Bayesian evaluation of dynamic signatures in operational conditions

Linden J., Bozza S. Marquis R., Taroni F.

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 have become more commonplace. These so-called ‘dynamic signatures’ (also known as ‘digitally captured signatures’) are much different from paper-based signatures. Not only does the medium of recording differ, but also the type, volume of data and features are different from the pattern-based evidence that makes up 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 [1], 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 was 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, Questioned Document Examination, Bayes’ factor, Feature Selection, Biometrics, Multivariate data

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 [1-3], due to many advantages linked to paperless approaches [4, 5]. Along with the increasing popularity of dynamic signatures, forensic document examiners have increasingly been faced with cases involving such signature types. The role of the forensic scientist is to assess the value of the measurements of signature features under competing, alternative hypotheses as to the signatures authorship. Forensic scientists have to provide this information in a transparent and intelligible way [6-8], with many authors advocating probabilistic frameworks and the use of the Bayes' factor (BF) [7, 9-13].

Recent progress in handwriting examination has been fundamental in restoring trust in a severely critiqued discipline [14-17]. 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 [13, 18-39]. Reliability claims were supported by statistical analyses related to mock cases built from large datasets containing known source writing [40-44]. There have also been many advances in methods for validation criteria [45-47]. Further progress in feature characterization could be achieved through computational techniques [18, 19, 48-54]. Forensic handwriting examination has been able to take advantage of this research [31, 55-57], 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 [58, 59] and the availability of guidance for communicating conclusions to courts of law [7, 9, 12, 60], only few attempts have been devoted to the quantification of the value of dynamic signatures [61-68]. Research in this field is often focused on technical aspects such as features' acquisition rather than methodological ones [69]. Experts have been criticized because of the lack of objectivity of their conclusions [14-16, 70, 71]. Probabilistic models for handwriting evidence evaluation do exist [72], and can be adapted to be used for the forensic examination of digitally captured signatures [73].

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 2.1, while the probabilistic model is summarized in Section 3.2. An overview of the selected features, as well as a detailed description of the experimental design is presented in Section 2.3 and 2.4, respectively. Data analysis and results are reported in Section 4 with attention to the model's accuracy and reproducibility faced with (i) limited control material (Section 3.1), (ii) alternative approaches for the joint evaluation of multiple features (Section 4.2), and (iii) alternative feature sets (Section 4.3). A summary of the research findings and of the benefits of a probabilistic model for digitally captured features evaluation is discussed in Section 4.

2. Materials and Methods

In this section, the data used in the study (Section 3.1) as well as the probabilistic model for the evaluation of features (Section 3.2) and the type of signature features are described. Finally, experimental designs for studying the impact of data sparsity, joint evaluation of multiple features and feature's selection are presented (Section 3.4).

2.1. Data collection

All signatures were acquired in standardized conditions. Participants were sitting on an adjustable office chair at a desk approximately 1m high. 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 were classified into: 1) genuine non-case-related signatures, 2) reference signatures (genuine case-related signatures) and 3) simulated signatures (known source forgeries).

1) 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 $\boldsymbol{z}_{ij} = \left(z_{ij1}, \dots, z_{ijp}\right)$, where $i = 1, \dots, m$ is the writer's identifier, $j = 1, \dots, n$ is the number of signatures collected for each writer and p is the number of observed variables. This dataset is denoted by b_g.

2) Reference signatures

For the reference signatures, three signers who did not participate in the acquisition of non-case-related signatures, were asked to sign regularly during a period of 18 months, in order to capture natural long- and short-term variations. Signers were chosen so as to have different styles represented: a "text-based" (signature 1), a "stylized" (signature 2) and a "mixed" (signature 3) signature [67]. Their signatures were also characterized by a different graphical and dynamic complexity. These signatures were used to generate fictive cases.

3) 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 1. These signatures will serve as part of the evidence (so as to have fictive cases where the proposition according to which the questioned signature is non-genuine) or as background data that is denoted $\boldsymbol{s}_{ij} = \left(s_{ij1}, \dots, s_{ijp}\right)$, where $i = 1, \dots, m$ is the forger's identifier, $j = 1, \dots, n$ is the number of delivered attempts (forgeries) and p is the number of observed variables. This information is shortened to s_x.

Denoted by x, the measurements on selected features characterizing one of the three genuine signatures were used to generate fictive cases. Denoted by y, the measurements on selected features characterizing a questioned signature were 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 1.

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

Table 1 - Summary of available case-related data.

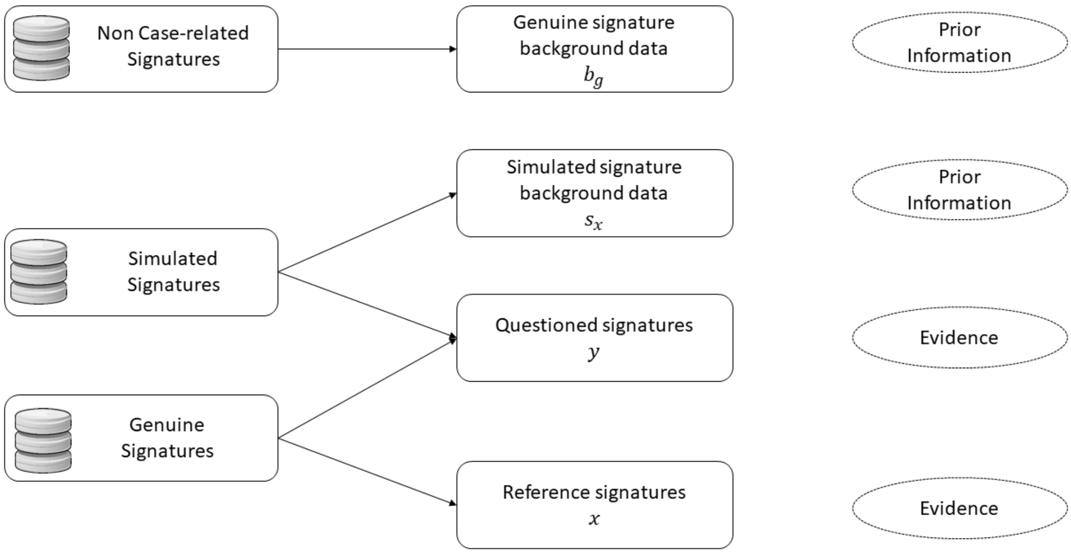

Figure 1 – Datasets used in this study (column 1) and their role in the evaluation of evidence computation (columns 2 and 3).

Every experiment involves the generation of 'random trials' from known source data. For every trial, there are drawn a total number of $q = l + m$ questioned signatures and a total number of r reference signatures. Questioned signatures are drawn (without replacement) from both simulated signatures $(l = 100)$ 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 (see Section 3.2). A Bayes' factor is then calculated for each one of the q fictive cases. 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.

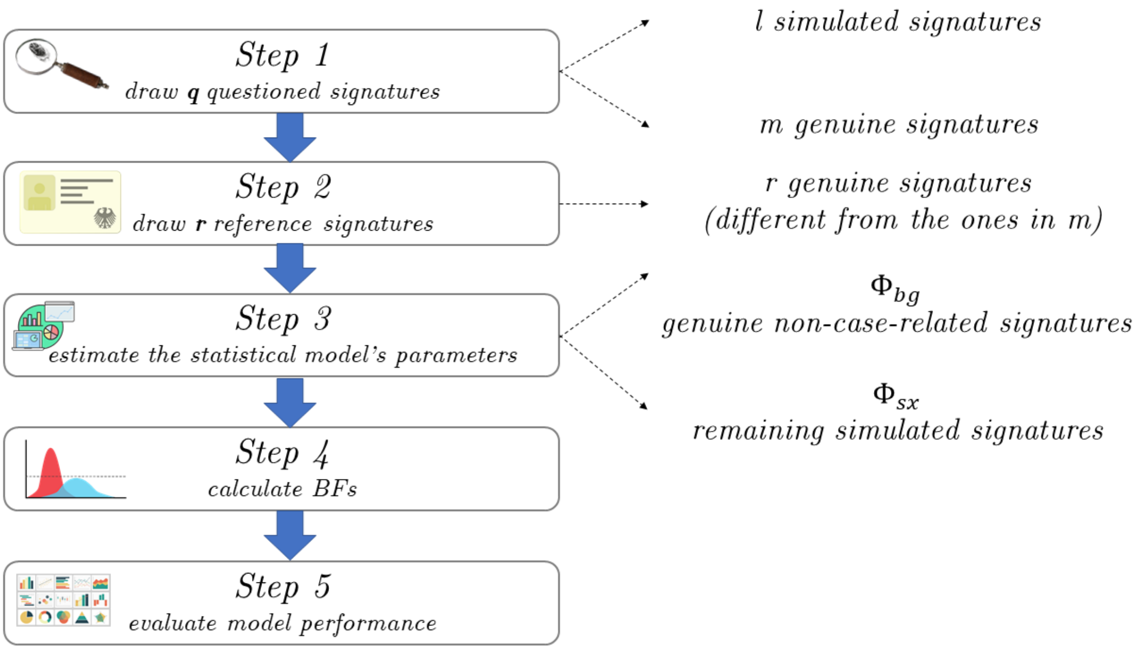

Figure 2 – Procedural description of one random trial, focusing on sampling and the parameter extraction. One random trial can be seen as being constituted of $q = l + m$ separate mock cases.

2.2. A probabilistic 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 can be used in most signature analysis cases:

- H_1: The signature on the questioned text (e.g. a contract) is a genuine signature made by a given source;
- H_2: The signature on the questioned 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}, $Z_{ij}{\sim}N(\theta_i, W)$ and $S_{ij}{\sim}N(\theta_i, W)$, where θ_i is the mean vector and W is the covariance matrix. A conjugate prior distribution is assumed for the model parameters (θ_i, W), that is

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

$$W{\sim}Wi(U, v),$$

(see [73] for a detailed description of the model). 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 $\boldsymbol{z}_{ij}$ (b_g, H_1 is true) or to simulated signatures $\boldsymbol{s}_{ij}$ (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, $(y|\theta, W) \sim N(\theta, W)$ and $(x|\theta, W) \sim N(\theta, W)$.

The Bayes' factor can be obtained as:

$$BF = \frac{\int_\psi f(y|\psi, H_1)\times f(\psi|x, \phi_{b_g}, H_1)\, d\psi}{\int_\psi f(y|\psi, H_2)\times f(\psi|x, \phi_{s_x}, H_2)\, d\psi}. \quad (1)$$

2.3. Features

In biometrics, features are classified into three broad categories [74]: global features, local features and segment features (see Table 1). According to Richiardi et al. [74] one is faced with global features if 'one feature is extracted for a whole signature, based on all sample points in the input domain' ([74] 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 [75]. 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 [61] 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.

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 aggregation 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.
SEGMENTAL FEATURE	One measurement per defined unit	The data is segmented according to a specific criterion. Every section has their own 'global' and 'local' features and can be compared unit by unit. Two examples are strokes (pen-ups) and lines between stops (velocity inversions).

Table 1 - Biometric feature type classification scheme.

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 3.4.

FEATURE	DESCRIPTION	FEATURE CLASS
HEIGHT	Height, measured vertically from minimum to maximum point	**Expansion Feature**
WIDTH	Width, measured horizontally from left- to right-most point	**Expansion Feature**
WH RATIO	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 Total time	**Time Feature**
UPTOT	Ratio of Up- to Total time	**Time Feature**
DOWNUP	Ratio of Down- to Uptime	**Time Feature**

Table 2 - 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 3 - Measured local features.

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 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) [9]. 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). This is quantified by:

$$\frac{s_A}{|\bar{x}_A|}, \qquad (2)$$

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 (2), 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 10'000 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 [34, 35] 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 [76, 77] and the assumption of independence can hardly be justified. Features showing a better performance when considered singularly were 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 3.2. 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 problem is the selection of discriminative features for the detection of simulated signatures. As no perfect algorithmic solution to feature selection exists, all trivariate feature combinations (34'220 possibilities) of the features in Table 2 and Table 3 were evaluated. The model dimensionality is set equal to 3 variables to keep the computational time short. The feature sets' performances were 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.

3. Results

3.1. Experiment 1 – The effect of data sparsity

The four best performing feature sets were selected for each reference signature. It must be highlighted that they were different for every signature. The average accuracy of the selected feature sets for the three reference signatures is represented in Figures 3 to 5, while the accuracy variance is represented in Figures 6 to 8.

Most feature sets have accuracy above 90%, with few exceptions where the reference signature's size is small (see Figures 5 and 6). 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 when more than 40 signatures are used for training. In general, a total number of 10 to 15 signatures appear as sufficient material 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 the presence of graphically different signatures, a viable alternative could be to divide them into 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.

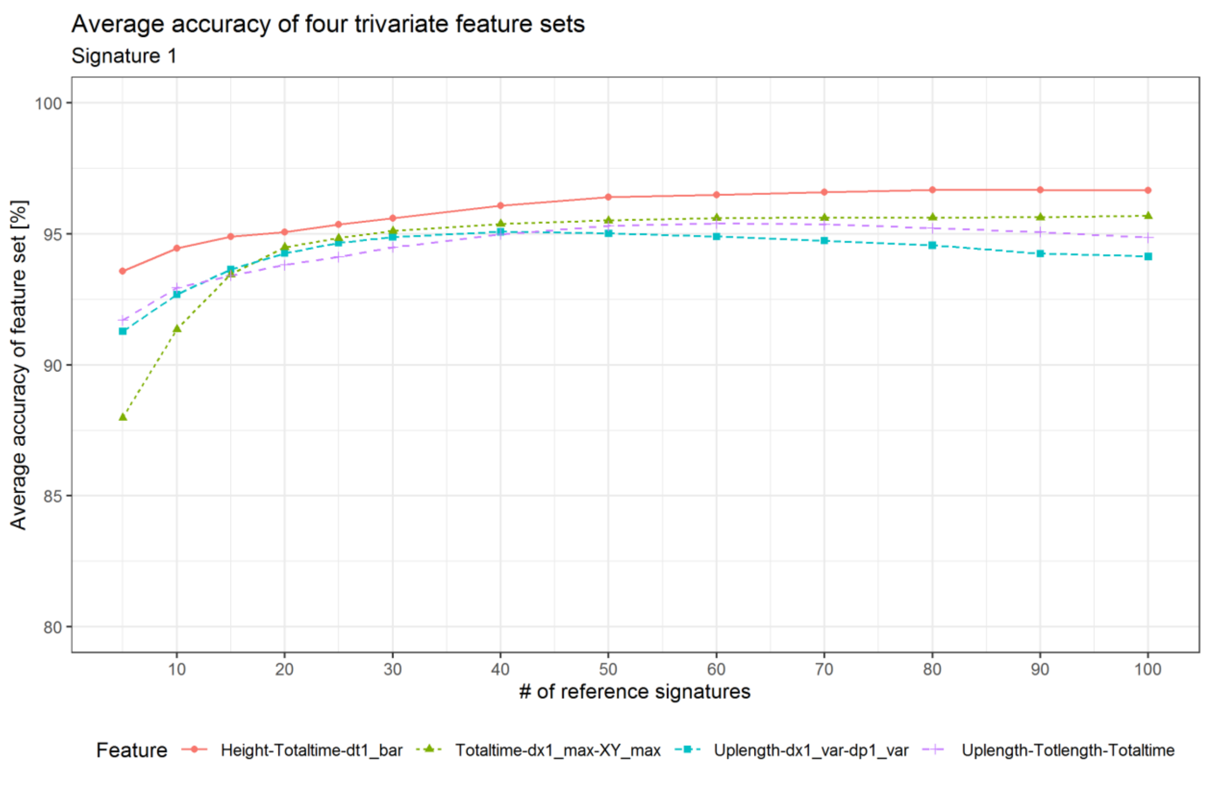

Figure 3 - Average accuracy for the four best performing trivariate feature sets over the 10'000 trials per experimental conditions for signature 1. For more detail on the features, see tables 2 and 3.

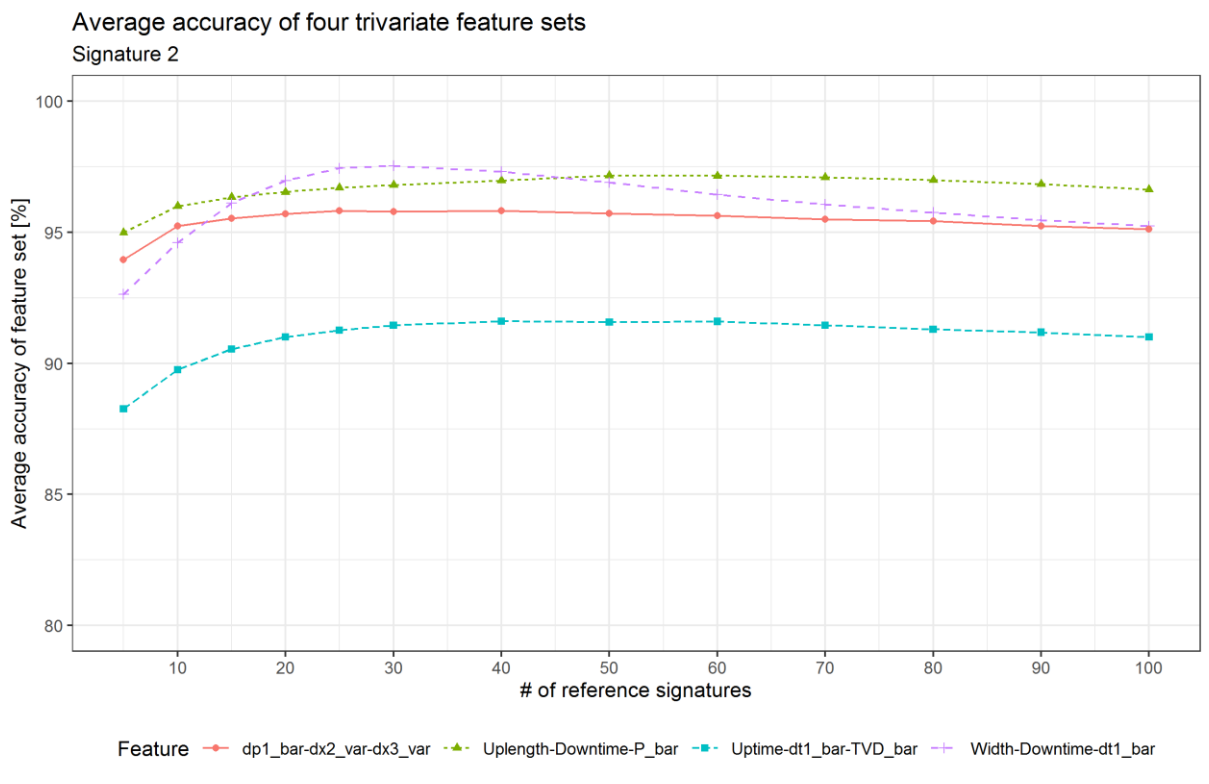

Figure 4 - Average accuracy for the four best performing trivariate feature sets over the 10'000 trials per experimental conditions for signature 2. For more detail on the features, see tables 2 and 3.

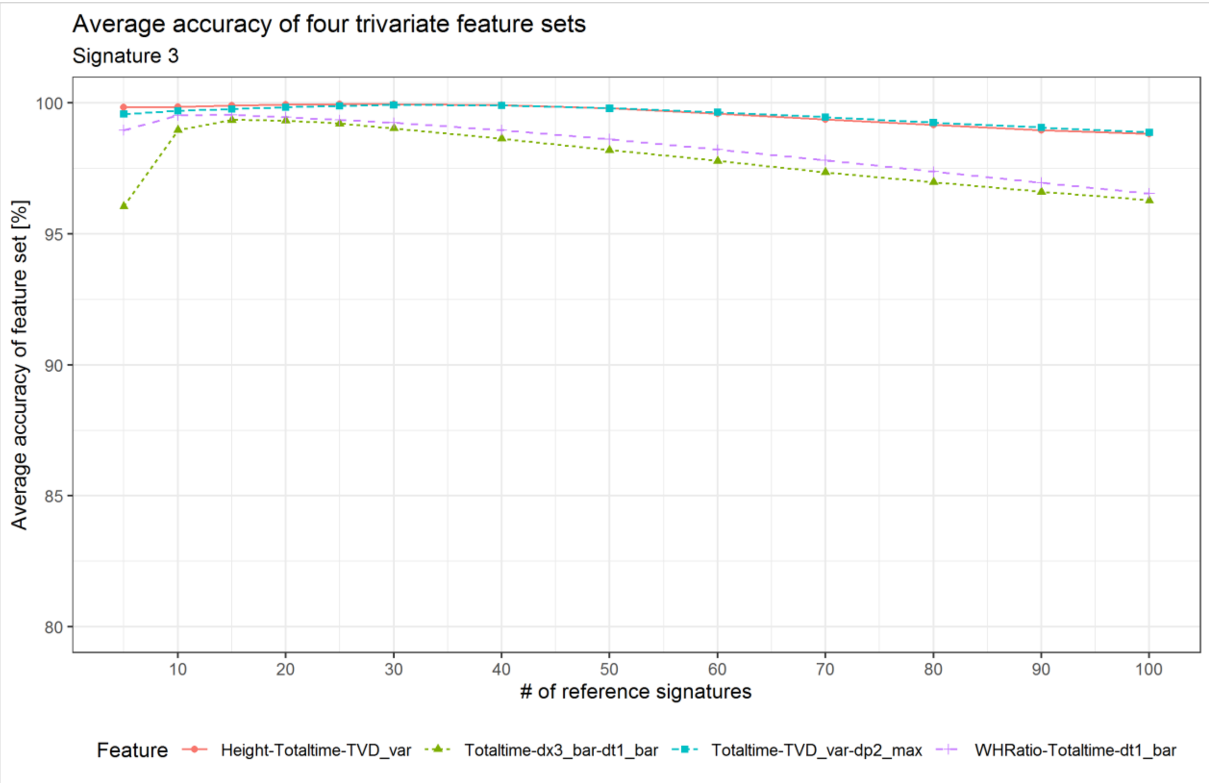

Figure 5 - Average accuracy for the four best performing trivariate feature sets over the 10'000 trials per experimental conditions for signature 3. For more detail on the features, see tables 2 and 3.

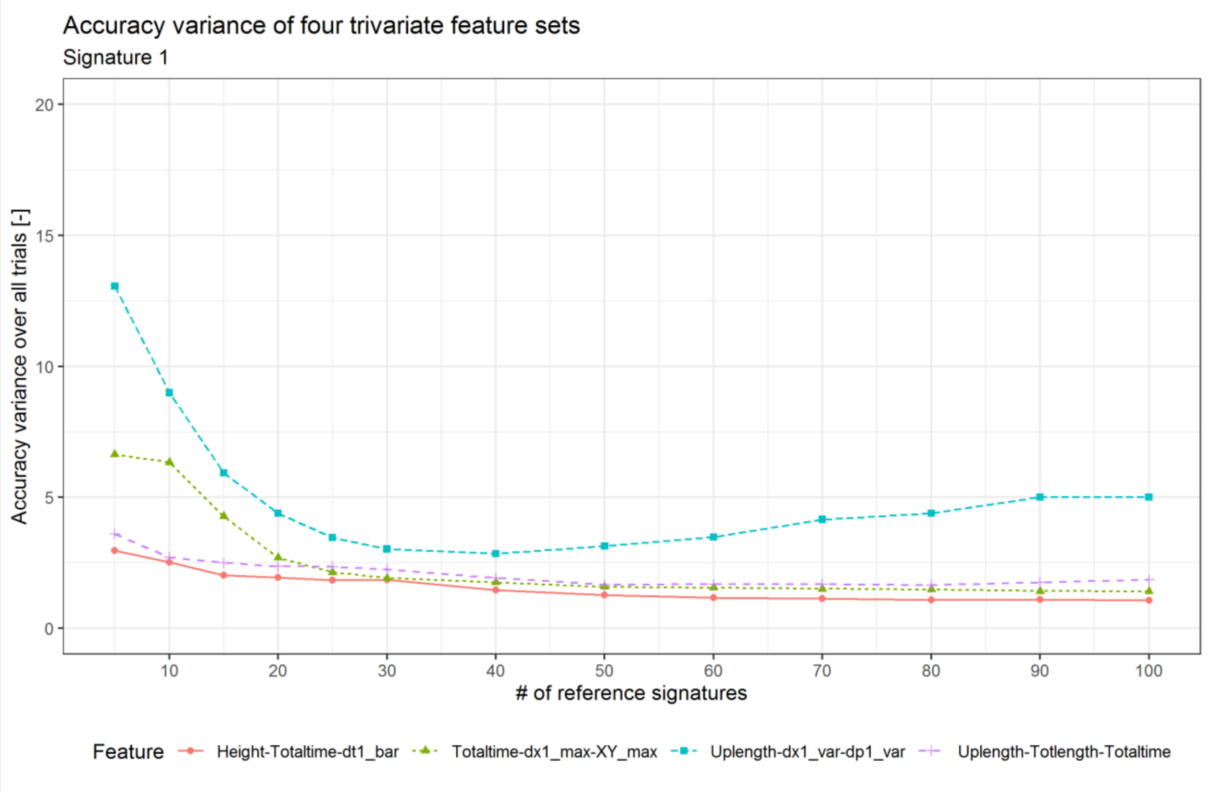

Figure 6 – Accuracy variance for the four best performing trivariate feature sets over the 10'000 trials per experimental conditions for signature 1. For more detail on the features, see tables 2 and 3.

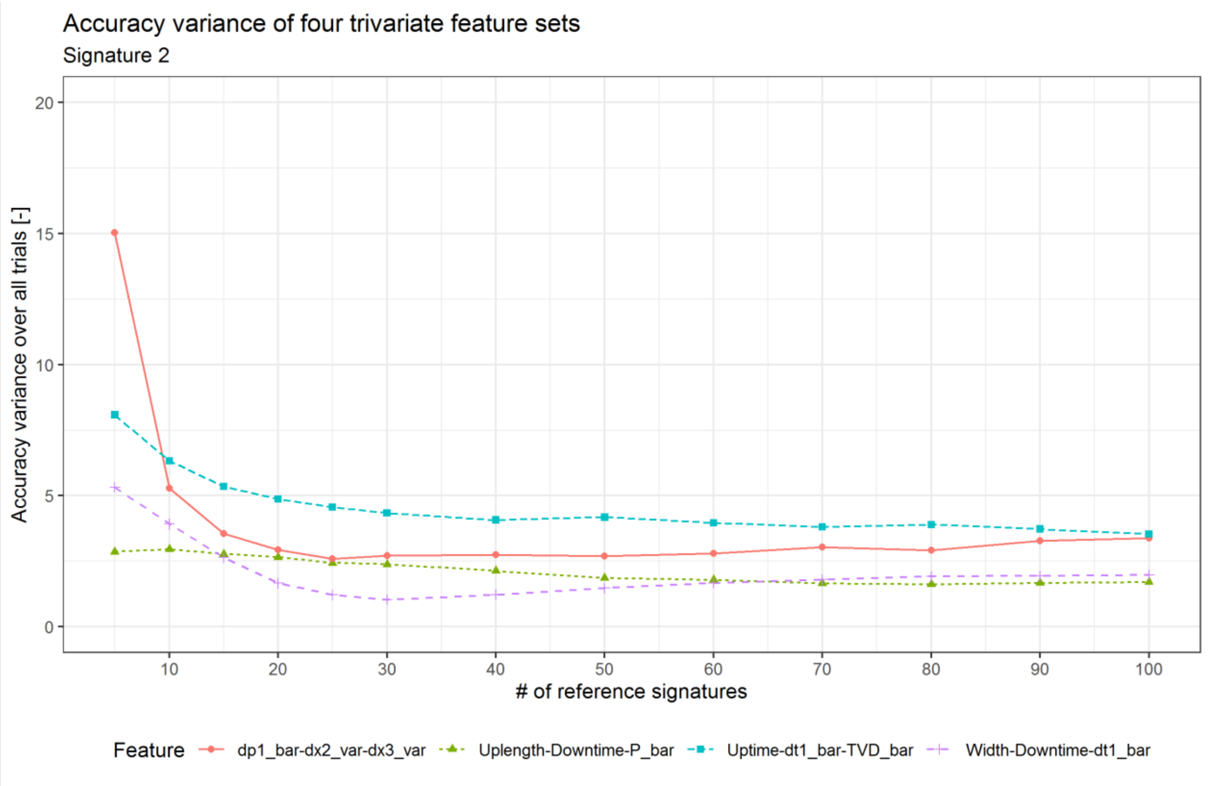

Figure 7 - Accuracy variance for the four best performing trivariate feature sets over the 10'000 trials per experimental conditions for signature 2. For more detail on the features, see tables 2 and 3.

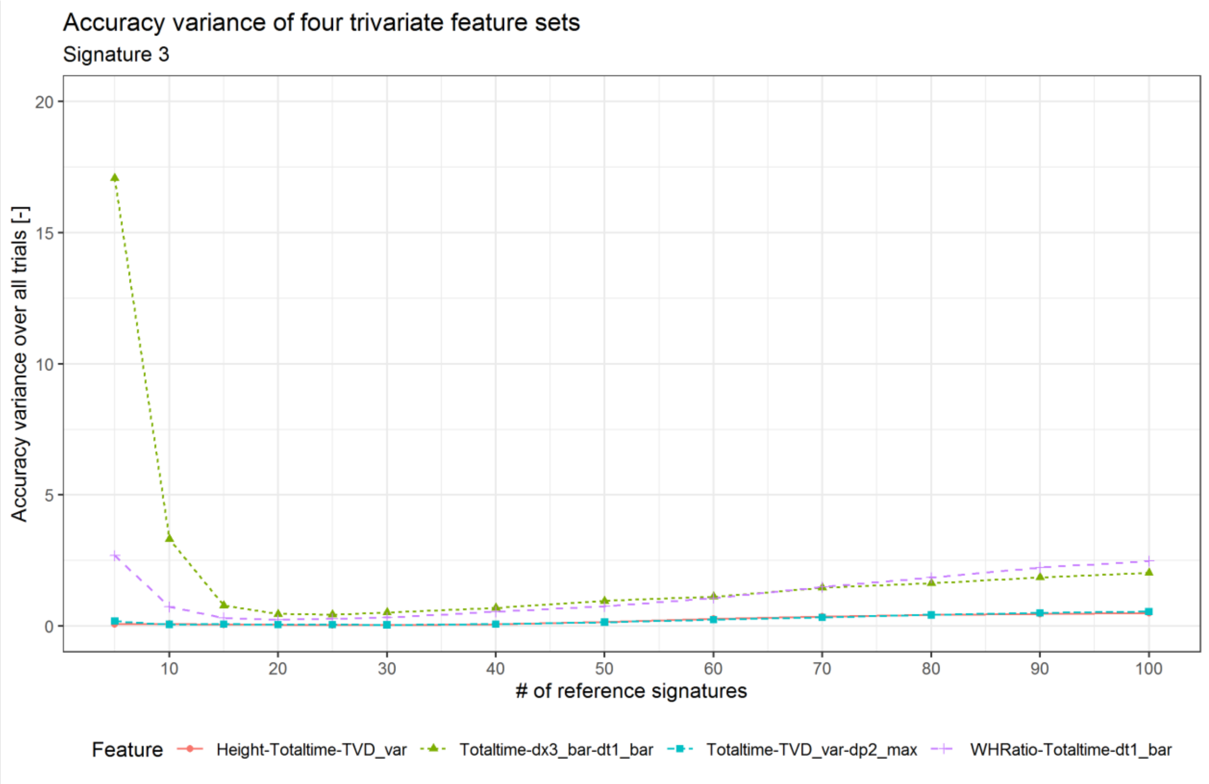

Figure 8 - Accuracy variance for the four best performing trivariate feature sets over the 10'000 trials per experimental conditions for signature 3. For more detail on the features, see tables 2 and 3.

3.2. Experiment 2 – Model dimension and combination of features

Forensic examiners have always argued that no single element in signatures is sufficient for detection of simulated signatures [59], and that multiple features should be observed and combined. Intuitively, one may expect that using more information always yields 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 4 to 6. 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 by about 2% when comparing univariate versus quadrivariate feature sets, while the 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 dimensions 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). An increasing variability is however 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.

	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 4 - Signature 1: model performances with feature sets of increasing dimension. Accuracy is the complement to the weighted average of RME.

	Univariate	**Bivariate**	**Trivariate**	**Quadrivariate**
Feature set	dt1_bar	P_bar, dy1_var	Totaltime, dx3_var, dt2_var	Totaltime, dx2_var, dx3_var, TAD_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 5 - Signature 2: model performances with feature sets of increasing dimension. Accuracy is the complement to the weighted average of RME.

	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 6 - Signature 3: model performances with feature sets of increasing dimension. Accuracy is the complement to the weighted average of RME.

The model performances reported in Tables 4 to 6 are related to experimental studies where the BF has been calculated as in (1) using the multivariate statistical model (see Section 3.2). 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 7 to 9. 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 Table 6, 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 from Tables 4 to 6. 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 highly related information. The decrease in value is an expected result 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, such an extreme BF is not indicative of high rate of reliability.

	Accuracy [%]	**RME(H1) [%]**	**RME(H2) [%]**
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 7 – 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).

	Accuracy [%]	**RME(H1) [%]**	**RME(H2) [%]**
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 8 - 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).

	Accuracy [%]	**RME(H1) [%]**	**RME(H2) [%]**
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 9 - 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).

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 pieces of information about the time of execution are highly correlated and the multiplicative approach yields BFs of increased magnitudes 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 7). 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 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 [9]. The so-called multiplicative strategy is not 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.

4.3. 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 10). The size of the control material 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 3.4. This approach has, however, some limits, as it cannot express the 'importance' of the contribution, nor directly express the complementarity of the features.

FEATURE TYPE	***SIGNATURE TYPE***		
	Mixed Style Medium complexity (signature 1)	*Stylized Low complexity (signature 2)*	*Text-based High complexity (signature 3)*
TIME FEATURES	100,00%	72,00%	46,22%
LENGTH FEATURES	0,00%	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 10 - 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.

The contribution of feature types varied with the reference signature and its complexity. If time-based 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 from the feature type. This appears clearly in Tables 11 and 12, 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. Information on the chosen features (time, pressure, speed, direction) is available on most dynamic signature hardware.

Although the results in Table 11 may seem to suggest that these feature sets work well, they are not optimal. The rates of misleading evidence vary a lot. A signature-specific feature selection is needed. As an example, Signature 3 had trivariate feature sets with 99.9% accuracy. These results are not reproducible by using the same feature set on other signatures.

Optimal performance is only achieved through a signature-specific feature selection. However, some feature sets appear as good compromises between applicability and performance results. Experiments confirm 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 11 – Performances of feature set including Totaltime, dt1_bar, TVD_bar. The accuracy is the opposite of the weighted average of the RMEs. This is a result as the number of samples between H_1 and H_2 was different.

	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 – Performances of feature set including Totaltime, P_var, dp3_max. The accuracy is the opposite of the weighted average of the RMEs. This is a result as the number of samples between H_1 and H_2 was different.

4. Conclusion

The study explored evidence evaluation in dynamic signature examination under operational conditions. Three specific 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. 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 the 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 produce more extreme (and unjustified) 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. Stylized signatures may perform better with dynamic features, while longer signature types may show to perform better 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 feature 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 were exclusively made of 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 mentioning is that dimensionality plays an important role in multivariate statistics [78]; 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. Alternatively, a score-based approach might be implemented. This would allow one to shrink the model dimensions and it may turn out to be a viable alternative to dealing 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 of 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 [6]. The model is versatile and applicable under operational conditions, and it may represent a valuable contribution to practice in handwriting examination. It may help the forensic handwriting examiner in providing - through the Bayes' factor - a quantified support in favor of or against a set of hypotheses about the origin of questioned documents. This numerical value guarantees that examiner's conclusions are sound because they are supported by strong scientific principles.

5. Bibliography

[1] Wacom Europe GmbH, Wacom Europe GmbH Business Solutions / Case Studies. <https://www.wacom.com/fr-fr/enterprise/business-solutions>, 2019 (accessed 08.04.2019.2019).
[2] Step Over International, Step Over International References. <https://www.stepover.com/references/>, 2019 (accessed 08.04.2019.2019).
[3] Topaz Systems Incorporated, Topaz Systems Inc. Case Studies. <https://www.topazsystems.com/casestudies.html>, 2019 (accessed 08.04.2019.2019).
[4] J. Lenz, Taking dynamic signatures seriously, Biometric Technology Today 2011(10) (2011) 9-11.
[5] J.-M. Lenz, Dynamic Signature Recognition : A sign of the times, Biometric Technology Today 17(2) (2008) 10-11.
[6] G. Edmond, B. Found, K. Martire, K. Ballantyne, D. Hamer, R. Searston, M. Thompson, E. Cunliffe, R. Kemp, M. San Roque, J. Tangen, R. Dioso-Villa, A. Ligertwood, D. Hibbert, D. White, G. Ribeiro, G. Porter, A. Towler, A. Roberts, Model forensic science, Australian Journal of Forensic Sciences 48(5) (2016) 496-537.
[7] B. Robertson, G.A. Vignaux, C.E.H. Berger, Interpreting Evidence: Evaluating Forensic Science in the Courtroom, John Wiley & Sons, Ltd2016.
[8] G. Jackson, D. Kaye, C. Neumann, A. Ranadive, V. Reyna, Communicating the Results of Forensic Science Examinations, in: C. Neumann, A. Ranadive, D.H. Kaye (Eds.) National Institute for Science and Technology (NIST), 2015.
[9] 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.
[10] Working Group on Statistics the Law of the Royal Statistical Society, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings : Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Practitioners Guide No 1, in: C. Aitken, P. Roberts, G. Jackson (Eds.) Royal Statistical Society,, https://rss.org.uk/membership/rss-groups-and-committees/sections/statistics-law/, 2010, p. 122.
[11] 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.
[12] I.W. Evett, Towards a uniform framework for reporting opinions in forensic science casework, Science & Justice 38(3) (1998) 198-202.
[13] F. Taroni, R. Marquis, M. Schmittbuhl, A. Biedermann, A. Thiery, S. Bozza, The use of the likelihood ratio for evaluative and investigative purposes in comparative forensic handwriting examination, Forensic Sci Int 214(1-3) (2012) 189-94.
[14] 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.
[15] M. Saks, H. VanderHaar, On the "General Acceptance" of Handwriting Identification Principles, (2005).
[16] D.M. Risinger, M. Saks, Science and Nonscience in the Courts, Daubert meets handwriting identification expertise, Iowa Law Review 82(1) (1996) 21-82.
[17] J.L. Mnookin, Scripting expertise: The history of handwriting identification evidence and the judicial construction of reliability, Virginia Law Review 87 (2001) 1723-1845.
[18] 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.
[19] S. Srihari, Computational Methods for Handwritten Questioned Document Examination, U.S. Department of Justice, 2010.
[20] S. Srihari, L. Graham, Survey of Computer Methods in Forensic Handwritten Document Examination, Eleventh Internation Graphonomics Society Conference, Scottsdale, AZ, 2003, pp. 278-281.

[21] S. Srihari, C. Huang, H. Srinivasan, On the discriminability of the handwriting of twins, J Forensic Sci 53(2) (2008) 430-46.
[22] S.N. Srihari, S.-H. Cha, H. Arora, S. Lee, Individuality of Handwriting, J Forensic Sci 47(4) (2002) 1-17.
[23] 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.
[24] S.N. Srihari, L. Meng, L. Hanson, Development of Individuality in Children's Handwriting, J Forensic Sci 61(5) (2016) 1292-300.
[25] S.N. Srihari, K. Singer, Role of automation in the examination of handwritten items, Pattern Recognition 47(3) (2014) 1083-1095.
[26] 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.
[27] C. Viard-Gaudin, Y. Tang, S. Srihari, H. Srinivasan, R. Zanibbi, Handwriting individualization using distance and rarity, Document Recognition and Retrieval XIX, SPIE, Burlingame, California, United States, 2012, pp. 82970R-82970R-10.
[28] 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.
[29] 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.
[30] 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.
[31] C.P. Saunders, L.J. Davis, J. Buscaglia, Using automated comparisons to quantify handwriting individuality, J Forensic Sci 56(3) (2011) 683-9.
[32] 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.
[33] R.D. Stoel, E.V. den Heuvel, L.C. Alewijnse, Structural equation modelling of subjective measures with an application to forensic handwriting examination, Science & Justice 50(1) (2010) 31-32.
[34] M.E. Johnson, T.W. Vastrick, M. Boulanger, E. Schuetzner, Measuring the Frequency Occurrence of Handwriting and Handprinting Characteristics, J Forensic Sci 62(1) (2017) 142-163.
[35] T.W. Vastrick, E. Schuetzner, K. Osborn, Measuring the Frequency Occurrence of Handwritten Numeral Characteristics, J Forensic Sci 63(4) (2018) 1215-1220.
[36] R. Marquis, M. Schmittbuhl, W.D. Mazzella, 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, Forensic Sci Int 150(1) (2005) 23-32.
[37] B.V.K. Vijaya Kumar, A. Kholmatov, S. Prabhakar, B. Yanikoglu, A.A. Ross, An individuality model for online signatures using global Fourier descriptors, 6944 (2008) 694407.
[38] F. Taroni, R. Marquis, M. Schmittbuhl, A. Biedermann, A. Thiery, S. Bozza, Bayes factor for investigative assessment of selected handwriting features, Forensic Sci Int 242 (2014) 266-73.
[39] A. Thiéry, Développement d'un processus de quantification et d'évaluation de caractères manuscrits : théorie et applications, Ecole des Sciences Criminelles, Université de Lausanne, 2014.
[40] C. Bird, B. Found, D. Rogers, Forensic document examiners' skill in distinguishing between natural and disguised handwriting behaviors, J Forensic Sci 55(5) (2010) 1291-5.
[41] B. Found, D. Rogers, The probative character of Forensic Handwriting Examiners' identification and elimination opinions on questioned signatures, Forensic Sci Int 178(1) (2008) 54-60.
[42] M. Liwicki, C.E.v.d. Heuvel, B. Found, M.I. Malik, Forensic Signature Verification Competition 4NSigComp2010 - Detection of Simulated and Disguised Signatures, 12th International Conference on Frontiers in Handwriting Recognition International Conference on Frontiers in Handwriting Recognition (ICFHR-2010), Kolkata, India, 2010, pp. 715-720.
[43] M. Liwicki, M.I. Malik, L. Alewijnse, E.v.d. Heuvel, B. Found, ICFHR 2012 Competition on Automatic Forensic Signature Verification (4NsigComp 2012), (2012) 823-828.

[44] J. Sita, B. Found, D. Rogers, Forensic Handwriting Examiners Expertise for Signature Comparison, J Forensic Sci 47(5) (2002) 1117-24.
[45] R. Haraksim, D. Ramos, D. Meuwly, C.E. Berger, Measuring coherence of computer-assisted likelihood ratio methods, Forensic Sci Int 249 (2015) 123-32.
[46] 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.
[47] D. Ramos, R. Haraksim, D. Meuwly, Likelihood ratio data to report the validation of a forensic fingerprint evaluation method, Data Brief 10 (2017) 75-92.
[48] A.K. Jain, A. Ross, Bridging the gap: from biometrics to forensics, Philos Trans R Soc Lond B Biol Sci 370(1674) (2015).
[49] 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.
[50] D. Meuwly, Forensic Individualisation from Biometric Data, Science & Justice 46(4) (2006) 205-213.
[51] D. Meuwly, R. Veldhuis, Forensic biometrics: From two communities to one discipline, 2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), 2012, pp. 1-12.
[52] A. Nautsch, D. Meuwly, D. Ramos, J. Lindh, C. Busch, Making Likelihood Ratios Digestible for Cross-Application Performance Assessment, IEEE Signal Processing Letters 24(10) (2017) 1552-1556.
[53] 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, Springer2017, pp. 305-327.
[54] M. Tistarelli, E. Grosso, D. Meuwly, Biometrics in Forensic Science: Challenges, Lessons and New Technologies, Springer International Publishing, Cham, 2014, pp. 153-164.
[55] K. Franke, L.R. Schomaker, C. Veenhuis, C. Taubenheim, I. Guyon, L. Vuurpijl, M. Van Erp, G. Zwarts, WANDA: A Generic Framework applied in Forensic Handwriting Analysis and Writer Identification.
[56] D.T. Gantz, M.A. Walch, FLASH ID Handwriting Derived Biometric Analysis Software, Measurement Science and Standards in Forensic Handwriting Analysis Conference, NIST.
[57] D.T. Gantz, C.P. Saunders, Quantifying the Effects of Database Size and Sample Quality on Measures of Individualization Validity and Accuracy in Forensics, U.S. Department of Justice, 2015.
[58] B. Found, C. Bird, The Modular Forensic Handwriting Method, Journal of Forensic Document Examination 26(1) (2016) 7-83.
[59] R. Huber, A. Headrick, Handwriting Identification: Facts and Fundamentals, CRC Press, Boca Raton, 1999.
[60] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A model for case assessment and interpretation, Science & Justice 38(3) (1998) 151-156.
[61] H. Harralson, Developments in handwriting and signature identification in the digital age, Routledge, Taylor & Francis Group, London, 2015.
[62] H.H. Harralson, Forensic Document Examination of Electronically Captured Signatures, Digital Evidence & Elec. Signature L. Rev. 9 (2012) 67-73.
[63] K.A. Nicolaides, Using acceleration plots in the forensic examination of electronically captured signatures, Journal of the American Society of Questioned Document Examiners 15(2) (2012) 29-43.
[64] W. Flynn, Executing a forensic examination of electronically captured signatures, Journal of the American Society of Questioned Document Examiners 15(1) (2012) 3-10.
[65] H. Harralson, L. Miller, Huber and Headrick's Handwriting Identification: Facts and Fundamentals, CRC Press, Boca Raton, 2018.
[66] M.P. Caligiuri, L.A. Mohammed, B. Found, D. Rogers, Nonadherence to the isochrony principle in forged signatures, Forensic Sci Int 223(1-3) (2012) 228-32.
[67] L. Mohammed, B. Found, M. Caligiuri, D. Rogers, Dynamic characteristics of signatures: effects of writer style on genuine and simulated signatures, J Forensic Sci 60(1) (2015) 89-94.

[68] L.A. Mohammed, B. Found, M. Caligiuri, D. Rogers, The dynamic character of disguise behavior for text-based, mixed, and stylized signatures, J Forensic Sci 56 Suppl 1 (2011) S136-41.
[69] J. Linden, R. Marquis, S. Bozza, F. Taroni, Dynamic signatures: A review of dynamic feature variation and forensic methodology, Forensic Sci Int 291 (2018) 216-229.
[70] 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.
[71] M.J. Saks, J.J. Koehler, The Individualization Fallacy in Forensic Science Evidence, Vanderbilt Law Review 61(1) (2008) 199-219.
[72] S. Bozza, F. Taroni, R. Marquis, M. Schmittbuhl, Probabilistic evaluation of handwriting evidence likelihood ratio for authorship, Applied Statistics. 57(3) (2008) 329–341.
[73] J. Linden, F. Taroni, R. Marquis, S. Bozza, Bayesian multivariate models for case assessment in dynamic signature cases, Forensic Sci Int 318 (2021) 110611.
[74] J. Richiardi, H. Ketabdar, A. Drygajlo, Local and global feature selection for on-line signature verification, in: B. Werner (Ed.) Eighth International Conference on Document Analysis and Recognition, IEEE Computer Society, Seoul, Korea, 2005, pp. 625-629.
[75] D. Impedovo, G. Pirlo, Automatic Signature Verification: The State of the Art, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38(5) (2008) 609-635.
[76] S. Edelman, T. Flash, A model of handwriting, Biological Cybernetics 57(1) (1987) 25-36.
[77] P. Viviani, C. Terzuolo, Trajectory determines movement dynamics, Neuroscience 7(2) (1982) 431-437.
[78] G. Zadora, A. Martyna, D. Ramos, C. Aitken, Statistical Analysis in Forensic Science Evidential Value of Multivariate Physicochemical Data, 1st Edition ed., John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom, 2014.

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.