Bayesian multivariate models for case assessment in dynamic signature cases

Authors : Jacques Linden[1], Silvia Bozza[2,1], Raymond Marquis[1], Franco Taroni[1]

1 : School of Criminal Justice, University of Lausanne, CH-1015 Lausanne Dorigny, Switzerland

2: Dipartimento di Economia, Università Ca' Foscari Venezia, Dorsoduro, 3246, 30123 Venezia VE, Italie

Full article reference :

Jacques Linden, Franco Taroni, Raymond Marquis, Silvia Bozza. Bayesian multivariate models for case assessment in dynamic signature cases. *Forensic Science International*, Volume 318, 2021, 110611, ISSN 0379-0738.

DOI :

https://doi.org/10.1016/j.forsciint.2020.110611

URL :

https://www.sciencedirect.com/science/article/pii/S0379073820304734

Bayesian multivariate models for case assessment in dynamic signature cases

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.

Highlights

- Description of the methodological and technical framework for the analysis of dynamic signatures expressed in the form of multivariate data
- Description of a probabilistic framework for the assessment of dynamic signature evidence through the quantification of the Bayes' Factor
- Assessment of model performances for signature evidence

1. Introduction

Forensic science has been strongly criticized in recent years. The highly debated NAS and PCAST reports [1-3] in the United States of America especially focus on the lack of statistical approaches in many pattern matching disciplines [2], as well as the lack of method validation. Many disciplines, handwriting and signature examination included, are indirectly concerned by these criticisms [1, 3-10], 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 [3]. 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 [11-25]. These systems offer more standardized and reproducible procedures [26], 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 [27-41] 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 [16, 42-45] 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 [46-49]. 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 [50]. 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 [51], glass evidence [40, 52], voice comparison [53], handwriting and static signature examination [54, 55]. 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 2.1. reviews the existing probabilistic models for forensic multivariate data, while a proposal for questioned dynamic signatures is illustrated in Section 2.2. The conditions and methods tested are presented in Section 3, while results and performance of the model are discussed in Section 4 . Section 5, finally, concludes the paper.

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[1] 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 assessing dynamic signature evidence, first a look at existing Bayesian multivariate models in forensic science applications (including the field of handwriting examination) is proposed.

2.1. Multivariate statistical models for forensic evidence

Several models for the evaluation of the evidence in the form of multivariate data can be found in the literature [20, 25, 40, 55, 56], although few publications deal with handwriting data [55] or signature data [54] 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

[1] While forgery is the more easily understood vocabulary, it conveys legal meaning. Therefore, forensic handwriting examiners prefer using 'simulated signatures' instead of forgeries. In this document, we use forgery and simulated signature interchangeably.

fragment. The existing probabilistic solutions rely on the assumption of sincerity[2] of available findings that implies independence between sources. This assumption is suitable for most physiological biometrics and for physical evidence. These models inherently focus on the inherent variability of features. For example, in handwriting examination [55], comparison between the questioned document and the 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 willful 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 [57] 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[3] 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 [40, 56] 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).

[2] Sincerity of the handwriting can be defined as the absence of intent to alter the 'natural' features of the handwriting. For a lack of better terminology, sincerity is going to be used to describe non-disguised or imitated signatures and writing.

[3] Propositions are very often denoted H_p and H_d for the prosecution and defense hypotheses, respectively. This is logical for criminal cases, but considering that questioned handwriting and signature cases are often civil procedures and oppose two parties rather than the prosecution and defendant, using numbers as subscript is felt to be more appropriate.

Consider a database $\{Z_{ij}\}$, with $i = 1, \dots, m$ and $j = 1, \dots, n$, where there are $p\ (> 1)$ collected features (e.g. chemical composition of glass fragments, or features of 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. Note that the within-source mean vector θ_i varies between sources $i = 1, \dots, m$, 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 handwriting or signature examination. Following Aitken & Lucy [40], 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:

$$Z_{ij} \sim N(\theta_i, W); \tag{1}$$

$$\theta_i \sim N(\mu, B), \tag{2}$$

where μ is the between-sources mean vector and B the between-sources 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:

$$(y|\theta_y, W) \sim N(\theta_y, W), \tag{3}$$

$$(x|\theta_x, W) \sim N(\theta_x, W). \tag{4}$$

The value of the evidence is computed as

$$BF = \frac{f(y, x|H_1)}{f(y, x|H_2)}, \tag{5}$$

where $f(y, x|H_1)$ is the marginal likelihood under hypothesis H_1 and $f(y, x|H_2)$ is the marginal likelihood under hypothesis H_2. If the questioned and reference material originate from the same source (i.e. H_1 holds), then the mean feature vectors are equal, $\theta_y = \theta_x = \theta$. The marginal likelihood $f(y, x|H_1)$ can be obtained as:

$$f(y, x|H_1) = \int_{\theta} f(y|\theta, W) f(x|\theta, W) f(\theta|\mu, B)\, d\theta. \tag{6}$$

Vice versa, if the questioned and reference material originate from different sources (i.e. H_2 holds), then the mean vectors within sources will be different, meaning $\theta_y \neq \theta_x$. The marginal likelihood $f(y, x|H_2)$ can be obtained as:

$$f(y, x|H_2) = \int_{\theta_y} f(y|\theta_y, W) f(\theta_y|\mu, B)\, d\theta_y \int_{\theta_x} f(x|\theta_x, W) f(\theta_x|\mu, B)\, d\theta_x. \tag{7}$$

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 (2) may be estimated starting from the available database $\{Z_{ij}\}$ as follows

$$f(\theta|\bar{z}_1, \dots, \bar{z}_m, B, h) = \frac{1}{m}\sum_{i=1}^{m} K(\theta|\bar{z}_i, B, h), \tag{8}$$

where $K(\cdot)$ is the Kernel function, $\bar{z}_i = \frac{1}{n}\sum_{j=1}^{n} z_{ij}$ are the group means and h is the smoothing parameter. Aitken and Lucy [40] propose using a Normal (Gaussian) kernel centered at the group means $\bar{z}_i$ with covariance matrix h^2B. 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 (6) and (7), respectively, where the Normal distribution for the between-source variability $f(\theta|\mu, B)$ in (2) is replaced by the kernel distribution $f(\theta|\bar{z}_1, \dots, \bar{z}_m, B, h)$ in (8).

For both the MVN and MVK models, the marginal likelihoods can be determined analytically [40]. Parameters μ, B and W are being estimated using available background data (e.g., Aitken & Lucy [40] 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. [55] 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 [58-60]. The extension proposed by Bozza et al. [55] 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:

$$Z_{ij} \sim N(\theta_i, W_i) \tag{9}$$

$$\theta_i \sim N(\mu, B) \tag{10}$$

$$W_i \sim IW(U, v),$$

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 v must be elicited. Bozza et al. [55] 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 v will allow to reduce(increase) the variability of the prior distribution.

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

$$(y|\theta_y, W_y) \sim N(\theta_y, W_y), \tag{11}$$

$$(x|\theta_x, W_x) \sim N(\theta_x, W_x). \tag{12}$$

The value of the evidence is given by the ratio of two marginal likelihoods under the competing propositions H_1 and H_2 as in (5).

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:

$$f(y, x|H_1) = \int_{\psi} f(y|\psi) f(x|\psi) f(\psi|H_1)\, d\psi, \tag{13}$$

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:

$$f(y, x|H_2) = \int_{\psi} f(y|\psi) f(\psi|H_2)\, d\psi \int_{\psi} f(x|\psi) f(\psi|H_2)\, d\psi. \tag{14}$$

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 (13) and (14) are no longer available in closed form since the integrals do not have an analytical solution. Bozza et al. [55] proposed to derive the marginal likelihood from the output of a Markov Chain Monte Carlo (MCMC) procedure, using Chib's method [61]. Other techniques such as bridge sampling [62] or importance sampling [63] can alternatively be used.

A schematic representation of the MVNIW model can be found in Figure 1. Note that the available background data is given by sincere, genuine material and is denoted by $\{Z_{ij}\}$.

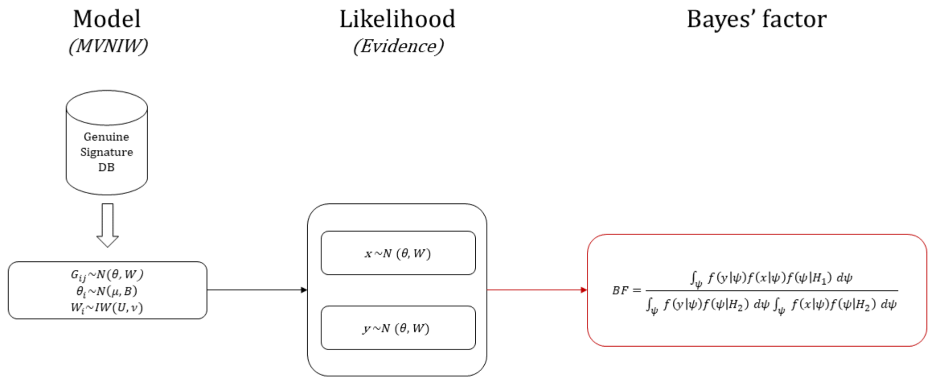

Figure 1 - Graphical Representation of the MVNIW Model. A database (DB) of genuine material $\{Z_{ij}\}$ *is available to estimate model parameters.*

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 (7) for models MVN/MVK and (14) 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

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

simplifies to

$$BF = \frac{f(y|x, H_1)}{f(y|H_2)}, \tag{15}$$

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 3.2

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 [47]. 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 defense 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.

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 [47-49].

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' [64-66] for instances where a simulator physically reproduces a signature. For the interested reader, a summary of these presentation attack types can be found in [67]. 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 [68-72]. 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 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 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 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 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.

Proposition	*Sub-Proposition*	*Scenario*	*Wording*	*Shortened name*
$\boldsymbol{H_1}$ *John Doe made the questioned signature*	$H_{1,1}$	Sincere	The questioned signature is a sincere signature made by John Doe	genuine signature
	$H_{1,2}$	Insincere Disguised	The questioned signature is a disguised signature made by John Doe	disguised signature
$\boldsymbol{H_2}$ *Someone else made the questioned signature*	$H_{2,1}$	Sincere Third Party	The questioned signature is a sincere signature made by a third party	randomly matching 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 1 – Generic Propositions for signature evaluation

2.2.2. The questioned signature model

As mentioned before, the existing proposals for handwriting examination (Section 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 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 (15) 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 (16).

$$BF_{Sig} = \frac{f(y, x|H_1)}{f(y, x|H_2)} = \frac{f(y|x, H_1)}{f(y|x, H_2)}. \tag{16}$$

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

$$Z_{ij} \sim N(\theta_i, W_i) \text{ and } S_{ij} \sim N(\theta_i, W_i). \tag{17}$$

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

$$\theta_i \sim N(\mu, \kappa W_i), \tag{18}$$

$$W_i \sim W(U, v), \tag{19}$$

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:

$$y \sim N(\theta_y, W_y) \text{ and } x \sim N(\theta_x, W_x). \tag{20}$$

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 3.2, to elicit the prior probability distributions.

The conditional distribution $f(y|x, H_1)$ is obtained as:

$$f(y|x, H_1) = \int_{\psi} f(y|\psi, H_1) f(x|\psi, H_1) f(\psi|\phi, H_1)\, d\psi \tag{21}$$

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 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 [71].

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 3.2. The conditional distribution $f(y|x, H_2)$ is obtained as:

$$f(y|x,H_2) = \int_{\psi} f(y|\psi,H_2) f(x|\psi,H_2)\, f(\psi|\phi,H_2)\, d\psi, \tag{22}$$

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 [71].

The Bayes' factor then becomes

$$BF_{Sig} = \frac{\int_{\psi} f(y|\psi,H_1) f(x|\psi,H_1) f(\psi|\phi,H_1)\, d\psi}{\int_{\psi} f(y|\psi,H_2) f(x|\psi,H_2)\, f(\psi|\phi,H_2)\, d\psi}. \tag{23}$$

A schematic representation of the model can be found in Figure 2.

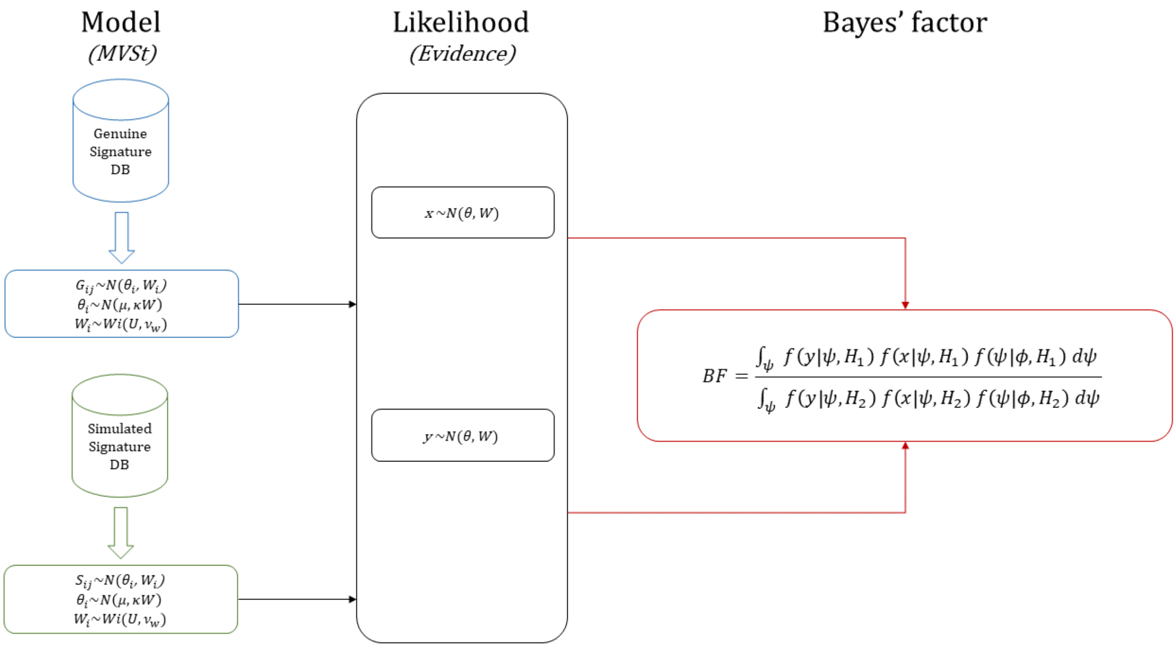

Figure 2 - Schematic representation of the MVSt Model. $\{G_{ij}\}$ and $\{S_{ij}\}$ are the genuine and simulated signature databases (DB).

3. Materials and Methods

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 [73].

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.

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 $i = 1, \dots, m$ and $j = 1, \dots, n$. 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 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 $i = 1, \dots, m$ and $j = 1, \dots, n$. 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.

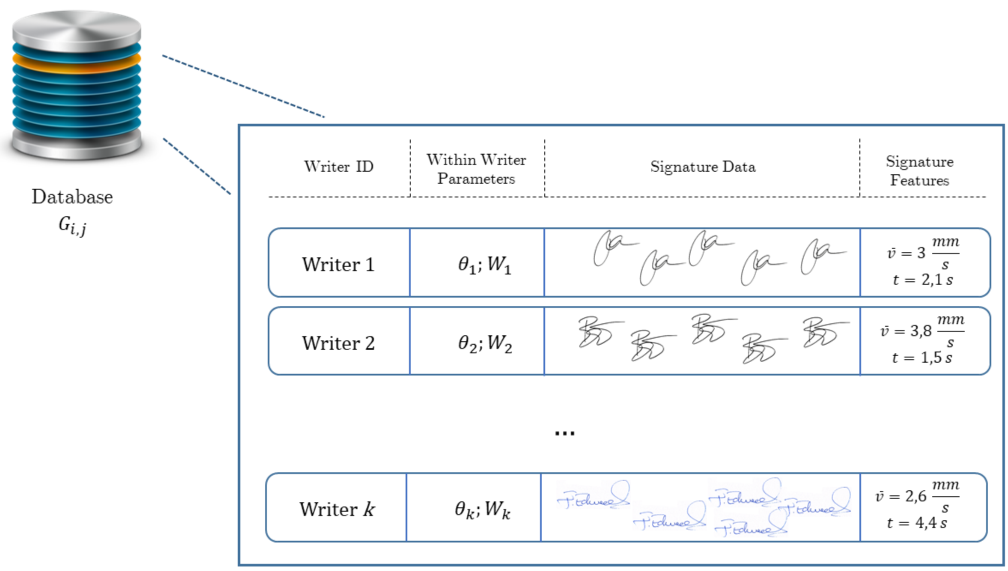

Figure 3 - Schematic representation of a signature population. The features $\bar{v}$ and t are the average pen speed and the signature duration respectively.

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 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 3. Every forger provided 10 forgeries.

Population	*Content*	*Role*
$\boldsymbol{y}$	Measurements from unknown source signature	Evidence
$\boldsymbol{x}$	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 2 - Summary of all data sources.

	Signature 1	*Signature 2*	*Signature 3*
# 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 3 - Summary of available data per signature; Specific datasets $\{Z_{i,j}\}$ and $\{S_{i,j}\}$.

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 [40, 56], 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 [55]. 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 are reproducible.

The selected signature features are summarized in Table 4. Features were selected as follows. A first feature set (Set N°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 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. (23) 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.

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 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	$\left\{\begin{matrix} 3, 5, 8, 10, 15, 20, 25, 30, 40, \\ 50, 60, 70, 75, 80, 90, 100 \end{matrix}\right\}$	$\{Genuine, Simulated\}$

Table 5 - Experimental condition summary.

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-Rate of Misleading Evidence).

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 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 4 and Figure 5, 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.

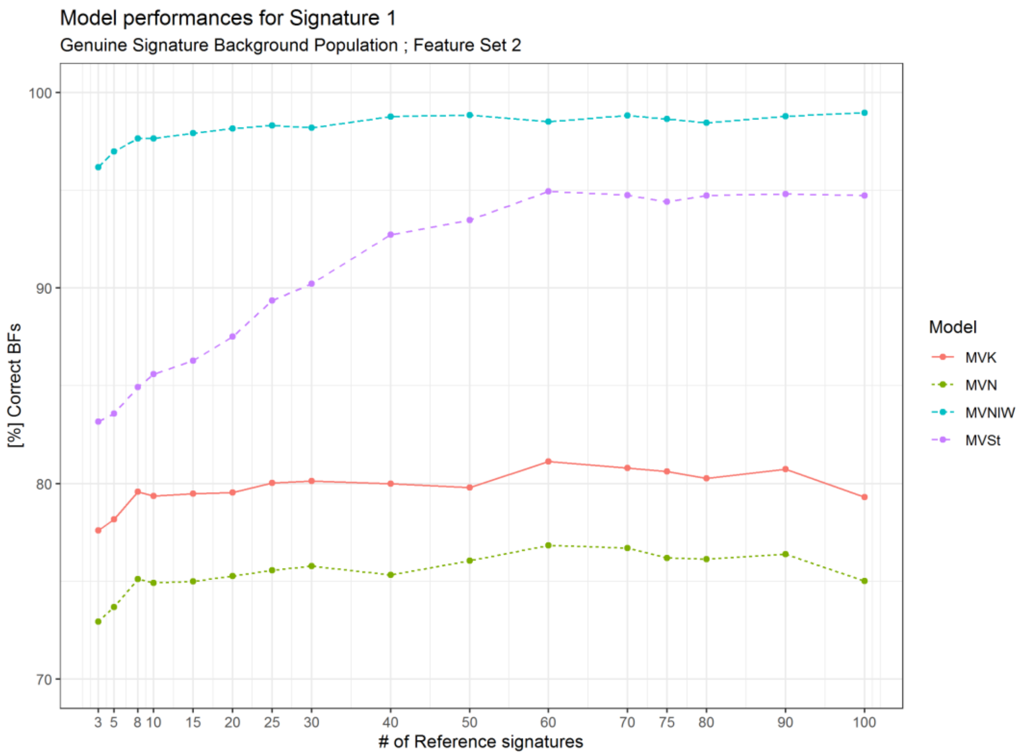

Figure 4 - Accuracies for Signature 1 and feature set 2

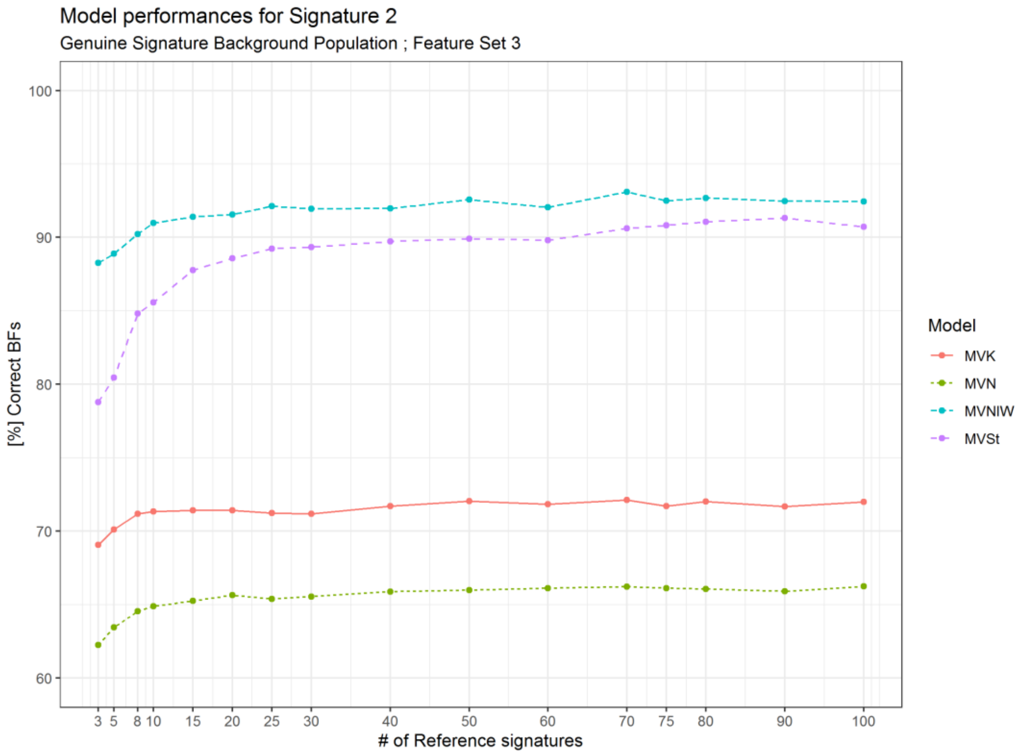

Figure 5 - Accuracies for Signature 2 with feature set 3

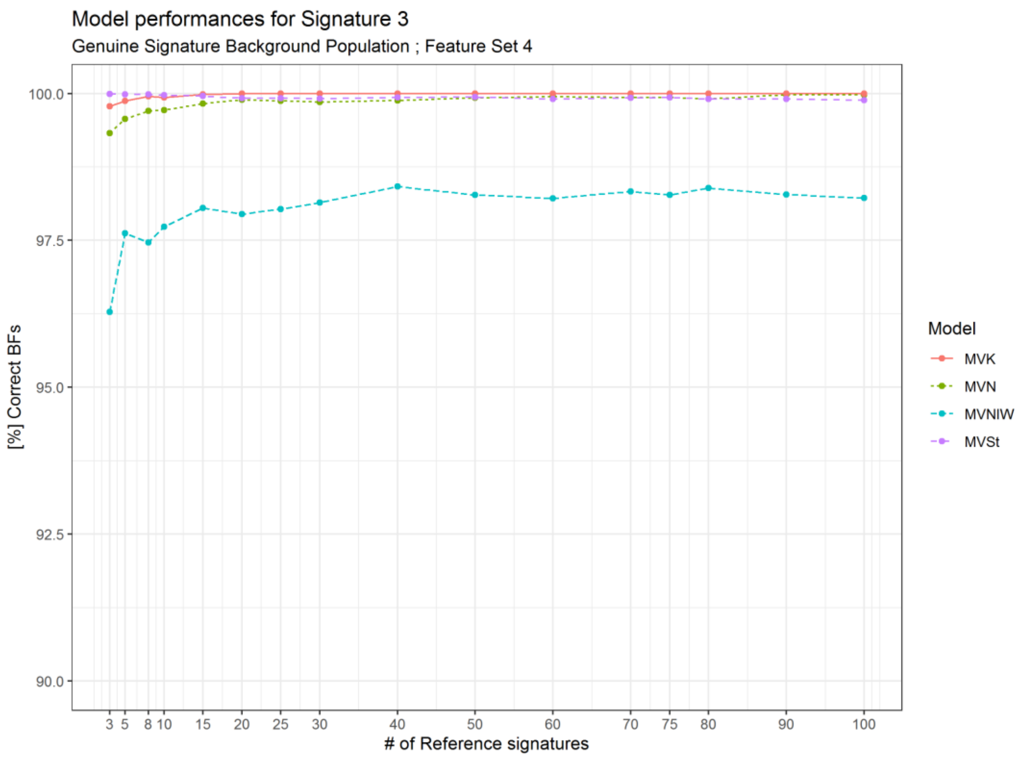

Figure 6 - Accuracies for Signature 3 with feature set 4

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 7 reports the results in terms of accuracy for signature 1. It can immediately 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.

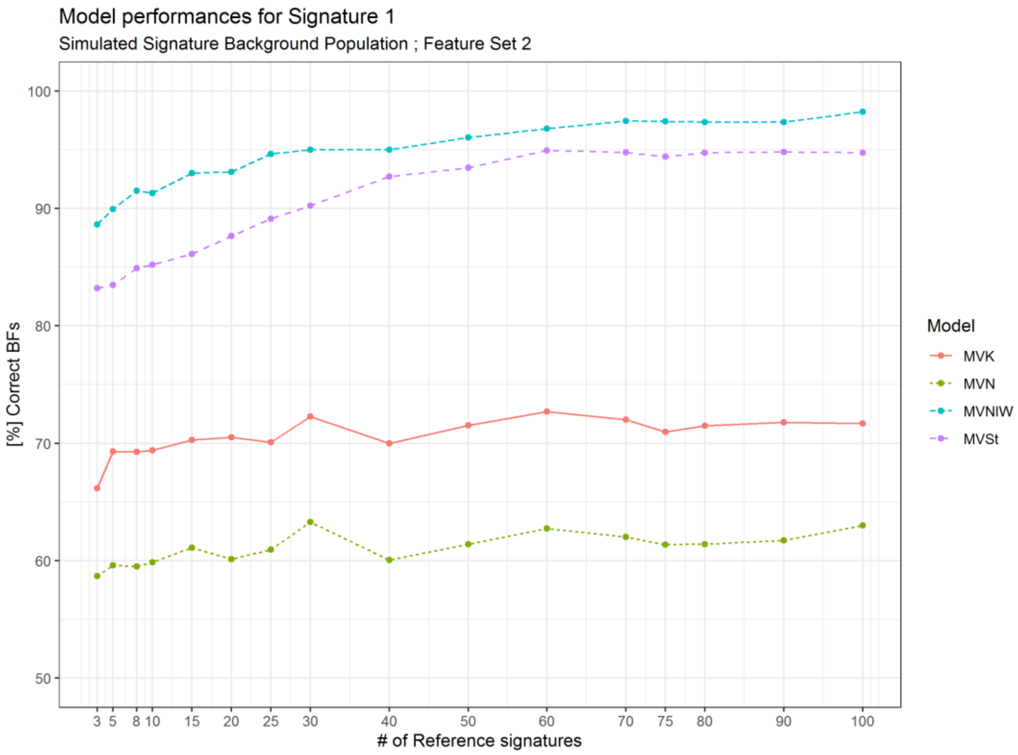

Figure 7 - Accuracies for Signature 1 with the s_x background data and feature set 2

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 [47]. 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.

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.

Bibliography

[1] National Research Council, Strengthening Forensic Science in the United States: A Path Forward, The National Academies Press, Washington, DC, 2009.
[2] Federal Bureau of Investigation, Comments on: President's Council of Advisors on Science and Technology REPORT TO THE PRESIDENT Forensic Science in Federal Criminal Courts: Ensuring Scientific Validity of Pattern Comparison Methods, FBI, 2016.
[3] 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, Office of Science and Technology Policy, 2016.
[4] G.S. Morrison, What should a forensic practitioner's likelihood ratio be? II, Science and Justice 57 (6) (2017) 472-476.
[5] G.S. Morrison, E. Enzinger, What should a forensic practitioner's likelihood ratio be?, Sci Justice 56 (5) (2016) 374-379.
[6] D.M. Risinger, M.P. Denbeaux, M.J. Saks, Exorcism of ignorance as proxy for rational knowledge: the lessons of handwriting identification "expertise", University of Pennsylvania Law Review 137 (1989) 731-792.
[7] D.M. Risinger, M. Saks, Science and Nonscience in the Courts, Daubert meets handwriting identification expertise, Iowa Law Review 82 (1) (1996) 21-82.
[8] M.J. Saks, D.L. Faigman, Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It, Annual Review of Law and Social Science 4 (1) (2008) 149-171.
[9] M.J. Saks, J.J. Koehler, The Individualization Fallacy in Forensic Science Evidence, Vanderbilt Law Review 61 (1) (2008) 199-219.
[10] J.L. Mnookin, Scripting expertise: The history of handwriting identification evidence and the judicial construction of reliability, Virginia Law Review 87 (2001) 1723-1845.
[11] L.J. Davis, C.P. Saunders, A. Hepler, J. Buscaglia, Using subsampling to estimate the strength of handwriting evidence via score-based likelihood ratios, Forensic Sci Int 216 (1-3) (2012) 146-57.
[12] A.B. Hepler, C.P. Saunders, L.J. Davis, J. Buscaglia, Score-based likelihood ratios for handwriting evidence, Forensic Sci Int 219 (1-3) (2012) 129-40.
[13] J.J. Miller, R.B. Patterson, D.T. Gantz, C.P. Saunders, M.A. Walch, J. Buscaglia, A Set of Handwriting Features for Use in Automated Writer Identification, J Forensic Sci 62 (3) (2017) 722-734.
[14] C. Neumann, 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, J Forensic Sci 60 (1) (2015) 252-6.
[15] C.P. Saunders, L.J. Davis, J. Buscaglia, Using automated comparisons to quantify handwriting individuality, J Forensic Sci 56 (3) (2011) 683-9.
[16] C.P. Saunders, L.J. Davis, A.C. Lamas, J.J. Miller, D.T. Gantz, Construction and evaluation of classifiers for forensic document analysis, The Annals of Applied Statistics 5 (1) (2011) 381-399.
[17] E. Enzinger, G.S. Morrison, 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, Sci Justice 56 (1) (2016) 42-57.
[18] G.S. Morrison, Calculation of forensic likelihood ratios: use of monte carlo simulations to compare the output of scorebased approaches with true likelihood-ratio values, In: Proceedings of the ICFIS 2014, Leiden University, Netherlands, Ed.(s) 2014,

[19] G.S. Morrison, E. Enzinger, Score based procedures for the calculation of forensic likelihood ratios - scores should take account of both similarity and typicality, Science and Justice 58 (1) (2018).
[20] J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, D. Ramos-Castro, J. Ortega-Garcia, Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems, Forensic Sci Int 155 (2-3) (2005) 126-40.
[21] D. Meuwly, D. Ramos, R. Haraksim, A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation, Forensic Sci Int 276 (2017) 142-153.
[22] D. Ramos, J. Gonzalez-Rodriguez, G. Zadora, C. Aitken, Information-theoretical assessment of the performance of likelihood ratio computation methods, J Forensic Sci 58 (6) (2013) 1503-18.
[23] D. Ramos, R.P. Krish, J. Fierrez, D. Meuwly, From Biometric Scores to Forensic Likelihood Ratios, in: M. Tistarelli, C. Champod (Eds.), Handbook of Biometrics for Forensic Science, Springer, 2017, pp. 305-327.
[24] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, A. Bromage-Griffiths, Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae, J Forensic Sci 52 (1) (2007) 54-64.
[25] J. Franco-Pedroso, D. Ramos, J. Gonzalez-Rodriguez, Gaussian Mixture Models of Between-Source Variation for Likelihood Ratio Computation from Multivariate Data, PLoS One 11 (2) (2016) e0149958.
[26] K. Franke, S.N. Srihari, Computational Forensics: An Overview, in: S.N. Srihari, K. Franke (Eds.) Computational Forensics, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 1-10.
[27] I.W. Evett, C.E.H. Berger, J.S. Buckleton, C. Champod, G. Jackson, Finding the way forward for forensic science in the US-A commentary on the PCAST report, Forensic Sci Int 278 (2017) 16-23.
[28] 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, A. Caliebe, A comment on the PCAST report: Skip the "match"/"non-match" stage, Forensic Sci Int 272 (2017) e7-e9.
[29] C.G.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, Second Edition, John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2004.
[30] A. Biedermann, S. Bozza, F. Taroni, C. Aitken, Reframing the debate: A question of probability, not of likelihood ratio, Sci Justice 56 (5) (2016) 392-396.
[31] A. Biedermann, P. Garbolino, F. Taroni, The subjectivist interpretation of probability and the problem of individualisation in forensic science, Science & Justice 53 (2) (2013) 192-200.
[32] F. Taroni, Inadequacies of posterior probabilities for the assessment of scientific evidence, Law, Probability and Risk 4 (1-2) (2005) 89-114.
[33] F. Taroni, P. Garbolino, A. Biedermann, C. Aitken, S. Bozza, Reconciliation of subjective probabilities and frequencies in forensic science, Law, Probability and Risk 17 (3) (2018) 243-262.
[34] C.E. Berger, J. Buckleton, C. Champod, I.W. Evett, G. Jackson, Evidence evaluation: a response to the court of appeal judgment in R v T, Sci Justice 51 (2) (2011) 43-9.
[35] A. Biedermann, T. Hicks, F. Taroni, C. Champod, C. Aitken, On the use of the likelihood ratio for forensic evaluation: response to Fenton et al, Sci Justice 54 (4) (2014) 316-8.
[36] J.S. Buckleton, C.M. Triggs, C. Champod, An extended likelihood ratio framework for interpreting evidence, Science & Justice 46 (2) (2006) 69-78.

[37] C. Champod, Fingerprint identification: advances since the 2009 National Research Council report, Philosophical Transactions: Biological Sciences 370 (1674) (2015) 1-10.
[38] C. Champod, I.W. Evett, G. Jackson, Establishing the most appropriate databases for addressing source level propositions, Science & Justice 44 (3) (2004) 153-164.
[39] D. Meuwly, Forensic Individualisation from Biometric Data, Science & Justice 46 (4) (2006) 205-213.
[40] C.G.G. Aitken, D. Lucy, Evaluation of trace evidence in the form of multivariate data, Appl. Statist. 53 (1) (2004) 109-122.
[41] B. Robertson, G.A. Vignaux, C.E.H. Berger, Interpreting Evidence: Evaluating Forensic Science in the Courtroom, John Wiley & Sons, Ltd, 2016.
[42] S. Srihari, C. Huang, H. Srinivasan, On the Discriminability of the Handwriting of Twins, (2007).
[43] S.N. Srihari, S.-H. Cha, H. Arora, S. Lee, Individuality of Handwriting, J Forensic Sci 47 (4) (2002) 1-17.
[44] S.N. Srihari, S.-H. Cha, S. Lee, The Discriminatory Power of Handwriting, in: P.B. Kantor, T. Kanungo, J. Zhou (Eds.) Document Recognition and Retrieval IX, Society of Photo Optical San Jose, CA, 2002.
[45] C. Viard-Gaudin, K. Das, S.N. Srihari, H. Srinivasan, R. Zanibbi, Questioned document workflow for handwriting with automated tools, 8297 (2012) 82970K-82970K-9.
[46] ENFSI, Best Practice Manual for the Forensic Examination of Handwriting Version 2, European Network of Forensic Science Institutes, http://enfsi.eu, 2018.
[47] ENFSI, ENFSI Guideline for evaluative reporting in forensic science, Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE), European Network of Forensic Science Institutes, http://enfsi.eu, 2015.
[48] Working Group on Statistics the Law of the Royal Statistical Society, Case Assessment and Interpretation of Expert Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Practitioner Guide No. 4, G. Jackson, C. Aitken, P. Roberts, Royal Statistical Society,, 2015, <https://rss.org.uk/membership/rss-groups-and-committees/sections/statistics-law/>.
[49] Working Group on Statistics the Law of the Royal Statistical Society, The Logic of Forensic Proof: Inferential Reasoning in Criminal Evidence and Forensic Science : Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Practitioner Guide No. 3, P. Roberts, C. Aitken, Royal Statistical Society,, 2014,
[50] J. Linden, R. Marquis, S. Bozza, F. Taroni, Dynamic signatures: A review of dynamic feature variation and forensic methodology, Forensic Science International 291 (2018) 216-229.
[51] C. Neumann, I.W. Evett, J. Skerrett, Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm, Journal of the Royal Statistical Society: Series A (Statistics in Society) 175 (2) (2012) 371-415.
[52] J.M. Curran, The Statistical Interpretation of Forensic Glass Evidence, International Statistical Review / Revue Internationale de Statistique 71 (3) (2003) 497-520.
[53] G.S. Morrison, J. Lindh, J.M. Curran, Likelihood ratio calculation for a disputed-utterance analysis with limited available data, Speech Communication 58 (2014) 81-90.
[54] L. Gaborini, A. Biedermann, F. Taroni, Towards a Bayesian evaluation of features in questioned handwritten signatures, Sci Justice 57 (3) (2017) 209-220.
[55] S. Bozza, F. Taroni, R. Marquis, M. Schmittbuhl, Probabilistic evaluation of handwriting evidence likelihood ratio for authorship, Applied Statistics. 57 (3) (2008) 329–341.
[56] C.G.G. Aitken, D. Lucy, Evaluation of trace evidence in the form of multivariate data Corrigendum, Appl. Statist. 53 (4) (2004) 665-666.

[57] J. Galbally, M. Gomez-Barrero, A. Ross, Accuracy evaluation of handwritten signature verification: Rethinking the random-skilled forgeries dichotomy, In: Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, Ed.(s) IEEE 2017, 302-310
[58] R. Huber, A. Headrick, Handwriting Identification: Facts and Fundamentals, CRC Press, Boca Raton, 1999.
[59] R. Morris, Forensic Handwriting Identification: Fundamental Concepts and Principles, 1st edition ed., Academic Press, 2000.
[60] K.M. Koppenhaver, Forensic Document Examination: Principles and Practice, Humana Press, 2007.
[61] S. Chib, I. Jeliazkov, Marginal Likelihood From the Metropolis–Hastings Output, Journal of the American Statistical Association 96 (453) (2001) 270-281.
[62] Q.F. Gronau, A. Sarafoglou, D. Matzke, A. Ly, U. Boehm, M. Marsman, D.S. Leslie, J.J. Forster, E.-J. Wagenmakers, H. Steingroever, A tutorial on bridge sampling, Journal of Mathematical Psychology 81 (2017) 80-97.
[63] P.W. Glynn, D.L. Iglehart, Importance Sampling for Stochastic Simulations, Management Science 35 (11) (1989) 1367-1392.
[64] R. Sanchez-Reillo, H.C. Quiros-Sandoval, I. Goicoechea-Telleria, W. Ponce-Hernandez, Improving Presentation Attack Detection in Dynamic Handwritten Signature Biometrics, IEEE Access 5 (2017) 20463-20469.
[65] International Organization for Standardization, ISO/IEC 30107-3:2017, Information technology -- Biometric presentation attack detection -- Part 3: Testing and reporting, International Organization for Standardization, 2017.
[66] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Presentation Attacks in Signature Biometrics: Types and Introduction to Attack Detection, 2019.
[67] M.I. Malik, Automatic Signature Verification Bridging the Gap between Existing Pattern Recognition Methods and Forensic Science, Department of Computer Science, Technische Universität Kaiserslautern, 2015.
[68] M.A. Ferrer, M. Diaz, C. Carmona-Duarte, A. Morales, A Behavioral Handwriting Model for Static and Dynamic Signature Synthesis, IEEE Trans Pattern Anal Mach Intell 39 (6) (2017) 1041-1053.
[69] J. Galbally, M. Diaz-Cabrera, M.A. Ferrer, M. Gomez-Barrero, A. Morales, J. Fierrez, On-line signature recognition through the combination of real dynamic data and synthetically generated static data, Pattern Recognition 48 (9) (2015) 2921-2934.
[70] J. Galbally, J. Fierrez, M. Martinez-Diaz, J. Ortega-Garcia, Evaluation of Brute-force Attack to Dynamic Signature Verification Using Synthetic Samples, In: Proceedings of the 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, Ed.(s) 2009, 131-135
[71] J. Galbally, J. Fierrez, J. Ortega-Garcia, R. Plamondon, Synthetic on-line signature generation. Part II: Experimental validation, Pattern Recognition 45 (7) (2012) 2622-2632.
[72] J. Galbally, R. Plamondon, J. Fierrez, J. Ortega-Garcia, Synthetic on-line signature generation. Part I: Methodology and algorithms, Pattern Recognition 45 (7) (2012) 2610-2621.
[73] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2015.