] R. Plamondon and M. Djioua. "A multi-level representation paradigm for handwriting stroke generation". In: *Hum Mov Sci* 25.4-5 (2006), pp. 586–607. ISSN: 0167-9457 (Print) 0167-9457 (Linking). DOI: 10.1016/j.humov.2006.07.004. URL: https://www.ncbi.nlm.nih.gov/pubmed/17023083.

[457] Réjean Plamondon, Adel M. Alimi, Pierre Yergeau, and Franck Leclerc. "Modelling velocity profiles of rapid movements: a comparative study". In: *Biol. Cybern.* 69.2 (1993), pp. 119–128. ISSN: 1432-0770. DOI: 10.1007/BF00226195. URL: https://doi.org/10.1007/BF00226195.

[458] Réjean Plamondon and Wacef Guerfali. "The generation of handwriting with delta-lognormal synergies". In: *Biol. Cybern.* 78 (1998), pp. 119–132. DOI: 10.1007/s004220050419.

[459] Réjean Plamondon, Xiaolin Li, and Moussa Djioua. "Extraction of delta-lognormal parameters from handwriting strokes". In: *Frontiers of Computer Science in China* 1.1 (2007), pp. 106–113. ISSN: 1673-7350 1673-7466. DOI: 10.1007/s11704-007-0009-0.

[460] Réjean Plamondon and Guy Lorette. "Automatic signature verification and writer identification — the state of the art". In: *Pattern Recognit.* 22.2 (1989), pp. 107–131. ISSN: 0031-3203. DOI: https://doi.org/10.1016/0031-3203(89)90059-9. URL: http://www.sciencedirect.com/science/article/pii/0031320389900599.

[461] Réjean Plamondon, Christian O'Reilly, Céline Rémi, and Thérésa Duval. "The lognormal handwriter: learning, performing, and declining". In: *Front. Psychol.* 4.945 (2013). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00945. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00945.

[462] Réjean Plamondon, Christian O'Reilly, Javier Galbally, Abdullah Almaksour, and Éric Anquetil. "Recent developments in the study of rapid human movements with the kinematic theory: Applications to handwriting and signature synthesis". In: *Pattern Recognit. Lett.* 35 (2014), pp. 225–235. ISSN: 01678655. DOI: 10.1016/j.patrec.2012.06.004.

[463] S. Planton, M. Jucla, F. E. Roux, and J. F. Demonet. "The "handwriting brain": a meta-analysis of neuroimaging studies of motor versus orthographic processes". In: *Cortex* 49.10 (2013), pp. 2772–87. ISSN: 1973-8102 (Electronic) 0010-9452 (Linking). DOI: 10.1016/j.cortex.2013.05.011. URL: http://www.ncbi.nlm.nih.gov/pubmed/23831432.

[464] Martin Potschka. *Peirce's Concept of Abduction (Hypothesis Formation) across His Later Stages of Scholarly Life*. Online Multimedia. 2017.

[465] D. Prattichizzo, L. Meli, and M. Malvezzi. "Digital Handwriting with a Finger or a Stylus: A Biomechanical Comparison". In: *IEEE Trans. Haptics* 8.4 (2015), pp. 356–370. ISSN: 1939-1412. DOI: 10.1109/TOH.2015.2434812.

[466] R Core Team. *R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.* Software. Version 3.6.3. Feb. 29, 2020.

[467] C. Rabasse, R. M. Guest, and M. C. Fairhurst. "A Method for the Synthesis of Dynamic Biometric Signature Data". In: *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on.* Vol. 1. 2007, pp. 168–172. ISBN: 1520-5363. DOI: 10.1109/ICDAR.2007.4378697.

[468] K. S. Radhika and S. Gopika. "Online and Offline Signature Verification: A Combined Approach". In: *Procedia Comput. Sci.* 46 (2015), pp. 1593–1600. ISSN: 18770509. DOI: 10.1016/j.procs.2015.02.089.

[469] D. Ramos. *Comparison of Robust Approaches to Forensic Voice Comparison Using Automated Systems*. Conference Paper. Lyon, France, 2018.

[470] D. Ramos and J. Gonzalez-Rodriguez. "Reliable support: Measuring calibration of likelihood ratios". In: *Forensic Sci. Int.* 230.1-3 (2013), pp. 156–69. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2013.04.014. URL: https://www.ncbi.nlm.nih.gov/pubmed/23664798.

[471] D. Ramos, J. Gonzalez-Rodriguez, G. Zadora, and C. Aitken. “Information-theoretical assessment of the performance of likelihood ratio computation methods”. In: *J. Forensic Sci.* 58.6 (2013), pp. 1503–18. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.12233. URL: https://www.ncbi.nlm.nih.gov/pubmed/23879526.

[472] D. Ramos, R. Haraksim, and D. Meuwly. “Likelihood ratio data to report the validation of a forensic fingerprint evaluation method”. In: *Data Brief* 10 (2017), pp. 75–92. ISSN: 2352-3409 (Print) 2352-3409 (Linking). DOI: 10.1016/j.dib.2016.11.008. URL: https://www.ncbi.nlm.nih.gov/pubmed/27981197.

[473] Daniel Ramos, Ram P. Krish, Julian Fierrez, and Didier Meuwly. “From Biometric Scores to Forensic Likelihood Ratios”. In: *Handbook of Biometrics for Forensic Science*. Ed. by Massimo Tistarelli and C. Champod. Springer, 2017, pp. 305–327. ISBN: 2191-6586 2191-6594. DOI: 10.1007/978-3-319-50673-9_14.

[474] C. Rathgeb, A. Uhl, and Christoph Busch. “Ageing effects and implications for biometric template protection”. In: *Age factors in biometric processing*. 2013. Chap. 17, pp. 321–341. DOI: 10.1049/PBSP010E_ch17.

[475] Ajita Rattani, Biagio Freni, Gian Luca Marcialis, and Fabio Roli. “Template Update Methods in Adaptive Biometric Systems: A Critical Review”. In: *Advances in Biometrics*. Ed. by Massimo Tistarelli and Mark S. Nixon. Springer Berlin Heidelberg, 2009, pp. 847–856. ISBN: 978-3-642-01793-3.

[476] O. Ribaux. *Police Scientifique - Le renseignement par la trace*. 1st. Vol. 1. Lausanne: Presses Polytechniques et Universitaires Romandes, 2014, p. 520.

[477] O. Ribaux, A. Baylon, C. Roux, O. Delemont, E. Lock, C. Zingg, and P. Margot. “Intelligence-led crime scene processing. Part I: Forensic intelligence”. In: *Forensic Sci. Int.* 195.1-3 (2010), pp. 10–6. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2009.10.027. URL: https://www.ncbi.nlm.nih.gov/pubmed/19932575.

[478] O. Ribaux and B. Talbot Wright. “Expanding forensic science through forensic intelligence”. In: *Sci. Justice* 54.6 (2014), pp. 494–501. ISSN: 1876-4452 (Electronic) 1355-0306 (Linking). DOI: 10.1016/j.scijus.2014.05.001. URL: http://www.ncbi.nlm.nih.gov/pubmed/25498939.

[479] J. Richiardi, Ketabdar Hamed, and A. Drygajlo. “Local and global feature selection for on-line signature verification”. In: *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. 2005, 625–629 Vol. 2. ISBN: 2379-2140. DOI: 10.1109/ICDAR.2005.152.

[480] J. Richiardi, H. Ketabdar, and A. Drygajlo. *Local and global feature selection for on-line signature verification*. Conference Paper. Seoul, Korea, 2005. DOI: 10.1109/ICDAR.2005.152.

[481] D Michael Risinger. “Cases Involving the Reliability of Handwriting Identification Expertise Since the Decision in Daubert”. In: *Tulsa Law Review* 43.2 (2007), pp. 477–596.

[482] D Michael Risinger and M. Saks. “Science and Nonscience in the Courts Daubert meets handwriting identification expertise”. In: *Iowa Law Review* 82.1 (1996), pp. 21–82.

[483] D. Michael Risinger. “Defining the ’Task at Hand’: Non-Science Forensic Science after Kumho Tire V. Carmichael”. In: *Washington & Lee Law Review* 57.1 (2000), p. 34. URL: https://ssrn.com/abstract=257546.

[484] D. Michael Risinger. “Goodbye to All That, or a Fool’s Errand, By One of the Fools: How I Stopped Worrying about Court Responses to Handwriting Identification (and Forensic Science in General) and Learned to Love Misinterpretations of Kumho Tire v. Carmichael”. In: *Tulsa Law Review* 43.2 (2007), pp. 447–476. URL: http://digitalcommons.law.utulsa.edu/tlr/vol43/iss2/10.

[485] D. Michael Risinger, Mark P. Denbeaux, and Michael J. Saks. “Exorcism of ignorance as proxy for rational knowledge: the lessons of handwriting identification "expertise"”. In: *University of Pennsylvania Law Review* 137 (1989), pp. 731–792.

[486] D. Michael Risinger and Michael J. Saks. “A House with No Foundation”. In: *Issues Sci. Technol.* 20.1 (2003), pp. 35–39. ISSN: 07485492, 19381557. URL: http://www.jstor.org/stable/43312396.

[487] D. Michael Risinger, Michael J. Saks, William C. Thompson, and Robert Rosenthal. “The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion”. In: *Calif. Law Rev.* 90.1 (2002), pp. 1–56. ISSN: 00081221. DOI: 10.2307/3481305. URL: www.jstor.org/stable/3481305.

[488] Paul Roberts and Colin Aitken. *The Logic of Forensic Proof: Inferential Reasoning in Criminal Evidence and Forensic Science*. Vol. 3. Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses. Royal Statistical Society, 2014, p. 161.

[489] Bernard Robertson, G. A. Vignaux, and Charles E. H. Berger. *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. John Wiley & Sons, Ltd, 2016. ISBN: 9781118492475. DOI: 10.1002/9781118492475.fmatter. URL: http://dx.doi.org/10.1002/9781118492475.fmatter.

[490] Sara Rosenblum, Batya Engel-Yeger, and Yael Fogel. “Reprint of ‘Age-related changes in executive control and their relationships with activity performance in handwriting’”. In: *Human Movement Science* 32.5 (2013), pp. 1056–1069. ISSN: 01679457. DOI: 10.1016/j.humov.2013.08.001.

[491] Q. Rossy, D. Décary-Hétu, O. Delémont, and M. Mulone. *The Routledge International Handbook of Forensic Intelligence and Criminology*. Taylor & Francis, 2017. ISBN: 9781134889020. URL: https://books.google.ch/books?id=QdNBDwAAQBAJ.

[492] S. Roux, T. J. McKeeff, G. Grosjacques, O. Afonso, and S. Kandel. “The interaction between central and peripheral processes in handwriting production”. In: *Cognition* 127.2 (2013), pp. 235–41. ISSN: 1873-7838 (Electronic) 0010-0277 (Linking). DOI: 10.1016/j.cognition.2012.12.009. URL: http://www.ncbi.nlm.nih.gov/pubmed/23454797.

[493] Tamás Rudas. *Handbook of Probability: Theory and Applications*. Thousand Oaks, California: SAGE Publications, Inc., 2008. DOI: 10.4135/9781452226620. URL: https://methods.sagepub.com/book/handbook-of-probability.

[494] N. Sae-Bae and N. Memon. "Quality of online signature templates". In: *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*. 2015, pp. 1–8. DOI: 10.1109/ISBA.2015.7126354.

[495] Napa Sae-Bae and N. Memon. "Online Signature Verification on Mobile Devices". In: *Information Forensics and Security, IEEE Transactions on* 9.6 (2014), pp. 933–947. ISSN: 1556-6013. DOI: 10.1109/TIFS.2014.2316472.

[496] Napa Sae-Bae, Nasir Memon, and Pitikhate Sooraksa. "Distinctiveness, complexity, and repeatability of online signature templates". In: *Pattern Recognit.* 84 (2018), pp. 332–344. ISSN: 00313203. DOI: 10.1016/j.patcog.2018.07.024.

[497] H. Sajid and S. c. S. Cheung. "VSig: Hand-gestured signature recognition and authentication with wearable camera". In: *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. 2015, pp. 1–6. DOI: 10.1109/WIFS.2015.7368566.

[498] M. J. Saks. "Commentary on: Srihari SN, Cha S-H, Arora H, Lee S. Individuality of handwriting. J Forensic Sci 2002; 47(4):856-72". In: *J Forensic Sci* 48.4 (2003), 916–8, author reply 919–20. ISSN: 0022-1198 (Print) 0022-1198. DOI: 10.1520/jfs2002428.

[499] M. J. Saks and H. VanderHaar. "On the "general acceptance" of handwriting identification principles". In: *J. Forensic Sci.* 50.1 (2005), pp. 119–26. ISSN: 0022-1198 (Print) 0022-1198.

[500] Michael J. Saks. "Forensic identification: From a faith-based "Science" to a scientific science". In: *Forensic Science International* 201.1 (2010), pp. 14–17. ISSN: 0379-0738. DOI: https://doi.org/10.1016/j.forsciint.2010.03.014. URL: http://www.sciencedirect.com/science/article/pii/S0379073810001064.

[501] Michael J. Saks and David L. Faigman. "Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It". In: *Annual Review of Law and Social Science* 4.1 (2008), pp. 149–171. ISSN: 1550-3585 1550-3631. DOI: 10.1146/annurev.lawsocsci.4.110707.172303.

[502] Michael J. Saks and Jonathan J. Koehler. "The Individualization Fallacy in Forensic Science Evidence". In: *Vanderbilt Law Review* 61.1 (2008), pp. 199–219.

[503] Raul Sanchez-Reillo, Helga C. Quiros-Sandoval, Ines Goicoechea-Telleria, and Wendy Ponce-Hernandez. "Improving Presentation Attack Detection in Dynamic Handwritten Signature Biometrics". In: *IEEE Access* 5 (2017), pp. 20463–20469. ISSN: 2169-3536. DOI: 10.1109/access.2017.2755771.

[504] C. P. Saunders. *Score-based Likelihood Ratios for Handwriting Evidence Evaluation*. Conference Paper. NIST, 2013.

[505] C. P. Saunders, L. J. Davis, and J. Buscaglia. “Using automated comparisons to quantify handwriting individuality”. In: *J. Forensic Sci.* 56.3 (2011), pp. 683–9. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/j.1556-4029.2011.01713.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/21391999.

[506] Christopher P. Saunders, Linda J. Davis, Andrea C. Lamas, John J. Miller, and Donald T. Gantz. “Construction and evaluation of classifiers for forensic document analysis”. In: *The Annals of Applied Statistics* 5.1 (2011), pp. 381–399. ISSN: 1932-6157. DOI: 10.1214/10-aoas379.

[507] S. Sayeed, R. Besar, and N. S. Kamel. “Dynamic Signature Verification Using Sensor Based Data Glove”. In: *8th International Conference on Signal Processing*. Vol. 3. 2006, p. 1. DOI: 10.1109/ICOSP.2006.345880.

[508] Tobias Scheidat, Karl Kümmel, and Claus Vielhauer. “Short Term Template Aging Effects on Biometric Dynamic Handwriting Authentication Performance”. In: *Communications and Multimedia Security*. Ed. by Bart De Decker and David W. Chadwick. Springer Berlin Heidelberg, 2006, pp. 107–116. ISBN: 978-3-642-32805-3.

[509] L. R. B. Schomaker and R. Plamondon. “The relation between pen force and pen-point kinematics in handwriting”. In: *Biol. Cybern.* 63.4 (1990), pp. 277–289. ISSN: 1432-0770. DOI: 10.1007/BF00203451. URL: https://doi.org/10.1007/BF00203451.

[510] N. J. Schweitzer and M. Saks. “The CSI Effect Popular Fiction about Forensic science affects the public’s expectations about real forensic science”. In: *Jurimetrics* 47 (2007), pp. 357–364.

[511] Emmanuelle Sciacca. “Etude de mentions manuscrites apposées dans des conditions non-conventionelles”. Thesis. 2010.

[512] Emmanuelle Sciacca, Jean-Claude Gilhodes, Pierre Margot, Marie-Blanche Langlois-Peter, and Jean-Luc Velay. “The range of handwriting variability under different writing conditions”. In: *Journal of Forensic Document Examination* 19 (2008), pp. 5–13.

[513] Emmanuelle Sciacca, Marie-Blanche Langlois-Peter, Pierre Margot, and J. L. Velay. “Effects of different postural conditions on handwriting variability”. In: *Journal of Forensic Document Examination* 21 (2011), pp. 51–60.

[514] Enric Sesa-Nogueras, Marcos Faundez-Zanuy, and Jiří Mekyska. “An Information Analysis of In-Air and On-Surface Trajectories in Online Handwriting”. In: *Cognitive Computation* 4.2 (2011), pp. 195–205. ISSN: 1866-9956 1866-9964. DOI: 10.1007/s12559-011-9119-y.

[515] Jungpil Shin and Toshiki Okuyama. “Detection of alcohol intoxication via online handwritten signature verification”. In: *Pattern Recognit. Lett.* 35 (2014), pp. 101–104. ISSN: 01678655. DOI: 10.1016/j.patrec.2012.07.016.

[516] J.A. Siegel, P.J. Saukko, and G.C. Knupfer. *Encyclopedia of forensic sciences*. Academic, 2000. URL: https://books.google.ch/books?id=bNrXAAAAMAAJ.

[517] J. Sita, B. Found, and D. K. Rogers. "Forensic handwriting examiners' expertise for signature comparison". In: *J. Forensic Sci.* 47.5 (2002), pp. 1117–24. ISSN: 0022-1198 (Print) 0022-1198.

[518] J. C. Sita and K. A. Taylor. "Eye movements during the handwriting of words: Individually and within sentences". In: *Human Movement Science* 43 (2015), pp. 229–38. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2015.01.011. URL: http://www.ncbi.nlm.nih.gov/pubmed/25687832.

[519] J. Skerrett, C. Neumann, and I. Mateos-Garcia. "A Bayesian approach for interpreting shoemark evidence in forensic casework: accounting for wear features". In: *Forensic Sci. Int.* 210.1-3 (2011), pp. 26–30. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2011.01.030. URL: https://www.ncbi.nlm.nih.gov/pubmed/21377817.

[520] Slate Online Magazine. *When did we start signing our names to authenticate documents?* Web Page. Mar. 11, 2011. URL: https://slate.com/news-and-politics/2011/03/when-did-we-start-signing-our-names-to-authenticate-documents.html.

[521] K. W. Smalldon and A. C. Moffat. "The Calculation of Discriminating Power for a Series of Correlated Attributes". In: *Journal of the Forensic Science Society* 13.4 (1973), pp. 291–295. ISSN: 0015-7368. DOI: https://doi.org/10.1016/S0015-7368(73)70828-8. URL: http://www.sciencedirect.com/science/article/pii/S0015736873708288.

[522] Daniel Lester Solomon. *Partially Bayes Estimates*. The Florida State University, 1968.

[523] Xiaoyu Song, Xinghua Xia, and Fangjun Luan. "Online Signature Verification Based on Stable Features Extracted Dynamically". In: *IEEE Trans. Syst., Man, Cybern., Syst.* 47.10 (2017), pp. 2663–2676. ISSN: 2168-2216 2168-2232. DOI: 10.1109/tsmc.2016.2597240.

[524] S. Srihari. *Computational Methods for Handwritten Questioned Document Examination*. Report. U.S. Department of Justice, 2010.

[525] S. Srihari, C. Huang, and H. Srinivasan. "On the discriminability of the handwriting of twins". In: *J. Forensic Sci.* 53.2 (2008), pp. 430–46. ISSN: 0022-1198 (Print) 0022-1198 (Linking). DOI: 10.1111/j.1556-4029.2008.00682.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/18366576.

[526] S. Srihari and C.G. Leedham. *Survey of Computer Methods in Forensic Handwritten Document Examination*. Conference Paper. Scottsdale, AZ, USA, 2003.

[527] S. N. Srihari, L. Meng, and L. Hanson. "Development of Individuality in Children's Handwriting". In: *J. Forensic Sci.* 61.5 (2016), pp. 1292–300. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.13158. URL: https://www.ncbi.nlm.nih.gov/pubmed/27431360.

[528] Sargur N. Srihari, Sung-Hyuk Cha, Hina Arora, and Sangjik Lee. "Individuality of Handwriting". In: *J. Forensic Sci.* 47.4 (2002), pp. 1–17.

[529] Sargur N. Srihari, Sung-Hyuk Cha, and Sangjik Lee. "The Discriminatory Power of Handwriting". In: *Document Recognition and Retrieval IX*. Ed. by Paul B. Kantor, Tapas Kanungo, and Jiangying Zhou. Vol. 4670. Society of Photo Optical, 2001.

[530] Sargur N. Srihari, Chen Huang, Harish Srinivasan, and Vivek Shah. "Biometric and Forensic Aspects of Digital Document Processing". In: *Digital Document Processing: Major Directions and Recent Advances*. Ed. by Bidyut B. Chaudhuri. London: Springer London, 2007, pp. 379–405. ISBN: 978-1-84628-726-8. DOI: 10.1007/978-1-84628-726-8_17. URL: https://doi.org/10.1007/978-1-84628-726-8_17.

[531] Sargur N. Srihari, Kamal Kuzhinjedathu, Harish Srinivasan, Chen Huang, and Danjun Pu. "Signature Verification Using a Bayesian Approach". In: *International Workshop on Computational Forensics*. Ed. by Sargur N. Srihari and Katrin Franke. Springer Berlin Heidelberg, 2008, pp. 192–203. ISBN: 978-3-540-85303-9.

[532] Sargur N. Srihari and Kirsten Singer. "Role of automation in the examination of handwritten items". In: *Pattern Recognit.* 47.3 (2014), pp. 1083–1095. ISSN: 00313203. DOI: 10.1016/j.patcog.2013.09.032.

[533] Wim Staat. "On Abduction, Deduction, Induction and the Categories". In: *Transactions of the Charles S. Peirce Society* 29.2 (1993), pp. 225–237. ISSN: 00091774, 15589587. URL: www.jstor.org/stable/40320413.

[534] Statisticat LLC. *LaplacesDemon Tutorial*. Manual. R package version 16.1.4. 2020. URL: https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software.

[535] M. Stauffer, P. Maergner, A. Fischer, R. Ingold, and K. Riesen. "Offline Signature Verification using Structural Dynamic Time Warping". In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 2019, pp. 1117–1124. ISBN: 2379-2140. DOI: 10.1109/ICDAR.2019.00181.

[536] StepOver International GmbH. *StepOver International Clientele and References*. Web Page. 2020. URL: https://www.stepover.com/references/.

[537] Hal S. Stern, Miriam Angel, Melvin Cavanaugh, Shuying Zhu, and Eric L. Lai. "Assessing the complexity of handwritten signatures". In: *Law, Probability and Risk* 17.2 (2018), pp. 123–132. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgy007.

[538] R. D. Stoel, E. Van den Heuvel, and L. C. Alewijnse. "Structural equation modelling of subjective measures with an application to forensic handwriting examination". In: *Sci. Justice* 50.1 (2010), pp. 31–32. ISSN: 13550306. DOI: 10.1016/j.scijus.2009.11.027.

[539] David A. Stoney. "What made us ever think we could individualize using statistics?" In: *J. Forensic Sci. Soc.* 31.2 (1991), pp. 197–199. ISSN: 00157368. DOI: 10.1016/s0015-7368(91)73138-1.

[540] Steven Strach, Michelle Novotny, Andrea Devlin, Claire Graydon, and Paul Westwood. *Development of a Draft Methodology for Electronic Signature Examination using Wacom SignatureScope*. Conference Paper. Pensacola Beach, Florida, 2016.

[541] J. J. Summers and J. G. Anson. "Current status of the motor program: revisited". In: *Hum Mov Sci* 28.5 (2009), pp. 566–77. ISSN: 1872-7646 (Electronic) 0167-9457 (Linking). DOI: 10.1016/j.humov.2009.01.002. URL: https://www.ncbi.nlm.nih.gov/pubmed/19230995.

[542] SWGDOC. *Standard for Examination of Handwritten Items*. Standard. 2013. URL: http://www.swgdoc.org/index.php/standards/published-standards.

[543] SWGDOC. *Standard for Scope of Work of Forensic Document Examiners*. Standard. 2013. URL: http://www.swgdoc.org/index.php/standards/published-standards.

[544] SWGDOC. *Standard Terminology for Expressing Conclusions of Forensic Document Examiners*. Standard. 2013. URL: http://www.swgdoc.org/index.php/standards/published-standards.

[545] SWGDOC. *Terminology Relating to the Examination of Questioned Documents*. Standard. 2013. URL: http://www.swgdoc.org/index.php/standards/published-standards.

[546] Lei Tang, Yuxun Fang, Qiuxia Wu, Wenxiong Kang, and Junhong Zhao. "Online Finger-Writing Signature Verification on Mobile Device for Local Authentication". In: 9967 (2016), pp. 409–416. ISSN: 0302-9743 1611-3349. DOI: 10.1007/978-3-319-46654-5_45.

[547] Yi Tang, Christian Viard-Gaudin, Sargur Srihari, Harish Srinivasan, and Richard Zanibbi. *Handwriting individualization using distance and rarity*. Conference Paper. Burlingame, California, United States, 2012. DOI: 10.1117/12.912035.

[548] C. C. Tappert, C. Y. Suen, and T. Wakahara. "The state of the art in online handwriting recognition". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 12.8 (1990), pp. 787–808. ISSN: 1939-3539. DOI: 10.1109/34.57669.

[549] S. Tariq, S. Sarwar, and W. Hussain. "Classification of Features into Strong and Weak Features for an Intelligent Online Signature Verification System". In: *First International Workshop on Automated Forensic Handwriting Analysis*. 2011.

[550] Saad Tariq, Saqib Sarwar, and Waqar Hussain. "Classification of Features into Strong and Weak Features for an Intelligent Online Signature Verification System". In: 768 (2011).

[551] F. Taroni. "Inadequacies of posterior probabilities for the assessment of scientific evidence". In: *Law, Probability and Risk* 4.1-2 (2005), pp. 89–114. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgi008.

[552] F. Taroni, A. Biedermann, S. Bozza, P. Garbolino, and C. Aitken. *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*. Wiley, 2014. ISBN: 9781118914748. URL: https://books.google.ch/books?id=v7cTBAAAQBAJ.

[553] F. Taroni, S. Bozza, and C. Aitken. "Decision Analysis in Forensic Science". In: *J. Forensic Sci.* 50.4 (2005), pp. 894–905. ISSN: 0022-1198 (Print) 0022-1198. DOI: 10.1520/JFS2004443.

[554] F. Taroni, P. Garbolino, A. Biedermann, C. Aitken, and S. Bozza. "Reconciliation of subjective probabilities and frequencies in forensic science". In: *Law, Probability and Risk* 17.3 (2018), pp. 243–262. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgy014.

[555] F. Taroni, R. Marquis, M. Schmittbuhl, A. Biedermann, A. Thiery, and S. Bozza. "Bayes factor for investigative assessment of selected handwriting features". In: *Forensic Sci. Int.* 242 (2014), pp. 266–73. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2014.07.012. URL: http://www.ncbi.nlm.nih.gov/pubmed/25117907.

[556] F. Taroni, R. Marquis, M. Schmittbuhl, A. Biedermann, A. Thiery, and S. Bozza. "The use of the likelihood ratio for evaluative and investigative purposes in comparative forensic handwriting examination". In: *Forensic Sci. Int.* 214.1-3 (2012), pp. 189–94. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2011.08.007. URL: http://www.ncbi.nlm.nih.gov/pubmed/21907508.

[557] Franco Taroni, Silvia Bozza, Alex Biedermann, and Colin Aitken. "Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio". In: *Law, Probability and Risk* 15.1 (2016), pp. 1–16. ISSN: 1470-8396 1470-840X. DOI: 10.1093/lpr/mgv008.

[558] Franco Taroni, Silvia Bozza, Alex Biedermann, G. Garbolino, and C. G. G. Aitken. *Data Analysis in Forensic Science: A Bayesian Decision Perspective*. John Wiley & Sons, 2010, p. 388. DOI: 10.1002/9780470665084.

[559] H. L. Teulings, J. L. Contreras-Vidal, G. E. Stelmach, and C. H. Adler. "Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control". In: *Exp. Neurol.* 146.1 (1997), pp. 159–70. ISSN: 0014-4886 (Print) 0014-4886. DOI: 10.1006/exnr.1997.6507.

[560] H. L. Teulings and L. R. Schomaker. "Invariant properties between stroke features in handwriting". In: *Acta Psychol (Amst)* 82.1-3 (1993), pp. 69–88. ISSN: 0001-6918 (Print) 0001-6918.

[561] Hans-Leo Teulings, Arnold J. W. M. Thomassen, and Gerard P. van Galen. "Invariants in Handwriting: The Information Contained in a Motor Program". In: *Advances in Psychology*. Ed. by Henry S. R. Kao, Gerard P. van Galen, and Rumjahn Hoosain. Vol. 37. North-Holland, 1986, pp. 305–315. ISBN: 0166-4115. DOI: https://doi.org/10.1016/S0166-4115(09)60089-3. URL: http://www.sciencedirect.com/science/article/pii/S0166411509600893.

[562] A. Thiery. "Développement d'un processus de quantification et d'évaluation de caractères manuscrits : théorie et applications". Thesis. 2014.

[563] A. Thiéry, R. Marquis, and I. Montani. "Statistical evaluation of the influence of writing postures on on-line signatures. Study of the impact of time". In: *Forensic Sci. Int.* 230.1-3 (2013), pp. 107–16. ISSN: 1872-6283 (Electronic) 0379-0738 (Linking). DOI: 10.1016/j.forsciint.2012.10.033. URL: http://www.ncbi.nlm.nih.gov/pubmed/23153802.

[564] W. C. Thompson. "How Should Forensic Scientists Present Source Conclusions?" In: *Seton Hall Law Review* 48.3 (2018), pp. 774–805.

[565] W. C. Thompson and E. J. Newman. "Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents". In: *Law Hum. Behav.* 39.4 (2015), pp. 332–49. ISSN: 1573-661X (Electronic) 0147-7307 (Linking). DOI: 10.1037/lhb0000134. URL: https://www.ncbi.nlm.nih.gov/pubmed/25984887.

[566] W. C. Thompson, J. Vuille, F. Taroni, and A. Biedermann. “After uniqueness: the evolution of forensic science opinions”. In: *Judicature* 102.1 (2018), pp. 18–27.

[567] William C. Thompson. “Discussion paper: Hard cases make bad law—reactions to R v T”. In: *Law, Probability and Risk* 11.4 (2012), pp. 347–359. DOI: 10.1093/lpr/mgs020. URL: https://doi.org/10.1093/lpr/mgs020.

[568] William C. Thompson, Suzanne O. Kaasa, and Tiamoyo Peterson. “Do Jurors Give Appropriate Weight to Forensic Identification Evidence?” In: *Journal of Empirical Legal Studies* 10.2 (2013), pp. 359–397. ISSN: 1740-1453. DOI: 10.1111/jels.12013. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jels.12013.

[569] M. Tistarelli and C. Champod. *Handbook of Biometrics for Forensic Science*. Springer International Publishing, 2017. ISBN: 9783319506739. URL: https://books.google.ch/books?id=aZONDgAAQBAJ.

[570] Massimo Tistarelli, Enrico Grosso, and Didier Meuwly. “Biometrics in Forensic Science: Challenges, Lessons and New Technologies”. In: Biometric Authentication. Springer International Publishing, 2014, pp. 153–164. ISBN: 978-3-319-13386-7.

[571] R. Tolosana, M. Gomez-Barrero, C. Busch, and J. Ortega-Garcia. “Biometric Presentation Attack Detection: Beyond the Visible Spectrum”. In: *IEEE Trans. Inf. Forensics Security* 15 (2020), pp. 1261–1275. ISSN: 1556-6021. DOI: 10.1109/TIFS.2019.2934867.

[572] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia. “Benchmarking desktop and mobile handwriting across COTS devices: The e-BioSign biometric database”. In: *PLoS One* 12.5 (2017), e0176792. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0176792. URL: https://www.ncbi.nlm.nih.gov/pubmed/28475590.

[573] R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez. “Optimal feature selection and inter-operability compensation for on-line biometric signature authentication”. In: *2015 International Conference on Biometrics (ICB)*. 2015, pp. 163–168. ISBN: 2376-4201. DOI: 10.1109/ICB.2015.7139047.

[574] R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez. “Update strategies for HMM-based dynamic signature biometric systems”. In: *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. 2015, pp. 1–6. DOI: 10.1109/WIFS.2015.7368583.

[575] Ruben Tolosana, Ruben Vera Rodriguez, Julian Fierrez, and Javier Ortega-Garcia. “Reducing the Template Aging Effect in On-Line Signature Biometrics”. In: *IET Biom.* (2019). ISSN: 2047-4938 2047-4946. DOI: 10.1049/iet-bmt.2018.5259.

[576] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, and Javier Ortega-Garcia. “BioTouchPass2: Touchscreen Password Biometrics Using Time-Aligned Recurrent Neural Networks”. In: *IEEE Trans. Inf. Forensics Security* 15 (2020), pp. 2616–2628. ISSN: 1556-6021. DOI: 10.1109/tifs.2020.2973832.

[577] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, and Javier Ortega-Garcia. "Presentation Attacks in Signature Biometrics: Types and Introduction to Attack Detection". In: 2019. DOI: 10.1007/978-3-319-92627-8_19.

[578] Ruben Tolosana, Ruben Vera-Rodriguez, Carlos Gonzalez-Garcia, Julian Fierrez, Aythami Morales, Javier Ortega-Garcia, Juan-Carlos Ruiz-Garcia, Sergio Romero-Tapiador, Santiago Rengifo, Miguel Caruana, Jiajia Jiang, Songxuan Lai, Lianwen Jin, Yecheng Zhu, Javier Galbally, Moises Diaz Cabrera, Miguel Angel Ferrer, Marta Gomez-Barrero, Ilya A. Hodashinsky, Konstantin S. Sarin, Artem Slezkin, Marina Bardamova, Mikhail Svetlakov, Mohammad Saleem, Cintia Lia Szücs, Bence Kovari, Falk Pulsmeyer, Mohamad Wehbi, Dario Zanca, Sumaiya Ahmad, Sarthak Mishra, and Suraiya Jabin. "SVC-onGoing: Signature Verification Competition". In: *CoRR* abs/2108.06090 (2021). arXiv: 2108.06090. URL: https://arxiv.org/abs/2108.06090.

[579] Ruben Tolosana, Ruben Vera-Rodriguez, Carlos Gonzalez-Garcia, Julian Fierrez, Santiago Rengifo, Aythami Morales, Javier Ortega-Garcia, Juan Carlos Ruiz-Garcia, Sergio Romero-Tapiador, Jiajia Jiang, et al. "ICDAR 2021 Competition on On-Line Signature Verification". In: *arXiv preprint arXiv:2106.00739* (2021).

[580] Ruben Tolosana, Ruben Vera-Rodriguez, Richard Guest, Julian Fierrez, and Javier Ortega-Garcia. "Exploiting Complexity in Pen- and Touch-based Signature Biometrics". In: *International Journal on Document Analysis and Recognition (IJDAR)* (2019). ISSN: 1433-2825. DOI: 10.1007/s10032-020-00351-3.

[581] Ruben Tolosana, Ruben Vera-Rodriguez, Javier Ortega-Garcia, and Julian Fierrez. "Preprocessing and Feature Selection for Improved Sensor Interoperability in Online Biometric Signature Verification". In: *IEEE Access* 3 (2015), pp. 478–489. ISSN: 2169-3536. DOI: 10.1109/access.2015.2431493.

[582] O. Tucha and K. W. Lange. "Effects of nicotine chewing gum on a real-life motor task: a kinematic analysis of handwriting movements in smokers and non-smokers". In: *Psychopharmacology (Berl)* 173.1-2 (2004), pp. 49–56. ISSN: 0033-3158 (Print) 0033-3158. DOI: 10.1007/s00213-003-1690-9.

[583] O. Tucha, L. Mecklinger, S. Walitza, and K. W. Lange. "Attention and movement execution during handwriting". In: *Human Movement Science* 25.4-5 (2006), pp. 536–52. ISSN: 0167-9457 (Print) 0167-9457 (Linking). DOI: 10.1016/j.humov.2006.06.002. URL: http://www.ncbi.nlm.nih.gov/pubmed/17011653.

[584] O. Tucha, S. Walitza, L. Mecklinger, D. Stasik, T. A. Sontag, and K. W. Lange. "The effect of caffeine on handwriting movements in skilled writers". In: *Human Movement Science* 25.4-5 (2006), pp. 523–35. ISSN: 0167-9457 (Print) 0167-9457 (Linking). DOI: 10.1016/j.humov.2006.06.001. URL: http://www.ncbi.nlm.nih.gov/pubmed/17010462.

[585] Peter V. Tytell. "Pen pressure as identifying characteristic of signatures: Verification from the Computer". In: *Journal of the American Society of Questioned Document Examiners* 1.1 (1998), pp. 21–31.

[586] United States Mississippi District Court. *American Family Life Assurance Company of Columbus v Biles*. Case. United States District Court, S.D. Mississippi, Jackson Division, 2011.

[587] United States Court of Appeals of Indiana. *Jessica Borjas v State Indiana*. Case. Court of Appeals of Indiana, 2011.

[588] United States Court Of Appeals Of Indiana. *Shawn P. Green v State of Indiana*. Case. COURT OF APPEALS OF INDIANA, 2011.

[589] T. W. Vastrick, E. Schuetzner, and K. Osborn. "Measuring the Frequency Occurrence of Handwritten Numeral Characteristics". In: *J. Forensic Sci.* 63.4 (2018), pp. 1215–1220. ISSN: 1556-4029 (Electronic) 0022-1198 (Linking). DOI: 10.1111/1556-4029.13678. URL: https://www.ncbi.nlm.nih.gov/pubmed/29098694.

[590] R. Vera-Rodriguez, R. Tolosana, J. Ortega-Garcia, and J. Fierrez. "E-biosign: stylus- and finger-input multi-device database for dynamic signature recognition". In: *3rd International Workshop on Biometrics and Forensics (IWBF 2015)*. 2015, pp. 1–6. DOI: 10.1109/IWBF.2015.7110242.

[591] Ruben Vera-Rodriguez, Julian Fierrez, and Javier Ortega-Garcia. "Dynamic Signatures as Forensic Evidence: A New Expert Tool Including Population Statistics". In: *Handbook of Biometrics for Forensic Science*. Ed. by Massimo Tistarelli and Christophe Champod. Cham: Springer International Publishing, 2017, pp. 329–349. ISBN: 978-3-319-50673-9. DOI: 10.1007/978-3-319-50673-9_15. URL: https://doi.org/10.1007/978-3-319-50673-9_15.

[592] Ruben Vera-Rodriguez, Ruben Tolosana, Javier Hernandez-Ortega, Alejandro Acien, Aythami Morales, Julian Fierrez, and Javier Ortega-Garcia. "Modeling the Complexity of Signature and Touch-Screen Biometrics using the Lognormality Principle". In: *The Lognormality Principle and its Applications*. World Scientific, 2018.

[593] Claus Vielhauer. "Handwriting biometrics: issues of integration in identification documents and sensor interoperability". In: *J. Electron. Imaging* 15.4 (2006), p. 041103. ISSN: 1017-9909. DOI: 10.1117/1.2399540.

[594] Claus Vielhauer, T. K. Basu, Jana Dittmann, and Pranab K. Dutta. "Finding meta data in speech and handwriting biometrics". In: *Electronic Imaging 2005*. Vol. 5681. SPIE, 2005, p. 12.

[595] Claus Vielhauer, Berrin Yanikoglu, Sonia Garcia-Salicetti, Richard Guest, and Stephen Elliott. "Special Section on Biometrics: Advances in Security, Usability, and Interoperability". In: *J. Electron. Imaging* 17.1 (2008), p. 011001. URL: https://doi.org/10.1117/1.2840774.

[596] B. V. K. Vijaya Kumar, Stephen J. Elliott, Salil Prabhakar, Eric P. Kukula, and Arun A. Ross. "A definitional framework for the human/biometric sensor interaction model". In: 7667 (2010), 76670H. ISSN: 0277786X. DOI: 10.1117/12.850595.

[597] B. V. K. Vijaya Kumar, Alisher Kholmatov, Salil Prabhakar, Berrin Yanikoglu, and Arun A. Ross. "An individuality model for online signatures using global Fourier descriptors". In: 6944 (2008), p. 694407. ISSN: 0277786X. DOI: 10.1117/12.777640.

[598] P. Viviani and C. Terzuolo. "Trajectory determines movement dynamics". In: *Neuroscience* 7.2 (1982), pp. 431–437. ISSN: 0306-4522. DOI: 10.1016/0306-4522(82)90277-9.

[599] C. S. Vorugunti, R. K. S. Gorthi, and V. Pulabaigari. "Online Signature Verification by Few-Shot Separable Convolution Based Deep Learning". In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 2019, pp. 1125–1130. ISBN: 2379-2140. DOI: 10.1109/ICDAR.2019.00182.

[600] C. S. Vorugunti, D. S. Guru, and V. Pulabaigari. "An Efficient Online Signature Verification Based on Feature Fusion and Interval Valued Representation of Writer Specific Features". In: *2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA)*. 2019, pp. 1–9. ISBN: 2640-0790. DOI: 10.1109/ISBA.2019.8778566.

[601] C. S. Vorugunti, P. Mukherjee, and V. Pulabaigari. "Online Signature Profiling using Generative Adversarial Networks". In: *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. 2020, pp. 894–896. ISBN: 2155-2509. DOI: 10.1109/COMSNETS48256.2020.9027369.

[602] C. S. Vorugunti, G. Devanur S, P. Mukherjee, and V. Pulabaigari. "OSVNet: Convolutional Siamese Network for Writer Independent Online Signature Verification". In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 2019, pp. 1470–1475. ISBN: 2379-2140. DOI: 10.1109/ICDAR.2019.00236.

[603] Wacom Europe GmbH. *List of Clientele using Wacom Signature Solutions*. Web Page. 2020. URL: http://signature.wacom.eu/wacom-signature/clients/.

[604] Wacom Co Ltd. *Wacom Business Solutions*. Web Page. 2020. URL: http://www.wacom.com/en-us/enterprise/business-solutions.

[605] Wacom Co. Ltd. *Wacom Signature Scope*. Visualization and Data Analysis. Version 1.44.0. Wacom Co. Ltd., 2020.

[606] Judie Walton. "Handwriting changes due to aging and Parkinson's syndrome". In: *Forensic Sci. Int.* 88 (1997), pp. 197–214.

[607] John Wann and Ian Nimmo-Smith. "The control of pen pressure in handwriting: A subtle point". In: *Human Movement Science* 10.2 (1991), pp. 223–246. ISSN: 0167-9457. DOI: https://doi.org/10.1016/0167-9457(91)90005-I. URL: http://www.sciencedirect.com/science/article/pii/016794579190005I.

[608] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

[609] Wikimedia Foundation. *Definition of 'Bayes Factor'*. Web Page. 2020. URL: https://en.wikipedia.org/wiki/Bayes_factor.

[610] Wikimedia Foundation. *Definition of 'Inference'*. Web Page. 2020. URL: https://en.wikipedia.org/wiki/Inference.

[611] Wikimedia Foundation. *Definition of 'Signature capture pad'*. Web Page. 2020. URL: https://de.wikipedia.org/wiki/Unterschriftenpad.

[612] E. J. Will. "Inferring relative speed of handwriting from the static trace". In: *Journal of Forensic Document Examination* 22.1 (2012).

[613] L. Wilson-Wilde, James Brandi, and Stephen Gutowski. *Forensic Science Standards: Interpretation and reporting in forensic investigations*. Conference Paper. 2012.

[614] Linzi Wilson-Wilde. "The future of the National Institute of Forensic Science – implications for Australia and New Zealand". In: *Aust. J. Forensic Sci.* 49.1 (2017), pp. 1–8. ISSN: 0045-0618. DOI: 10.1080/00450618.2016.1253773. URL: https://doi.org/10.1080/00450618.2016.1253773.

[615] Alan M. Wing. "Motor control: Mechanisms of motor equivalence in handwriting". In: *Curr. Biol.* 10.6 (2000), R245–R248. ISSN: 0960-9822. DOI: http://dx.doi.org/10.1016/S0960-9822(00)00375-4.

[616] Anna Woch, Rejean Plamondon, and Christian O'Reilly. "Kinematic characteristics of bidirectional delta-lognormal primitives in young and older subjects". In: *Human Movement Science* 30.1 (2011), pp. 1–17. ISSN: 0167-9457. DOI: https://doi.org/10.1016/j.humov.2009.10.006. URL: http://www.sciencedirect.com/science/article/pii/S0167945710000977.

[617] X. Wu, A. Kimura, B. K. Iwana, S. Uchida, and K. Kashino. "Deep Dynamic Time Warping: End-to-End Local Representation Learning for Online Signature Verification". In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 2019, pp. 1103–1110. ISBN: 2379-2140. DOI: 10.1109/ICDAR.2019.00179.

[618] X. Wu, A. Kimura, S. Uchida, and K. Kashino. "Prewarping Siamese Network: Learning Local Representations for Online Signature Verification". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019, pp. 2467–2471. ISBN: 2379-190X. DOI: 10.1109/ICASSP.2019.8683036.

[619] Xinghua Xia, Zhili Chen, Fangjun Luan, and Xiaoyu Song. "Signature alignment based on GMM for on-line signature verification". In: *Pattern Recognit.* 65 (2017), pp. 188–196. ISSN: 00313203. DOI: 10.1016/j.patcog.2016.12.019.

[620] Xinghua Xia, Xiaoyu Song, Fangun Luan, Jungang Zheng, Zhili Chen, and Xiaofu Ma. "Discriminative feature selection for on-line signature verification". In: *Pattern Recognit.* 74 (2018), pp. 422–433. ISSN: 00313203. DOI: 10.1016/j.patcog.2017.09.033.

[621] Tsuguya Yamanami, Takahiko Funahashi, and Toshiaki Senda. *Position Detecting Apparatus*. Ed. by United States Patent and Trademark Office. Patent. Wacom Co. Ltd., 1989.

[622] Dit-Yan Yeung, Hong Chang, Yimin Xiong, Susan George, Ramanujan Kashi, Takashi Matsumoto, and Gerhard Rigoll. "SVC2004: First International Signature Verification Competition". In: *Biometric Authentication*. Ed. by David Zhang and Anil K. Jain. Springer Berlin Heidelberg, 2004, pp. 16–22. ISBN: 978-3-540-25948-0.

[623] Shiyang Yu and Frank Zenker. "Peirce Knew Why Abduction Isn't IBE—A Scheme and Critical Questions for Abductive Argument". In: *Argumentation* 32.4 (2018), pp. 569–587. ISSN: 1572-8374. DOI: 10.1007/s10503-017-9443-9. URL: https://doi.org/10.1007/s10503-017-9443-9.

[624] G. Zadora, A. Martyna, D. Ramos, and C. Aitken. *Statistical Analysis in Forensic Science Evidential Value of Multivariate Physicochemical Data*. 1st Edition. Vol. 1. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom: John Wiley & Sons, Ltd, 2014, p. 333. ISBN: 978-0-470-97210-6. DOI: 10.1002/9781118763155.

[625] Grzegorz Zadora. "Evaluation of the evidential value of physicochemical data by a Bayesian network approach". In: *J. Chemom.* 24.7-8 (2010), pp. 346–366. ISSN: 08869383. DOI: 10.1002/cem.1307.

[626] Farhana Javed Zareen and Suraiya Jabin. "Authentic mobile-biometric signature verification system". In: *IET Biom.* 5.1 (2016), pp. 13–19. ISSN: 2047-4938 2047-4946. DOI: 10.1049/iet-bmt.2015.0017.

Appendix

Appendix A - Glossary and Abbreviations

Glossary

Authentication (Biometrics) Biometric authentication is a technique for identifying the person accessing a secured asset, be it a physical space, computer software or hardware, as being indeed who they claim to be by comparing their unique biological features such as fingerprints, palm print, retina scan, or voice pattern recognition with their corresponding features in the database and to grant the person access only when there is a match. Authentication involves a one-to-one match that compares a query sample against a template of the claimed biometric identity in the database. The claim is either accepted or rejected. [383]

Bayesian The adjective Bayesian describes the fact of being, relating to, or involving statistical methods that assign probabilities or distributions to events (such as rain tomorrow) or parameters (such as a population mean) based on experience or best guesses before experimentation and data collection and that apply Bayes' theorem to revise the probabilities and distributions after obtaining experimental data. [377]

Bayes Factor In statistics, the use of Bayes factors is a Bayesian alternative to classical hypothesis testing. Bayesian model comparison is a method of model selection based on Bayes factors. The models under consideration are statistical models. The aim of the Bayes factor is to quantify the support for a model over another, regardless of whether these models are correct. [609]

Biometrics Automated recognition of individuals based on their behavioral and biological characteristics. [383]

Digitizer A digitizer is a hardware device that receives analog information, such as sound or light, and records it digitally. Usually, the information is stored in a file on a computing device. This process is called digitization. [93]

Dynamic Signature Synonym of online signature. Digitized and discretized signature in list form, containing temporal and kinetic data. Their data is a recording of a signature, in signal form. Signature written using handwriting capture technology. In capturing an "online" signature, temporal, dynamic characteristics are recorded such as duration and velocity. [242]

Feature Representation (Descriptor) of a characteristic, extracted from raw biometric data. There are two main categories of features:

- **Global feature** Handwriting feature extracted for the entire signature unit. [242]
- **Local feature** Handwriting feature based on a single data point. [242]

Feature Extraction Features are extracted from large amounts of data resulting in a major reduction of data. Used in pattern recognition and the image processing phase of the handwriting verification process. [242]

Feature Set Synonym of Feature Vector. An ordered set of different features. The mathematical space associated with a feature in extraction of handwriting features. The feature or associated image may be represented by pixels. [242]

Feature Vector Synonym of Feature Set. An ordered set of different features. The mathematical space associated with a feature in extraction of handwriting features. The feature or associated image may be represented by pixels. [242]

Freehand Simulation Simulation method involving the attempted imitation of a handwriting or signature based upon an available model handwriting. The simulation is performed through memory of the model or observation of the model. [242]

Identification (Biometrics) Biometric identification/recognition refers to the process of automatic recognition of individuals based on their physiological and/or behavioural characteristics. Identification/recognition involves one-to-many matching that compares a query sample against all the biometric templates in the database to output the identity or the possible identity list of the input query. In this scenario, it is often assumed that the query sample belongs to the persons who are registered in a database. [383]

Inference Inferences are steps in reasoning, moving from premises to logical consequences; etymologically, the word infer means to "carry forward". Inference is theoretically traditionally divided into deduction and induction, a distinction that in Europe dates at least to Aristotle (300s BCE). Deduction is inference deriving logical conclusions from premises known or assumed to be true, with the laws of valid inference being studied in logic. Induction is inference from particular premises to a universal conclusion. A third type of inference is sometimes distinguished, notably by Charles Sanders Peirce, distinguishing abduction from induction. [610]

Likelihood Ratio A likelihood ratio is the ratio of any two specified likelihoods. In statistics, the likelihood function (often simply called the likelihood) measures the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters. It is formed from the joint probability distribution of the sample, but viewed and used as a function of the parameters only, thus treating the random variables as fixed at the observed values.

- The likelihood ratio is central to likelihoodist statistics: the law of likelihood states that degree to which data (considered as evidence) supports one parameter value versus another is measured by the likelihood ratio.
- In frequentist inference, the likelihood ratio is the basis for a test statistic, the so-called likelihood-ratio test. By the Neyman–Pearson lemma, this is the most powerful test for comparing two simple hypotheses at a given significance level. Numerous other tests can be viewed as likelihood-ratio tests or approximations thereof. The asymptotic distribution of the log-likelihood ratio, considered as a test statistic, is given by Wilks' theorem.
- The likelihood ratio is also of central importance in Bayesian inference, where it is known as the Bayes factor, and is used in Bayes' rule. Stated in terms of odds, Bayes' rule is that the posterior odds of two alternatives, A_1 and A_2 , given an event B, is the prior odds, times the likelihood ratio. [192]

Online Signature Synonym of dynamic signature. Digitized and discretized signature in list form, containing temporal and kinetic data. Their data is a recording of a signature, in signal form.

Offline Signature Synonym of static signature. Digitized signature in image form, containing pixel (color) information.

Static Signature Synonym of offline signature. Digitized signature in image form, containing pixel (color) information.

Signature (Capture) Pad A signature (capture) pad is a device that electronically captures a person's handwritten signature on an LCD touchpad using a pen-type stylus. [270, 611]

Verification (Biometrics) Synonym of Authentication. Biometric authentication is a technique for identifying the person accessing a secured asset, be it a physical space, computer software or hardware, as being indeed who they claim to be by comparing their unique biological features such as fingerprints, palm print, retina scan, or voice pattern recognition with their corresponding features in the database and to grant the person access only when there is a match. Verification involves a one-to-one match that compares a query sample against a template of the claimed biometric identity in the database. The claim is either accepted or rejected. [383]

Abbreviations

ASV Automated Signature Verification

AFSP Association of Forensic Service Providers

AAFS American Academy of Forensic Science

BF Bayes Factor

BPM Best Practice Manual

BWV Between Writer Variation

CAI Case Assessment and Interpretation

CTP Capacitive Touch Panel

CV Crossvalidation

DCS Digitally Captured Signature

DNA Deoxyribonucleic Acid

DSC Digital Signature Certificate

DSV Dynamic Signature Verification

DTW Dynamic Time Warping

EMR Electromagnetic Resonance

ESC École des Sciences Criminelles

ENFSI European Network of Forensic Science Institutes

ENFHEX European Network of Forensic Handwriting Experts

GMM Gaussian Mixture Model

HMM Hidden Markov Model

ISO International Standardization Organization

LOESS Locally Estimated Scatterplot Smoothing

LR Likelihood Ratio

NAS National Academy of Sciences

NIST National Institute of Standards and Technology

NN Neural Network (Adversarial, Artificial, Convolutional, Recurrent, Siamese, ...)

OSV Online Signature Verification

PC Personal Computer

PCAST President's Council of Advisors on Science and Technology

RRS Repeated Random Sampling

RSS Royal Statistical Society

RTP Resistive Touch Panel

STR Short Tandem Repeat

SVM Support Vector Machine

UNIL Université de Lausanne

WWV Within Writer Variation

Appendix B - Source Code

The thesis makes use of self written code in the R statistical programming language. Over a thousand lines of code were written for the project and the various publications included in the present thesis. In order to limit an excessive page number, only the core routines related to the probabilistic model, found in chapter 6, are represented here.

The code presented in the following sections relies on two package dependencies: ggplot2 [608] for visualization and plotting, as well as LaplacesDemon [534] for the non-central multivariate student-t distribution density function.

Multivariate Normal Model with unknown mean and covariance

```r
# Normal-Wishart Prior and Normal Likelihood #
# Unknown mean and covariance #

# Multi.NormWish.prior #
# function to elicit prior parameters from population

multi.NormWish.prior <- function(Dataset) # use prior population
{
  # Step 1 - Set virtual sample size for prior strength

  # kappa_0 determines how strongly the prior information impacts the results
  # kappa may dominate the updating process if the prior information is >>> than the data
  kappa_0 <- 1

  # Step 2 - determine sample mean and covariance, as well as other parameters
  mu_0 <- colMeans(Dataset)
  # Mean vector
  sigma_0 <- do.call(what=cbind,args=lapply(Dataset,sd))
  # standard deviation vector
  nu_0 <- ncol(Dataset)
  # Degrees of freedom (DoF; set at minimum (number of variables), as discussed with Gaborini)
  T_0 <- cov(Dataset)
  # covariance matrix

  # Step 3 - Return complete prior parameter list
  result <- list("kappa_0"=kappa_0,"nu_0"=nu_0,"mu_0"=mu_0,"sigma_0"=sigma_0,"T_0"=T_0)
  return(result)
}

# multi.NormWish.post #
# function to update model parameters to posterior parameters

multi.NormWish.post <- function(priorParamList, Dataset)
  # use evidence (reference and questioned samples) as dataset and prior parameters
{
  # Step 1 - Determine evidence sample parameters
  N <- nrow(Dataset)
  dim <- ncol(Dataset)
  DataMean <- colMeans(Dataset)
  MeanDiff <- (priorParamList[["mu_0"]]-DataMean)
  # intermediary result ; calculated and stored for simplicity
  DataDiff <- do.call(cbind,lapply(1:ncol(Dataset),function(x) Dataset[,x]-mean(Dataset[,x])))
  # intermediary result ; calculated and stored for simplicity

  # Step 2 - Update parameters
  kappa_n <- priorParamList[["kappa_0"]]+N
  nu_n <- priorParamList[["nu_0"]]+N
  mu_n <- (priorParamList[["kappa_0"]] * priorParamList[["mu_0"]] + N*DataMean)/
    (priorParamList[["kappa_0"]]+N)
  S <- crossprod(DataDiff)
  T_n <- priorParamList[["T_0"]] + S + ((priorParamList[["kappa_0"]]*N)/
                                          (priorParamList[["kappa_0"]]+N)*MeanDiff%*%(t(MeanDiff)))

  # Step 3 - return the updated parameters
  result <- list("kappa_n"=kappa_n,"nu_n"=nu_n,"mu_n"=mu_n,"S"=S,"T_n"=T_n,"n"=N,"d"=dim)
  return(result)
}

# multi.NormWish.postpred #
# function to determine posterior predictive model parameters

multi.NormWish.postpred <-  function(postParamList)
{
  dof_pp <- postParamList[["nu_n"]]-postParamList[["d"]]+1
  cov_pp <- postParamList[["T_n"]]*(postParamList[["kappa_n"]]+1)/(postParamList[["kappa_n"]]*postParamList[["nu_n"]]-
      postParamList[["d"]]+1)
  result <- list("dof_pp"=dof_pp,"mu_pp"=postParamList[["mu_n"]],"cov_pp"=cov_pp)
  return(result)
  }
```

BF Calculation

```
################### multi.drawfromMVSt #################
# function to generate BFs drawn form simulated artificial data drawn from the distributions

multi.drawfromMVSt <- function(fitHp=NULL,fitHd=NULL,output="BF",num=100000)
{
  # Step 1 - Transform parameters into usable form for LaplacesDemon's non central student-t distribution
  d1 <- as.integer(fitHp[['dof_pp']])
  d2 <- as.integer(fitHd[['dof_pp']])
  m1 <- as.vector(fitHp[['mu_pp']])
  m2 <- as.vector(fitHd[['mu_pp']])
  S1 <- as.matrix(fitHp[['cov_pp']])*((d1-2)/d1)
  S2 <- as.matrix(fitHd[['cov_pp']])*((d2-2)/d2)

  # Step 2 - Draw and convert a large number of artificial 'samples' from the distributions
  DrawHp <-  rmvt(n=num,df=d1,mu=m1,S=S1)
  DrawHd <-  rmvt(n=num,df=d2,mu=m2,S=S2)
  DrawHp <- as.matrix(DrawHp)
  DrawHd <- as.matrix(DrawHd)

  # Step 3 - Evaluate the samples using both distributions to obtain probabilities
  HpP <-  dmvt(x=DrawHp,df=d1,mu=m1,S=S1)
  HpD <-  dmvt(x=DrawHp,df=d2,mu=m2,S=S2)
  HdP <-  dmvt(x=DrawHd,df=d1,mu=m1,S=S1)
  HdD <-  dmvt(x=DrawHd,df=d2,mu=m2,S=S2)

  # Step 4 - Output samples or calculate and output BF
  if (output=="draw")
  {
    result <- list("HpP"=HpP,"HpD"=HpD,"HdP"=HdP,"HdD"=HdD)
  }
  else if (output=="BF")
  {
    LRHp <-  data.frame("BF"=HpP/HpD,"Type"="H1")
    LRHd <-  data.frame("BF"=HdP/HdD,"Type"="H2")
    Likelihood <-  rbind(LRHp,LRHd)
    result <- list("ProbHp"=LRHp, "ProbHd"=LRHd,"BF"=Likelihood[,1],Likelihood[,2])
  }
  return(result)
}

# multi.getLR.MVSt #
# function to calculate BFs from empirical data

multi.getLR.MVSt<- function(fitHp,fitHd,H1Sample,H2Sample)
{
  # Step 1 - conform parameters for the LaplacesDemon formulation of the student-t distribution
  d1 <- as.integer(fitHp[['dof_pp']])
  d2 <- as.integer(fitHd[['dof_pp']])
  m1 <- as.vector(fitHp[['mu_pp']])
  m2 <- as.vector(fitHd[['mu_pp']])
  S1 <- as.matrix(fitHp[['cov_pp']])*((d1-2)/d1)
  S2 <- as.matrix(fitHd[['cov_pp']])*((d2-2)/d2)

  # Step 2 - Conform Data for for the LaplacesDemon formulation of the student-t distribution
  H1 <- as.matrix(H1Sample)
  H2 <- as.matrix(H2Sample)

  # Step 3 - Calculate probabilities for the empirical samples
  HpP <-  dmvt(x=H1,df=d1,mu=m1,S=S1)
  HpD <-  dmvt(x=H1,df=d2,mu=m2,S=S2)
  HdP <-  dmvt(x=H2,df=d1,mu=m1,S=S1)
  HdD <-  dmvt(x=H2,df=d2,mu=m2,S=S2)
  LRHp <-  data.frame("BF"=HpP/HpD,"Type"="H1")
  LRHd <-  data.frame("BF"=HdP/HdD,"Type"="H2")
  Likelihood <-  rbind(LRHp,LRHd)
  result <- list("ProbHp"=LRHp, "ProbHd"=LRHd,"BF"=Likelihood[,1],Likelihood[,2])
  return(result)
}

# multi.MVSt.BF #
# function to calculate BF for casework data using multivariate student-t distribution
# fitHx = model parameters, Evidence are reference and questioned specimens

multi.MVSt.BF <- function(fitHp,fitHd,Evidence)
{
  # Step 1 - conform parameters for LaplacesDemon package
  d1 <- as.integer(fitHp[['dof_pp']])
  d2 <- as.integer(fitHd[['dof_pp']])
```

```
  m1 <- as.vector(fitHp[['mu_pp']])
  m2 <- as.vector(fitHd[['mu_pp']])
  S1 <- as.matrix(fitHp[['cov_pp']])*((d1-2)/d1)
  S2 <- as.matrix(fitHd[['cov_pp']])*((d2-2)/d2)

  # Step 2 - Enter data and calculate probabilities
  QData <- as.matrix(Evidence)
  H1Q <-  dmvt(x=QData,df=d1,mu=m1,S=S1)
  H2Q <-  dmvt(x=QData,df=d2,mu=m2,S=S2)

  # Step 3 - Calculate Bayes Factor and return results
  result <-  list("BF"=H1Q/H2Q,"Num"=H1Q,"Denom"=H2Q)
  return(result)
}
```

Model Checking

```r
# multi.NormWish.ppCheck.data #
# function to generate artificial data and check model adequacy
# Posterior predictive checks (Krushke, 2010) check intravariability and model fit

multi.NormWish.ppCheck.data <- function(PriorData,Data,var.name,nrep=1)
{
  # Step 1 - Subfunction definition

  # Univariate ppcheck for multivariate distribution fitting
  # subfunction to plot the details in an overlaid, transparent combination of histogram
  # and density functions
  # Histogram = Test Data, function curve = Predicted data density
  vis.PPCheck <- function(TestData,var.name,parameters)
  {
    arguments <- list(nu=parameters[['dof_pp']],mu=parameters[['mu_pp']][[var.name]],sigma=sqrt(parameters[['cov_pp']][var.name,var.name]))
    result <-  ggplot(TestData,aes(x=TestData[[var.name]])) +theme_bw() +
      geom_histogram(alpha = 0.6,colour='black',fill='blue',aes(y = ..density..), position = 'identity')+
      labs(x=paste0(var.name," values/scores"))+
      #scale_x_continuous(limits=c(0,10000))+
      stat_function(fun=dst,colour="blue",args=arguments) +
      ggtitle(paste0("Posterior predictive check for ",var.name))
    return(result)
  }

  # subfunction to loop function if several variables are inputted
  # generates a plot per variable
  vis.perVariable <- function(TestData,var.name,parameters)
  {
    result <- lapply(var.name,function(x) vis.PPCheck(TestData=TestData,var.name=x,parameters=parameters))
    return(result)
  }

  # subfunction to loop ppcheck to test with different test and training data sets
  check.repetition <- function(PriorData,TestData,var.name)
  {
    PDat <- as.data.frame(PriorData[,var.name])
    TestDat <- as.data.frame(TestData[,var.name])
    colnames(TestDat) <- colnames(PDat) <- var.name # redefine variable names for plotting
    Sample <- split.signatureData(TestDat,method="RRS",type=TestData[["Type"]][1],test.proportion=0.2) # split dataset into two parts by repeated random sampling, with 20% of data used for testing and 80% for training
    parameters <-  multi.NormWish(PriorData=PDat,Data=Sample[['Training.Set']])[['Post.Pred.Param']] # update parameters
    return(vis.perVariable(TestData=Sample[['Test.Set']],var.name=var.name,parameters=parameters))
  }

  # Step 2 -  Execution of ppcheck
  repetitionlist <- lapply(1:nrep,function(x) check.repetition(PriorData=PriorData,TestData = Data,var.name=var.name))
  variablelist.plots <- lapply(1:length(var.name),function(x) lapply(repetitionlist,function(y) y[[x]]))
  names(variablelist.plots)=var.name

  # Step 3 - Multiplotting
  # Plots several plots in a single graph
  for (i in 1:length(variablelist.plots))
  {
    multiplot(plotlist=variablelist.plots[[i]],cols=2)
  }
  # return(variablelist.plots)
}

# multi.NormWish.ppCheck.model #
# function to generate artificial data and check model adequacy for both populations
# visualization and check of inter and intra variability

multi.NormWish.ppCheck.model <- function(PriorData,H1Data,H2Data,var.name,nrep=1,check.type="pBayes")
{
  # Step 1 - Input check
  # check for function input errors; stop function if erroneous input provided
  if (!check.type%in%c("iBayes","pBayes"))
  {
    stop("Unknown 'check.type', please use 'iBayes' or 'pBayes'",.call=T)
  }

  # Step 2 - subfunction definition
  # Visualization function for Data and models of both populations
  vis.ModelCheck <- function(H1Data,H2Data,var.name,H1parameters,H2parameters)
  {
    H1 <- cbind(H1Data,"H1")
```

```
    colnames(H1)[ncol(H1)] <- "Origin"
    H2 <- cbind(H2Data,"H2")
    colnames(H2)[ncol(H2)] <- "Origin"
    Data <- rbind(H1,H2)

    # Subfunction for plotting
    # plotting histograms with two colors, transparency and legend
    plot.hists <- function(Data,var.name,H1parameters,H2parameters)
    {
      # redefine axes, if necessary
      # xmin <- min(Data[[var.name]])
      # xmax <- max(Data[[var.name]])

      #summarize model parameters for fuction call
      H1arguments <- list(nu=H1parameters[['dof_pp']],mu=H1parameters[['mu_pp']][[var.name]],sigma=sqrt(H1parameters[['
          cov_pp']][var.name,var.name]))
      H2arguments <- list(nu=H2parameters[['dof_pp']],mu=H2parameters[['mu_pp']][[var.name]],sigma=sqrt(H2parameters[['
          cov_pp']][var.name,var.name]))

      # plot functions and data on a single plot
      result <-  ggplot(Data,aes(x=Data[[var.name]],fill=as.factor(Origin))) +theme_bw() +
        # xlim(xmin,xmax) +
        geom_histogram(alpha = 0.6,colour='black',aes(y = ..density..), position = 'identity')+
        stat_function(fun=dst,colour='red',args=H1arguments) + stat_function(fun=dst,colour="blue",args=H2arguments)+
        labs(x=paste0(var.name," values/scores"))+  scale_fill_discrete(name="Data Type",labels=c("H1","H2"))
      ggtitle(paste0("Posterior predictive check for ",var.name))
      return(result)
    }
    # loop function to produce output for every variable (univariate)
    result <- lapply(var.name,function(x) plot.hists(Data=Data,var.name=x,H1parameters=H1parameters,H2parameters=
        H2parameters))
    return(result)
  }

  # Subfunction to split data into trainign and testing data, as well as determine model parameters
  # used to loop if multiple checks on different datasets are required
  check.repetition <- function(PriorData,H1Data,H2Data,var.name,check.type)
  {
    # select relevant data and rename variable names for plotting
    H1temp <- as.data.frame(H1Data[,var.name])
    H2temp <- as.data.frame(H2Data[,var.name])
    PDat <- as.data.frame(PriorData[,var.name])
    colnames(H1temp) <- colnames(H2temp) <- colnames(PDat)<- var.name

    # separate data for training and testing, based on repeated random sampling (RRS)
    H1Sample <- split.signatureData(H1temp,method="RRS",type=H1Data[["Type"]][1],test.proportion = 0.2)
    H2Sample <- split.signatureData(H2temp,method="RRS",type=H2Data[["Type"]][1],test.proportion = 0.2)
    # update parameters
    params <- Bayes.Method(PDat,H1Sample[["Training.Set"]],H2Sample[["Training.Set"]],method=check.type)
    H1parameters <-  params[["H1"]][['Post.Pred.Param']]
    H2parameters <-  params[["H2"]][['Post.Pred.Param']]
    # return data and parameters
    return(vis.ModelCheck(H1Data=H1Sample[['Test.Set']],H2Data=H2Sample[['Test.Set']],var.name=var.name,H1parameters=
        H1parameters,H2parameters=H2parameters))
  }

  # Step 3 - Function execution, looping and plotting
  repetitionlist <- lapply(1:nrep,function(x) check.repetition(PriorData=PriorData,H1Data = H1Data,H2Data=H2Data,var.
      name=var.name,check.type=check.type))
  variablelist.plots <- lapply(1:length(var.name),function(x) lapply(repetitionlist,function(y) y[[x]]))
  names(variablelist.plots)=var.name
  # plot all repetitions and variables in a single plot
  for (i in 1:length(variablelist.plots))
  {
    multiplot(plotlist=variablelist.plots[[i]],cols=2)
  }
  # return(variablelist.plots)
}
```

Sampling and Wrappers

```
# split.signatureData #
# function used to split a dataset according to a number of criteria

split.signatureData <- function(Data,method,type=NA,test.proportion=0.2,WriterID=1,n.train=1,n.test=1)
{
  # Step 1 - Check input for errors
  if (!(method%in%c("RRS","Writer","Size")))
  {
    stop("Unknown Method, please use 'RRS', 'Writer' or 'Size'")
  }

  if(is.na(type))
  {
    type <- Data[["Type"]][1]
  }

  if (!(type%in%c("Feature","Score")))
  {
    stop("Unknown LR Type, please use 'feature' or 'score'")
  }

  # Step 2 - Subfunction definitions
  # Repeated Random Sampling; Separate data into training and test set according to proportions
  # use the proportion of test set, generally 20% of the data
  RRS <- function(Data,test.proportion=0.2)
  {
    n.train.samp <- ceiling(nrow(Data)*(1-test.proportion))
    sample.indx <- sample(nrow(Data),size=n.train.samp,replace=F)
    training.subset <-  1:nrow(Data)%in%sample.indx
    test.subset <- !training.subset
    training <- subset(Data,subset=training.subset)
    test <- subset(Data,subset=test.subset)
    return(list("Training.Set"=training,"Test.Set"=test))
  }

  # Writer selection; Select all specimens from a specific (or random) writer
  WriterSplit <- function(Data,WriterID,type)
  {
    if(WriterID=="random")
    {
      if (type=="Feature")
      {
        possibleWriters <- unique(Data[,2])
      }
      else
      {
        possibleWriters <- unique(c(Data[,2],Data[,6]))
      }
      WriterID <- sample(possibleWriters,size=1)
    }
    if(type=="Feature")
    {
      selector <- Data[,2]==WriterID
    }
    else
    {
      selector1 <- Data[,2]==WriterID
      selector2 <- Data[,6]==WriterID
      selector <- selector1|selector2
    }
    selector.inverse <- !selector
    training.set <- subset(Data,subset=selector.inverse)
    test.set <- subset(Data,subset=selector)
    # select samples from the WriterID supplied in the function as test set
    return(list("Training.Set"=training.set,"Test.Set"=test.set))
  }

  # Divide data set according to the number of data in each set
  # specify the number of signatures for training and test set, then drawn randomly
  SampleSize <- function(Data,n.train,n.test)
  {
    DrawSubsample <- function(Data,n)
    {
      if(n>nrow(Data))
      {
        stop(paste0("Sample is too small for the desired subsample size; Sample size: ",nrow(Data),"Subsample :", n,sep
            =""),call.=T)
      }
```

```
      sample.indx <- sample(nrow(Data),size=n,replace=F)
      training.subset <-  1:nrow(Data)%in%sample.indx
      test.subset <- !training.subset
      res.subset <- subset(Data,subset=training.subset)
      res.rest <- subset(Data,subset=test.subset)
      return(list("Subset"=res.subset,"Rest"=res.rest))
    }
    firstSample <- DrawSubsample(Data=Data,n=n.train)
    secondSample <- DrawSubsample(Data=firstSample[[2]],n=n.test)
    return(list("Training.Set"=firstSample[[1]],"Test.Set"=secondSample[[1]]))
  }

  # Step 3 - Separate Dataset according to chosen criteria
  if (method=="RRS")
  {
    result <-  RRS(Data=Data,test.proportion=test.proportion)
  }
  else if (method=="Writer")
  {
    result <-  WriterSplit(Data=Data,WriterID=WriterID,type=type)
  }
  else
  {
    result <- SampleSize(Data=Data,n.train=n.train,n.test=n.test)
  }
  return(result)
}

# multi.NormWish #
# Wrapper function for updating and parameter extraction

multi.NormWish <- function(PriorData,Data)
  # use prior data and evidence
{
  priors <- multi.NormWish.prior(PriorData)
  posteriors <- multi.NormWish.post(priors,Data)
  predictives <- multi.NormWish.postpred(posteriors)
  result <- list("PriorParam"=priors,"PostParam"=posteriors,"Post.Pred.Param"=predictives)
  return(result)
}

# Bayes.Method #
# Wrapper for parameter extraction functions
# parameter updating in accodance with a traditional (pBayes) or
# non-traditional (iBayes) use of the evidence and alternative propositions

Bayes.Method <- function(PriorData,H1Data,H2Data,method="pBayes")
{
  # Step 1 - Check input for errors
  if (!method%in%c("pBayes","iBayes"))
  {
    stop("unknown method entered, use 'pBayes' or 'iBayes'",call.=T)
  }

  # Step 2 - Extract and return parameters
  if (method=="iBayes")
  {
    H1parameters <-  multi.NormWish(PriorData=PriorData,Data=H1Data)
    H2parameters <-  multi.NormWish(PriorData=PriorData,Data=H2Data)
  }
  else
  {
    H1parameters <-  multi.NormWish(PriorData=PriorData,Data=H1Data)
    H2parameters <-  multi.NormWish(PriorData=H2Data,Data=H1Data)
  }
  return(list("H1"=H1parameters,"H2"=H2parameters))
}
```

Appendix C - Population Statistics

The thesis uses two databases, one of case-unrelated genuine signatures and one of case-related forgeries as background data for the Bayesian model. The forged signature population is further separated by signature model (signature 1, 2 and 3). The detailed description of the population and acquisition conditions can be found in chapter 5. In this appendix, we represent some descriptive statistics of both populations for general information.

Genuine case-unrelated signature population

All writers in this population are in an age range from 22 to 41 years. This age range is considered stable. Due to missing data, laterality of the participants cannot be reported.

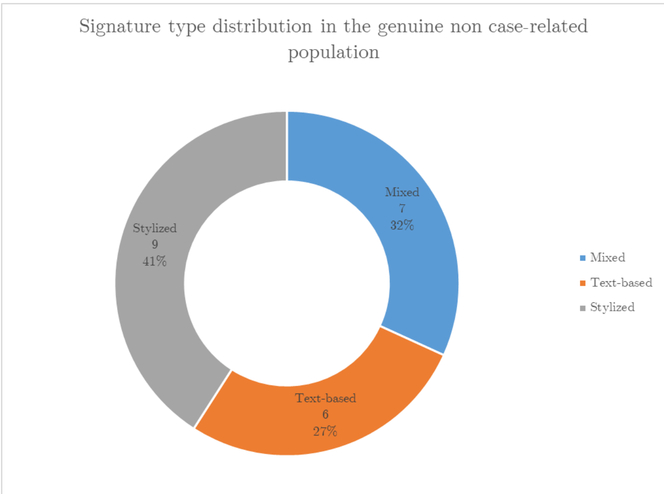
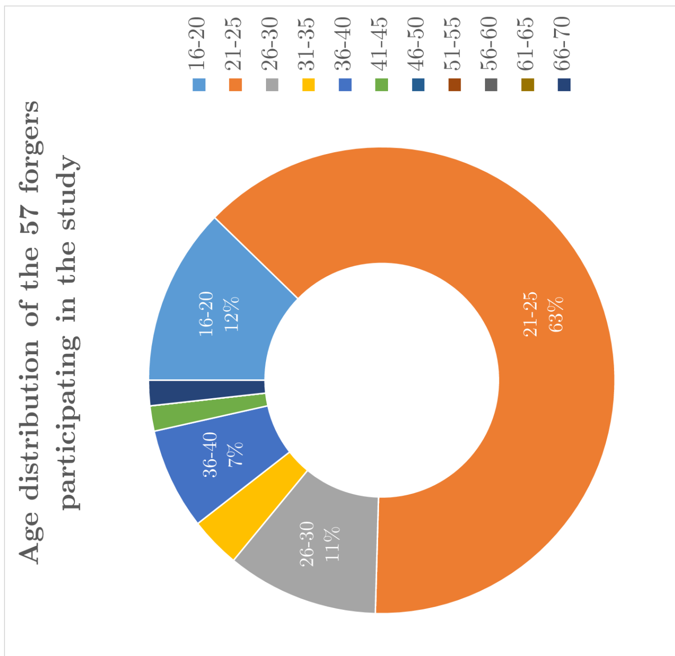
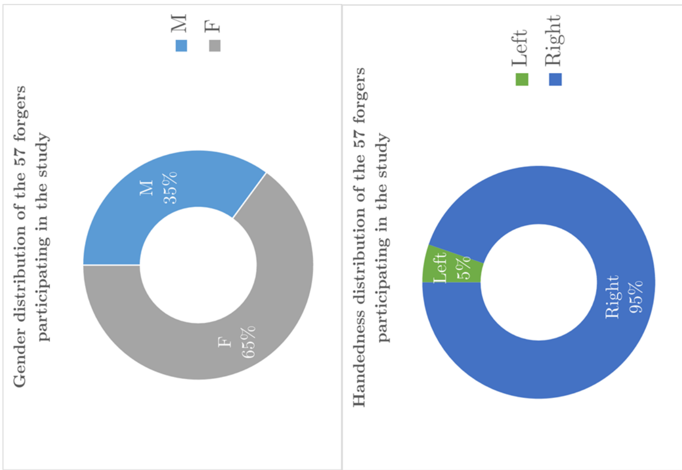


Figure C.1: Descriptive statistics of signature types present in the genuine background population

Case-related forged signature population



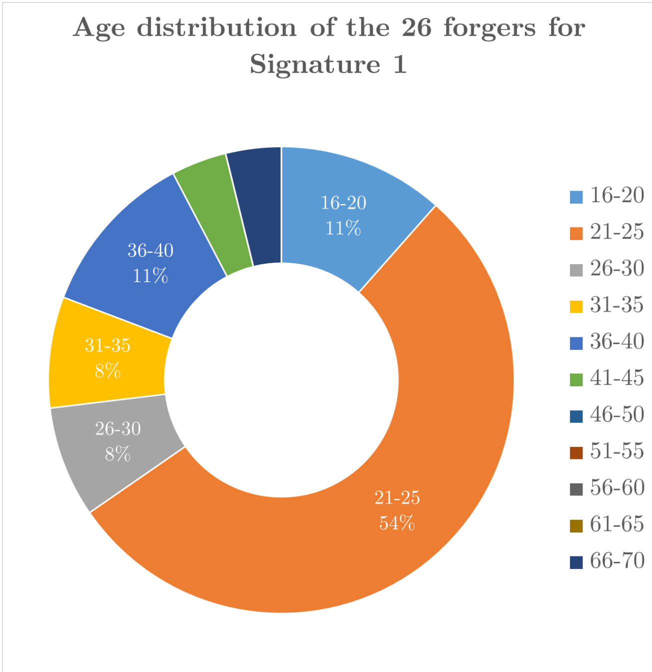
(a)



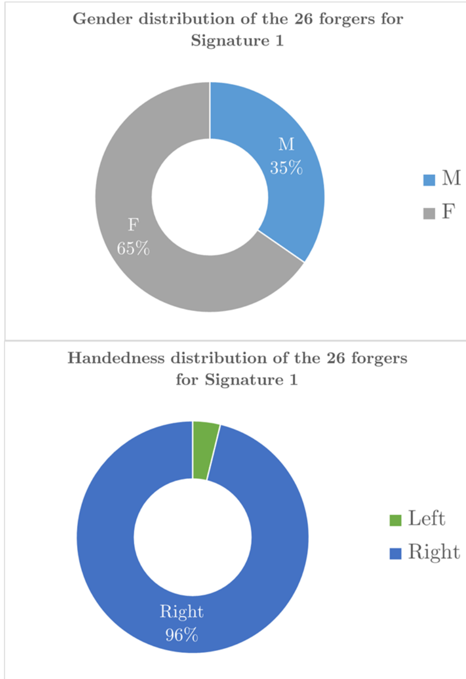
(b)

Figure C.2: Descriptive statistics of all forgers participating in the study

Signature 1 Subpopulation



(a)



(b)

Figure C.3: Descriptive statistics of forgers of Signature 1

Signature 2 Subpopulation

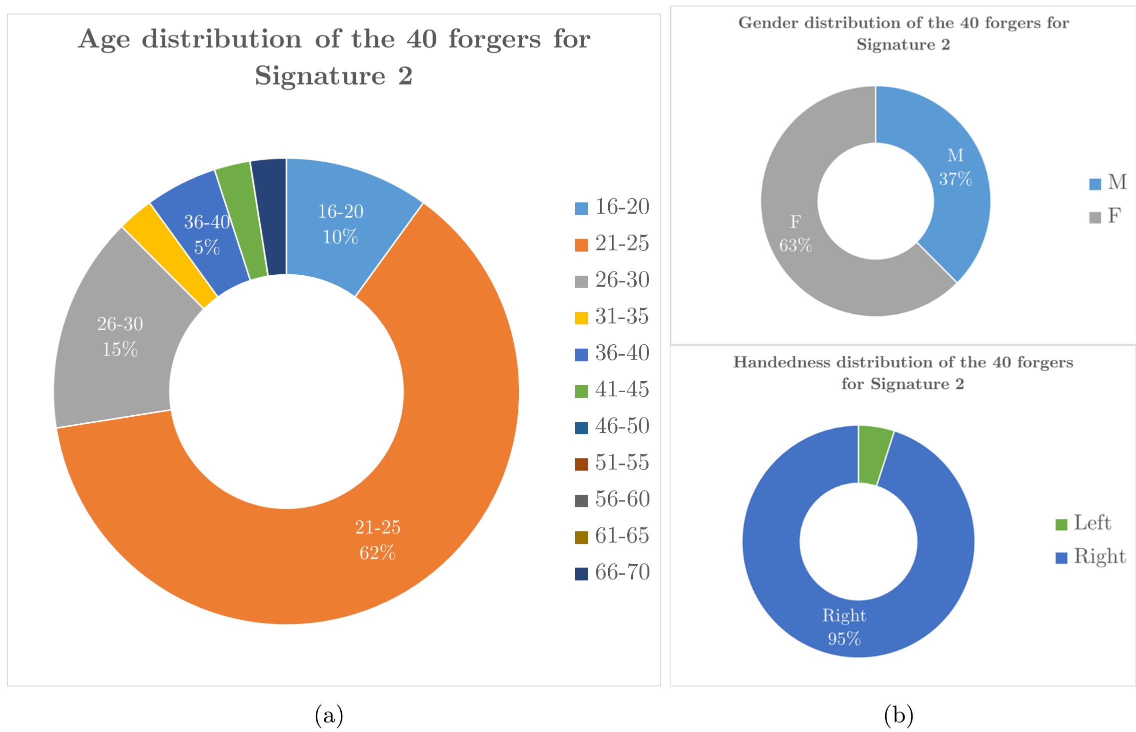
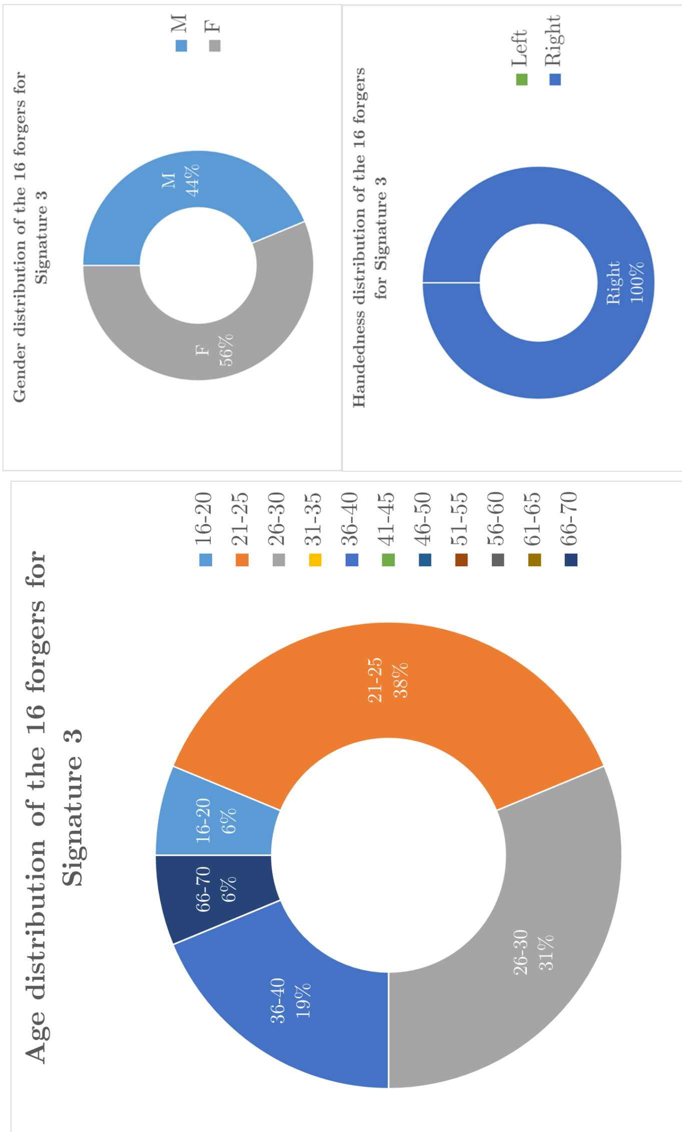


Figure C.4: Descriptive statistics of forgers of Signature 2

Signature 3 Subpopulation



(a) (b)

Descriptive statistics of forgers of Signature 3