PLOS ONE

RESEARCH ARTICLE

# Classification and Verification of Handwritten Signatures with Time Causal Information Theory Quantifiers

**Osvaldo A. Rosso**[1,2,3]*, **Raydonal Ospina**[4], **Alejandro C. Frery**[5]

**1** Instituto de Física, Universidade Federal de Alagoas (UFAL), Maceió, AL, Brazil, **2** Instituto Tecnológico de Buenos Aires (ITBA), and CONICET, Ciudad Autónoma de Buenos Aires, Argentina, **3** Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Santiago, Chile, **4** Centro de Ciências Exatas e da Natureza, Departamento de Estatística, Universidade Federal de Pernambuco (UFPE), Recife, PE, Brasil, **5** Laboratório de Computação Científica e Análise Numérica, Universidade Federal de Alagoas (UFAL), Maceió, AL, Brazil

* oarosso@gmail.com





**Citation:** Rosso OA, Ospina R, Frery AC (2016) Classification and Verification of Handwritten Signatures with Time Causal Information Theory Quantifiers. PLoS ONE 11(12): e0166868. doi:10.1371/journal.pone.0166868

**Editor:** Zhong-Ke Gao, Tianjin University, CHINA

**Received:** April 9, 2016

**Accepted:** November 5, 2016

**Published:** December 1, 2016



**Data Availability Statement:** The MCYT-100 database is publicly available by request through the consortium homepage. Contact person: Javier Ortega-Garcia, Director, Biometrics Research Lab. (ATVS) Dpto. Ingenieria Informatica, Escuela Politecnica Superior, Universidad Autonoma de Madrid (http://atvs.ii.uam.es).

**Funding:** The authors are grateful to CONICET (Argentina, OAR), CNPq (447229/2014-1, Brazil, ACF) and FACEPE (Brazil, RO) for partial funding of this research. No individuals employed or contracted by the funders (other than the named authors) played any role in: study design, data collection and analysis, decision to publish, or

## Abstract

We present a new approach for handwritten signature classification and verification based on descriptors stemming from time causal information theory. The proposal uses the Shannon entropy, the statistical complexity, and the Fisher information evaluated over the Bandt and Pompe symbolization of the horizontal and vertical coordinates of signatures. These six features are easy and fast to compute, and they are the input to an One-Class Support Vector Machine classifier. The results are better than state-of-the-art online techniques that employ higher-dimensional feature spaces which often require specialized software and hardware. We assess the consistency of our proposal with respect to the size of the training sample, and we also use it to classify the signatures into meaningful groups.

## Introduction

The word *biometrics* is associated to human traits or behaviors which can be measured and used for individual recognition. In fact, the biometry recognition, as a personal authentication signal processing, can be used in situations or instances where users need to be security identified [1]. These kind of systems can either verify or identify.

Two types of biometrics can be defined according to the personal traits considered: *a) physical/physiological* which take into account the biological traits of users, like fingerprints, iris, face, hand, etc. *b) behavioral*, those which consider dynamic traits such as, voice, handwritten evidence and particular expressions. Biometric systems are attractive because of the enhanced security [1] provided by two main facts: (i) users do not have to remember passwords or carry access keys, (ii) it is difficult to steal, imitate or generate genuine biometric data.

The way we sign has the widest social and legal acceptance among pure behavioral biometric traits [2–6]. People sign every day to verify their identity, as this does not require any invasive measurement. Allegedly, this identification and identity verification modality is the most attacked.

preparation of the manuscript. There was no additional external funding received for this study.

**Competing Interests:** The authors have declared that no competing interests exist.

Signatures are written by moving a pen over a surface, e.g., paper or a digitizing device. Handwritten signature verification is a problem in which the input signature (a test signature) is classified as genuine or forged. Although signatures are intended to serve as identity verification, the same person's signature varies due to a number of factors and conditions.

Hilton [7] found that signatures have three main attributes: form, movement, and variation; movement being the most important. The author found that little variations occur over time once a signature style has been adopted. The signing processes can be described at high level as how the the brain recovers information from long term memory in which parameters such as size, shape, timing, etc., are specified, without any particular attention to detail. Genuine signatures are associated to a spurt of neural activity, whereas the forgery signatures are the result of deliberate handwriting which is characterized by a conscious attempt to reproduce [8, 9].

Two opposite mechanisms describing the signing process can be found in the literature. Longstaff and Heath [10] found evidence of chaotic behavior on the underlying dynamics of time series related to velocity profiles of handwritten texts. In opposition, most of the research in the field of signal verification considers the input information as well described by a random process, e.g. Hidden Markov Models [2–6]. Then, the dynamic input information acquired through a time sampling procedure must be considered as a discrete time random sequence.

Offline signature verification is based solely on the signature image, while online procedures require additional information. Our procedure exploits only the temporal information present in the signature coordinates and, thus, can be termed *quasi-offline*.

Following [2–6], we describe the three main stages of our work:

- **Data acquisition and pre-processing.** We perform *quasi-offline* recognition, as we only employ information about coordinates and do not require pressure, speed or pen-up movements data.
- **Feature extraction.** We tackle the problem with *parameter features*: signatures are characterized as a six-dimensional vector extracted from the original data.
- **Classification.** Our approach is related to *distance-based classifiers*, as we will make decisions based on the similarity of the features extracted from the test signature to a description of an ensemble of genuine signatures.

Our proposal relies on the use of time causal quantifiers based on information theory for the characterization of quasi-offline handwritten signatures: *normalized permutation Shannon entropy*, *permutation statistical complexity* and *permutation Fisher information measure*. These quantifiers have proved to be useful in the identification of chaotic and stochastic dynamics throughout the associated time series [11, 12]. Details and further references are provided in the Supplementary Information S1 File. Their evaluation is simple and fast, making them apt for the signature verification problem. We apply our proposal to the well-know MCYT online signature data base [13], but we only use time causal information about their trajectories.

We refer to "time causal information" to attest that the only causal information we use comes from the time ordering of the data. Mutual Information, Conditional Entropy, Transfer Entropy and other similar measures are excellent for identifying and quantifying relationships between processes, e.g. synchronization, causality, etc. [14]. This is not the case in our study, as we do not employ any other process apart from the observed coordinates along time. Those information theory measures would be of great value if we had data about, for instance, the neural activity that leads to the signatures, but we do not.

Our proposal consists, thus, in using features extracted from a nonparametric transformation of two time series. Other recent techniques have been proposed for the analysis of time series as, for instance, transforming them into complex networks [15–19], and using multiscale

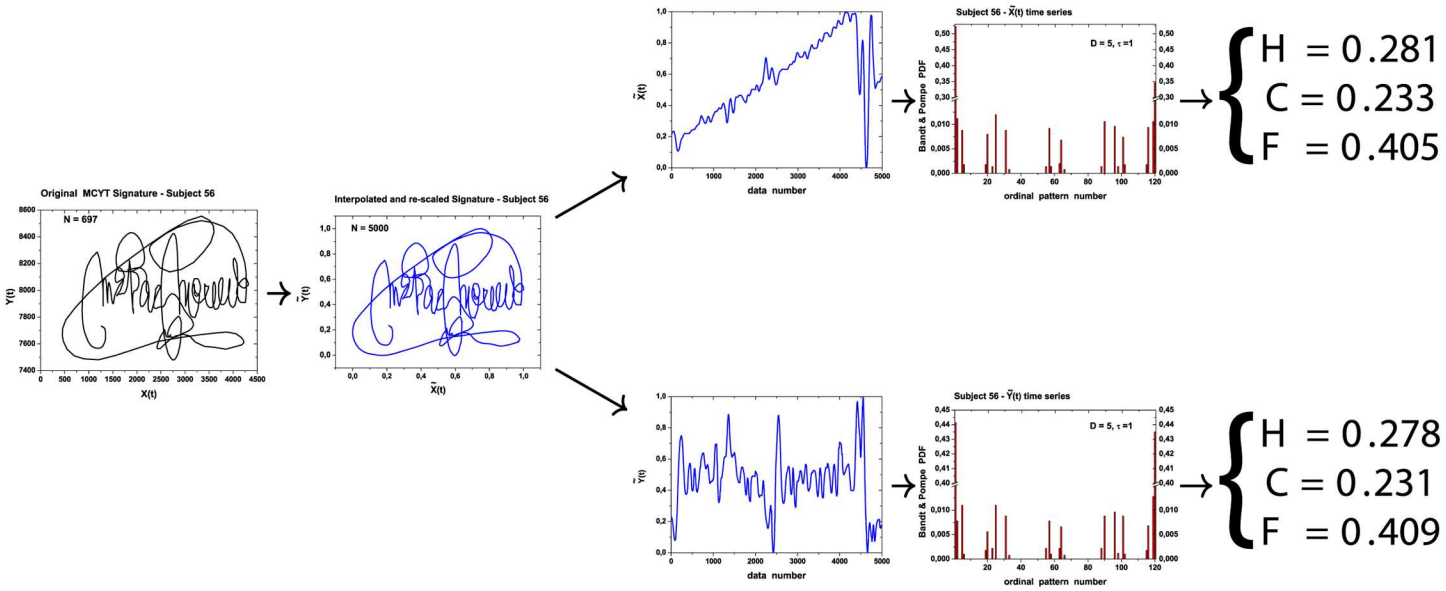


**Fig 1. Diagram of the proposed procedure: original signature, interpolation, X and Y coordinates as time series, Band & Pompe histograms, entropy, statistical complexity and Fisher information.**

doi:10.1371/journal.pone.0166868.g001

analysis [20]. These, and other similar approaches, produce excellent results at the price of heavy computational overload.

Fig 1 sketches the complete workflow of our proposal. Signatures are the input; they are first scaled to fit an unitary square, and interpolated in order to have same number of data for all subjects. Then, the time series of both horizontal and vertical writing processes are extracted. These time series are then represented in a nonparametric manner using a time causal descriptor: the Bandt and Pompe symbolization [21]. A histogram of these symbols is then built for each coordinate, and information theory quantifiers are computed from these histograms: normalized Shannon entropy, Fisher's information measure, and statistical complexity. After an exploratory data analysis, we show that simple dendrograms based on these quantifiers reveal meaningful groups of signatures. The signature stability of each of these groups is also evaluated. Finally, we propose using a One-Class Support Vector Machine for signature verification, and we show that this approach has better performance than state-of-the-art classifiers defined in feature spaces ten times larger than ours. With this, our proposal attains better results in less computational time for an application that, besides being relevant, requires fast responses.

Next section describes the database used in this study. In addition to the usual data flow, we present an exploratory data analysis (EDA) of the features that enhances their appropriateness for this problem. The expressiveness and usefulness of these descriptors for the problem of signature classification and verification follows in the sequence: we experiment their application to the test-bed.

## Handwritten signatures database

The present study is carried out on the freely available and widely used handwritten signatures database MCYT. In the following paragraph, we reproduce the main protocol and

methodological details of the MCYT data base acquisition published by Ortega-Garcia and coworkers in [13, 22, 23].

> "The acquisition of each on-line signature is accomplished dynamically using a graphics tablet. The signatures are acquired on a WACOM© graphic tablet, model INTUOS A6 USB. The tablet resolution is 2540 lines/in (100 lines/mm), and the precision is ±0.25 mm. The maximum detection height is 10 mm (so also pen-up movements are considered), and the capture area is 127 mm (width) × 97 mm (height). This tablet provides the following discrete-time sequences: *a)* position $x_t$ in the $x$-axis, *b)* position $y_t$ in the $y$-axis, and *c)* also the time series corresponding to the pressure $p_t$ applied by the pen, as well as the azimuth $\gamma_t$ and altitude $\varphi_t$ angles of the pen with respect to the tablet, not used in the present work. The sampling frequency is set to 100 Hz. Taking into account the Nyquist sampling criterion and that the maximum frequencies of the related biomechanical sequences are always under 20-30 Hz [24], this sampling frequency leads to a precise discrete-time signature representation. The signature corpus comprises genuine and shape-based highly skilled forgeries with natural dynamics [13, 23]. The forgeries are produced requesting each contributor to imitate other signers by writing naturally. For this task, they were given the printed signature to imitate and were asked not only to imitate the shape, but also to generate the imitation without artifacts such as breaks or slow-downs. Each signer contributes with 25 genuine signatures in five groups of five signatures each, and is forged 25 times by five different imitators. Since signers are concentrated in a different writing task between genuine signature sets, the variability between client signatures from different acquisition sets is expected to be higher than the variability of signatures within the same set. The total number of contributors in the MCYT is 330, and the total number of signatures present in the signature database is 16,500, half of them genuine signatures and the rest forgeries."

We used the MCYT-100 subset of the database, which includes 100 subjects and for each one, 25 genuine and 25 skilled forged signatures. The only data we use are the $x$- and $y$-coordinates time series.

Fig 2 presents examples of six subjects, being the first two columns genuine and the third column forgery signatures. In particular, one must note that the time series' lengths are quite variable. We pre-processed each time series as follows: *a)* the coordinates were re-scaled into the unit square $[0, 1] \times [0, 1]$; *b)* the original total number of data for each time series is expanded to $M = 5000$ points using a cubic Hermite polynomial. In this way, for each subject $k$ ($k = 1, \ldots, 100$) and associated signatures $j$ ($j = 1, \ldots, 25$) we will analyze two time series, denoted by $\mathbf{X}_j^{(k;\alpha)} = \{0 \leq \tilde{x}_{j;i}^{(k;\alpha)} \leq 1, i = 1, \ldots, M\}$ and $\mathbf{Y}_j^{(k;\alpha)} = \{0 \leq \tilde{y}_{j;i}^{(k;\alpha)} \leq 1, i = 1, \ldots, M\}$, in which the supra-index $\alpha = G, F$ denotes genuine and forgery signature, and $\tilde{x}$ and $\tilde{y}$ are the interpolated values, respectively.

## Signature features and exploratory data analysis

Handwritten classification and verification is an important and challenging problem due to two main factors. First, intra-personal variation in speed, pressure and inclination can be large, as signature consistency is often poor. Second, we can only obtain few samples from one person and no forgeries in practice. The reliability of extracted features is, thus, difficult to assess.

Developing an efficient and effective system for data acquisition is a challenging task. The volume of their databases grows boundlessly and soon becomes unmanageable, so reducing the raw data to parsimonious forms, without loosing important information, is at the core of

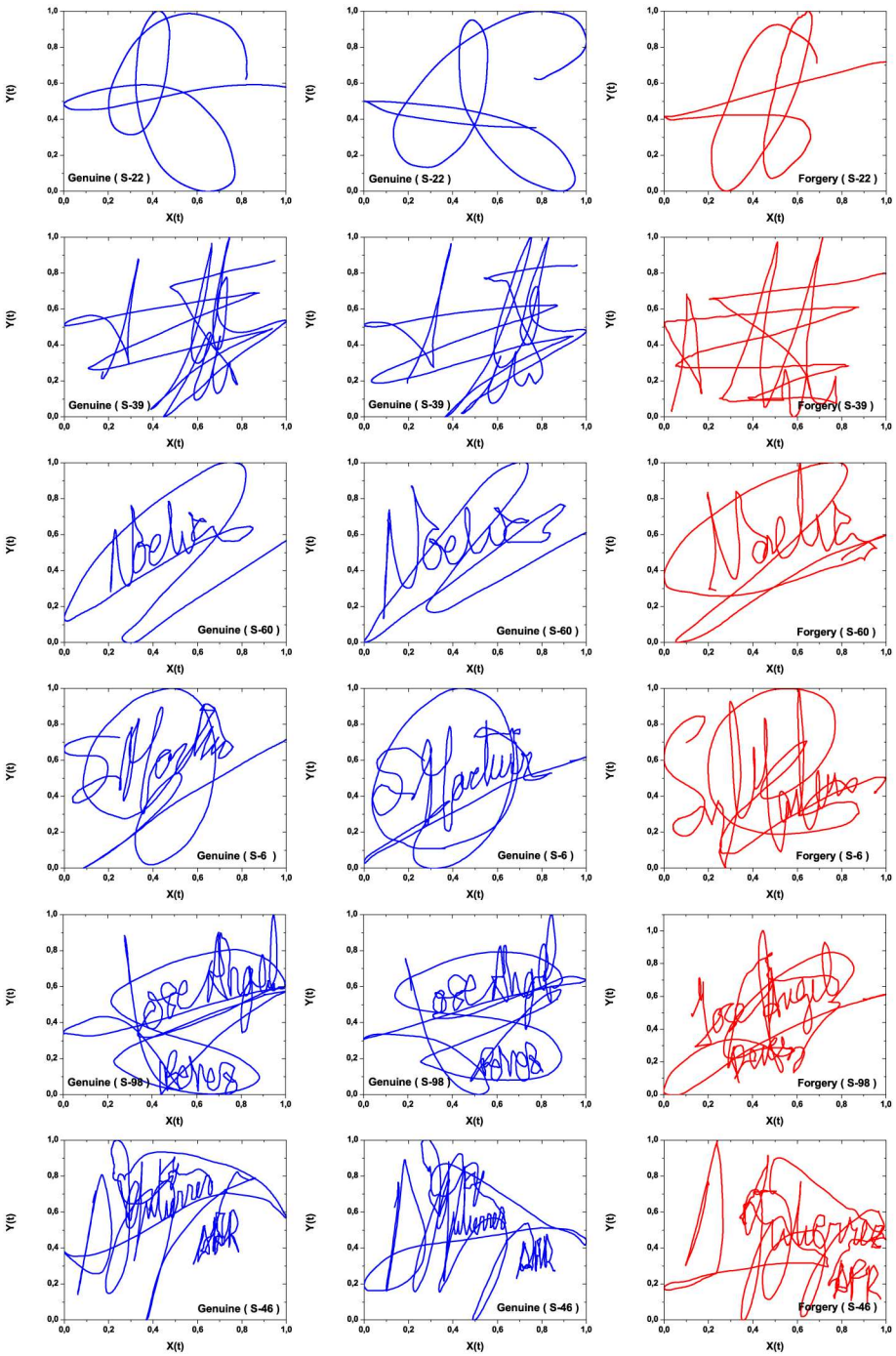


**Fig 2. Six different subjects signatures from the MCYT database.** Two genuine signatures (left, blue) and a skilled forgery (right, red). The two first signatures were classified as H1A and H1B, the following two to types H2A and H2B, and the last two to types H3A and H3B; cf. Sec. Signature classification.

doi:10.1371/journal.pone.0166868.g002

intelligent solutions. We aim at discovering relevant low-dimensional features that, albeit promoting the reduction of data, are able to differentiate forgery from authentic signatures.

In this work we employ time causal information theory quantifiers; see details in the Supplementary Information S1 File. For each of the $k$ subjects ($k = 1, \ldots, 100$) in the database and its $j$ associated signatures (25 genuine and 25 skilled forgery), two time series $\mathbf{X}_j^{(k;\alpha)}$ and $\mathbf{Y}_j^{(k;\alpha)}$ are extracted and transformed into Bandt and Pompe's PDFs with pattern length (embedding dimension) $D = 5$ and time lag $\tau = 1$ [21].

We denoted these PDFs as:

$$\begin{aligned} P_{X;j}^{(k;\alpha)} &= \text{Bandt and Pompe's PDF of } \mathbf{X}_j^{(k;\alpha)}|_{D,\tau}, \text{ and} \\ P_{Y;j}^{(k;\alpha)} &= \text{Bandt and Pompe's PDF of } \mathbf{Y}_j^{(k;\alpha)}|_{D,\tau}, \end{aligned}$$

in which $j = 1, \ldots, 25$, and $\alpha = G, F$ identify genuine and skilled forgery signatures, respectively.

We chose $D = 5$ after trying other values: $D = 3, 4$ led to too coarse histograms (not enough bins), while $D = 6$ (that requires counting 720 cases) produced too many zero-count bins. Note that the condition $M \gg D!$ is satisfied with $D = 5$. We used unlagged data ($\tau = 1$) after checking that there were not significant changes with lagged $\tau = 2, 3$ series.

We computed the normalized permutation Shannon entropy $\mathcal{H}$, the permutation statistical complexity $\mathcal{C}$, and the permutation Fisher information measure $\mathcal{F}$ from these PDFs, and the obtained values are denoted as:

$$\begin{aligned} \mathcal{H}_{X;j}^{(k;\alpha)} &= \mathcal{H}[P_{X;j}^{(k;\alpha)}], \quad \mathcal{H}_{Y;j}^{(k;\alpha)} = \mathcal{H}[P_{Y;j}^{(k;\alpha)}]; \\ \mathcal{C}_{X;j}^{(k;\alpha)} &= \mathcal{C}[P_{X;j}^{(k;\alpha)}], \quad \mathcal{C}_{Y;j}^{(k;\alpha)} = \mathcal{C}[P_{Y;j}^{(k;\alpha)}]; \\ \mathcal{F}_{X;j}^{(k;\alpha)} &= \mathcal{F}[P_{X;j}^{(k;\alpha)}], \quad \mathcal{F}_{Y;j}^{(k;\alpha)} = \mathcal{F}[P_{Y;j}^{(k;\alpha)}]. \end{aligned}$$

We performed Exploratory Data Analysis (EDA) on these information theory quantifiers looking for simple descriptions of the data. We also used the Pearson correlation to measure the association between features. This analysis was performed using the R language and platform version 3.2.1 (http://www.R-project.org).

Fig 3 shows a scatterplot of the entropy for both the genuine and skilled forgery signatures. The 5000 points correspond to 25 genuine signatures (in blue) and 25 forgery signatures (in red) for each of the 100 subjects. Both types of signatures show similar association (Correlation): $\text{Corr}(\mathcal{H}_{X;j}^{(k;G)}, \mathcal{H}_{Y;j}^{(k;G)}) = 0.9665$ and $\text{Corr}(\mathcal{H}_{X;j}^{(k;F)}, \mathcal{H}_{Y;j}^{(k;F)}) = 0.9770$. The entropies of both types of signatures are overlapped and scattered elliptically. However, the bivariate mean and dispersion values differ.

Entropies are less dispersed in the genuine than in the skilled forgery signatures, a signal of the separability between them. Marginal density plots show the distribution of entropy for each coordinate of both types of signatures. These plots, in spite of being limited due to its marginal nature, reveal several modes, and suggest different dispersion patterns.

Fig 4 shows the contour plots of bivariate kernel density estimates for the entropy in genuine and forgery signatures. A number of features are immediately noticeable. The dispersion in the former group is much smaller than in the latter (less than 0.4). The kernel density estimates reveal skewness and a mild multimodality in the joint distribution of the data. Quite many points that are far from these curves and cluster centers. These points correspond to abnormal local estimates obtained in heterogeneous blocks, possibly induced by the presence of clusters. The modes in genuine signatures are smaller than in forgery signatures, and this may be used as discriminatory measure. Similar results are obtained for the Complexity and

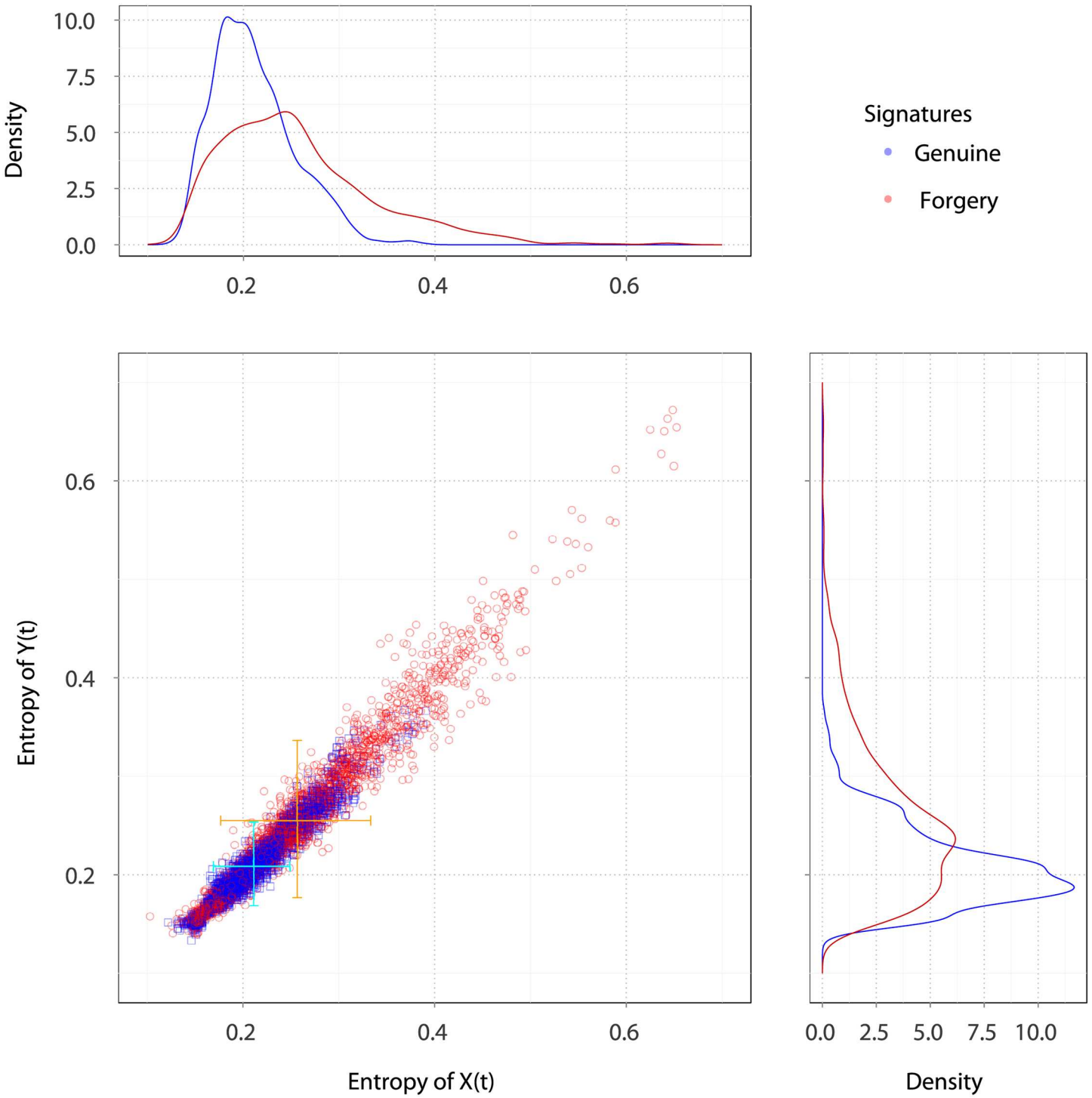


**Fig 3. Scatter plot with marginal kernel density estimates of entropy quantifiers in both trajectory coordinates time series X and Y.** Genuine (blue) and skilled forgery signatures (red points), 100 subjects. Marginal kernel densities depict the distribution of entropy quantifiers along both axes.

doi:10.1371/journal.pone.0166868.g003

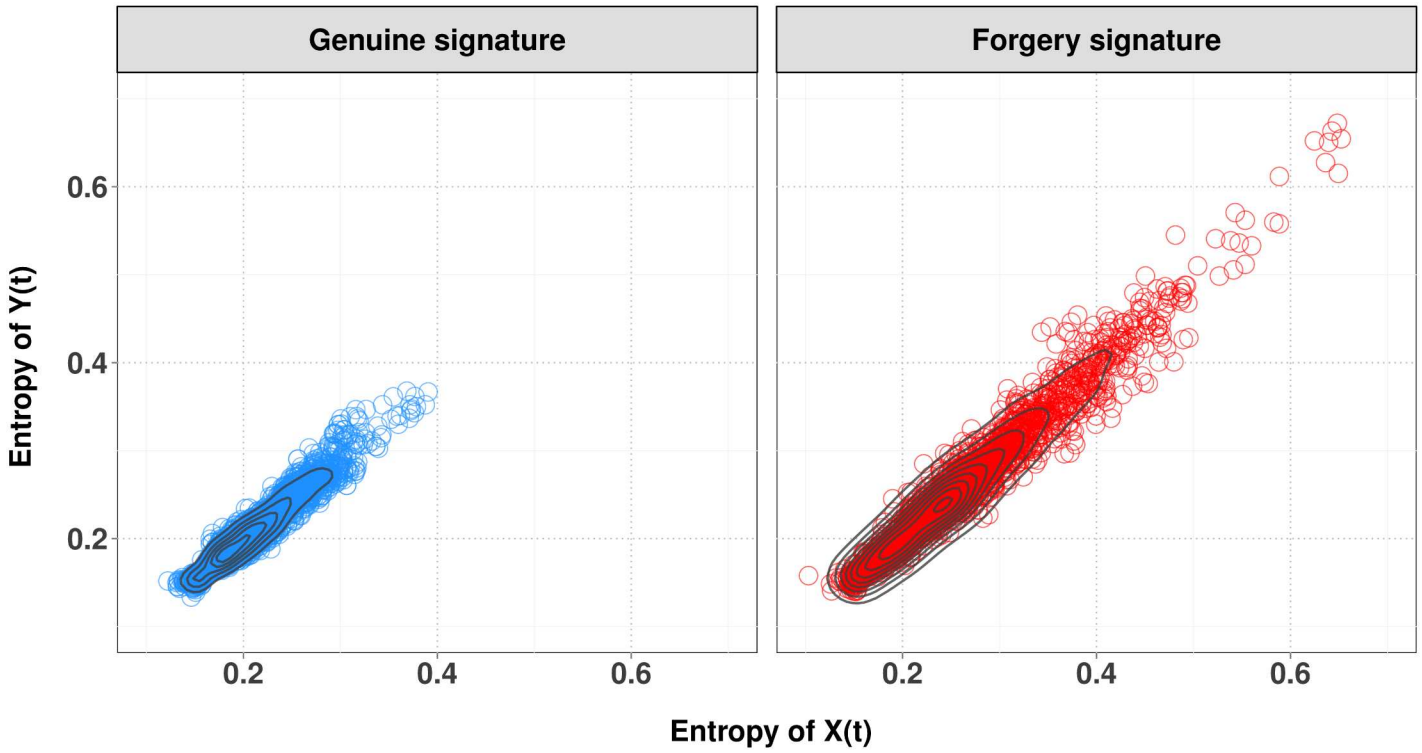


**Fig 4. Contour plot superimposed on the scatterplot of entropy quantifiers for genuine (right panel) and skilled forgery signatures (left panel).**

doi:10.1371/journal.pone.0166868.g004

the Fisher information; these are reported in the Supplementary Information, see Figs A, B, C and D in S3 File, respectively.

## Signature classification

As pointed out by Boulétreal *et al.* [25], a signature is characterized by two aspects: *a)* a conscious one associated to the pattern signature; and *b)* an unconscious one which leads spontaneous movements constituting the drawing. These two factors produce high variability, being the amount of signature variability strongly writer-dependent. In fact, the signature *variability* or, conversely, the signature *stability* can be considered an important indicator for writer characterization [26]. Houmani and Garcia-Salicetti [26] argue that signature stability is required in genuine signatures to characterize a writer: signature variability reduces the ability to identify forgery. Also, complex enough signatures are required to guarantee a certain level of security, in the sense that the more complex a signature is, the more difficult it will be to forge it [26].

Boulétreal and collaborators [25, 27] propose a signature complexity measure related to signature legibility and based on fractal dimension. They classify writer styles into: highly cursive, very legible, separated, badly formed, and small writings, using only genuine signatures. Unfortunately, such resulting categories were not confronted to classifiers for performance analysis.

We classified the one hundred genuine signatures in the MCYT-100 data base with causal information theory quantifiers: Normalized permutation Shannon entropy, permutation statistical complexity and permutation Fisher information measure of both **X** and **Y** trajectories. The mean and standard deviation values were clustered using the neighbor-joining method

and an automatic Hierarchical Clustering with the Euclidean distance-based dissimilarity matrix. Each feature was treated independently, and the results are shown as circular dendrograms. Fig 5 shows the results of clustering the entropy. With this, we distinguish three classes of genuine signatures denoted by H1, H2, and H3.

The H1 group is the first group to form, i.e., the one comprised of the most similar individuals. It is formed below the 25% level, and it is composed by two subgroups: H1A and H1B. The H1A group is formed exclusively by oversimplified signatures made by mere loops without identifiable letters. It encompasses the following subjects: 1, 16, 17, 22, 23, 27, 29, 37, 83. The same group is formed when the other features are used. The H1B group is comprised of the following subjects: 2, 5, 8, 10, 19, 21, 24, 28, 32, 35, 36, 39, 43, 48, 49, 51, 55, 58, 59, 64, 69, 70, 74, 77, 89. Although these are simplified signatures, traces of letters and/or more complex curves appear and differentiate them from the members of H1A.

The H2 group is formed approximately at the 32% level, and, again, it is comprised of two distinct groups: H2A and H2B. The subjects that make the H2A group are: 4, 7, 12, 15, 18, 20, 30, 31, 34, 38, 40, 41, 42, 52, 57, 60, 62, 66, 67, 68, 71, 73, 75, 79, 80, 81, 86, 87, 91, 96, 100. It is composed by signatures with traces that resemble letters, but that are not perfectly identifiable, and that include circling traces of large or moderate size. Signatures in this group are kind of framed by large loops. The H2B group is similar to the previous one, i.e., it is formed by signatures with large and medium size circling traces, but with more identifiable letters than in the previous groups. Names and surnames are more readable in this group than in previous ones. It is formed by the following signatures: 6, 9, 13, 25, 33, 45, 50, 63, 65, 76, 78, 82, 84, 85, 88, 92, 94, 95, 97, 99.

The H3 group is formed at, approximately, the 43% level by the fusion of two other highly unbalanced subgroups: one, H3A, with only two subjects (44, 46) and the other, H3B, with thirteen subjects (3, 11, 14, 26, 47, 53, 54, 56, 61, 72, 90, 93, 98). These two clusters form at approximately the same level. The former is composed of calligraphic signatures where vertical traces predominate over horizontal ones. The latter is composed of highly cursive signatures, with separation between the surname and the family name.

The same results of clustering was obtained with the Manhattan (norm $\mathcal{L}_1$) and Maximum distances ($\mathcal{L}_\infty$ norm), showing that entropy is an expressive and stable quantifier. Similar analyses were carried with the permutation statistical complexity and permutation Fisher information (presented in Supplementary Information Figs A and B in S4 File). Complexity produces the same clusters identified by entropy, so it adds no new information. The Fisher information measure forms the same H1A group that was identified by the entropy, but with less cohesion, at about 15%. In other words, these nine subjects are more similar locally than globally. As with entropy, three main groups form at similar levels. The members of these clusters are slight variations of those identified using entropy, with very similar structure.

Table 1 presents the mean and standard deviation of the three quantifiers over the 25 genuine and 25 skilled forgery signatures (**X** and **Y** time series) for each of the typical subjects, split in types H1, H2, and H3. These data reveal interesting tendencies. Genuine signatures present quantifiers values lower than those corresponding to forgery signatures, and the latter also exhibit larger standard deviation. This may be explained by the imitative character of these signatures, however it deserves closer studies.

The classification into subclasses of genuine signatures was also carried by the parallelepiped algorithm [28], arguably the simplest model-free classification procedure. Entropy leads to clusters with nice interpretability. Fig 6 shows the regions that define the three classes identified by the dendrogram based on entropy presented in Fig 5. All subclasses are well separated by disjoint boxes, except H1B and H2A that overlap slightly but without compromising the

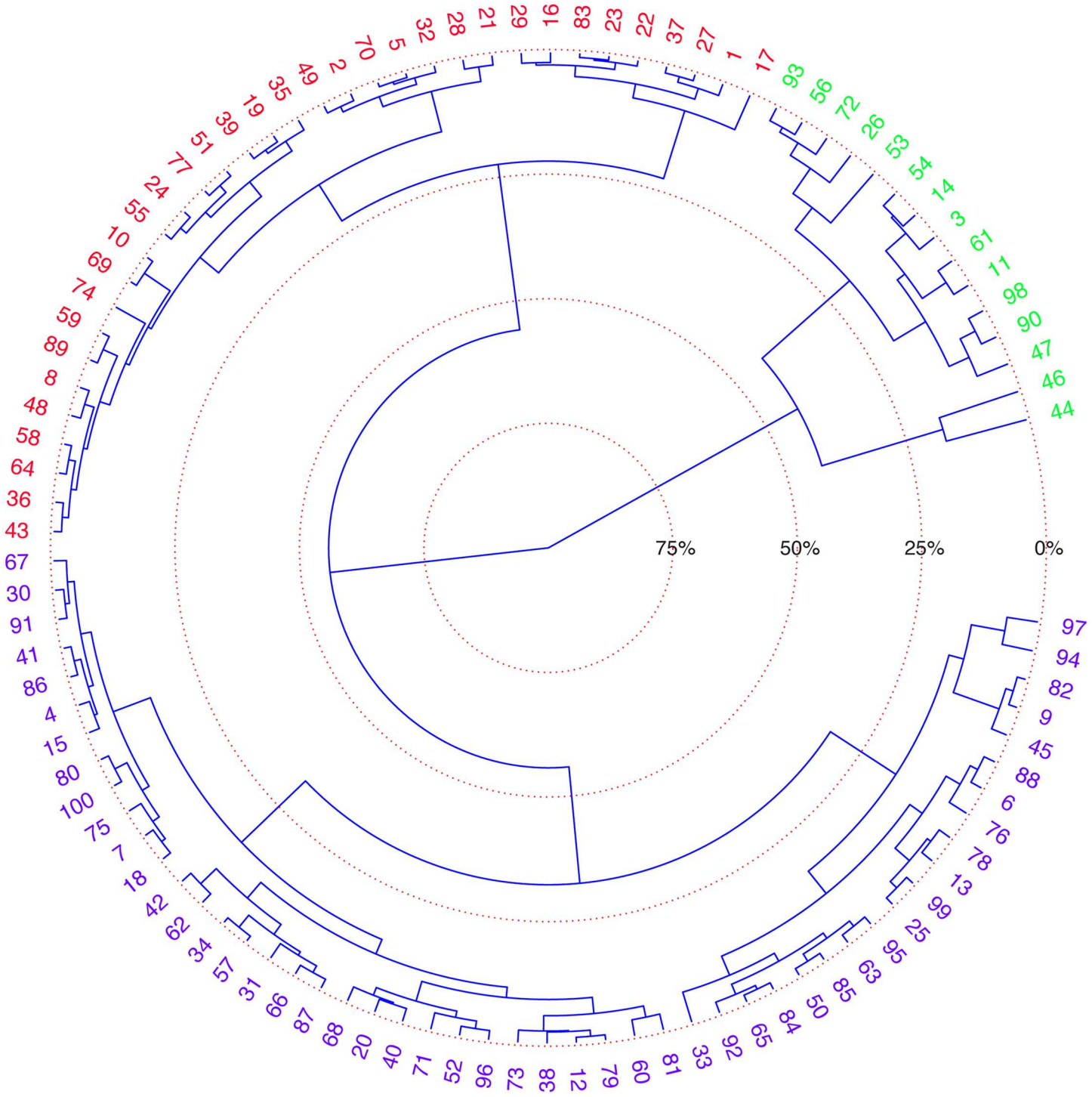


**Fig 5. Neighbor-joining, rooted, circular dendrogram clustering of genuine signatures by entropy: H1, H2, and H3, in red, blue, and green, respectively.**

doi:10.1371/journal.pone.0166868.g005

**Table 1. Sample mean and standard deviation (S.D.) of the time series quantifiers for the 25 genuine (G) and 25 skilled forged (F) signatures, for each of the typical subjects: H1A, H1B, H2A, H2B, H3A, and H3B (same order as in Fig 2).**

| | | | | | Entropy | | Complexity | | Fisher Information | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Type* | *Sub–Type* | Subject | Coordinate | Class | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| H1 | H1A | 22 | **X** | F | 0.1568 | 0.0052 | 0.1490 | 0.0039 | 0.4688 | 0.0070 |
| | | | | G | 0.1519 | 0.0019 | 0.1457 | 0.0015 | 0.4766 | 0.0035 |
| | | | **Y** | F | 0.1595 | 0.0071 | 0.1511 | 0.0052 | 0.4665 | 0.0097 |
| | | | | G | 0.1512 | 0.0042 | 0.1447 | 0.0037 | 0.4734 | 0.0046 |
| | H1B | 39 | **X** | F | 0.2212 | 0.0384 | 0.1941 | 0.0257 | 0.4286 | 0.0147 |
| | | | | G | 0.1749 | 0.0037 | 0.1620 | 0.0028 | 0.4497 | 0.0029 |
| | | | **Y** | F | 0.2270 | 0.0449 | 0.1980 | 0.0296 | 0.4277 | 0.0153 |
| | | | | G | 0.1776 | 0.0043 | 0.1644 | 0.0031 | 0.4491 | 0.0035 |
| H2 | H2A | 60 | **X** | F | 0.2482 | 0.0593 | 0.2112 | 0.0365 | 0.4212 | 0.0107 |
| | | | | G | 0.2010 | 0.0056 | 0.1803 | 0.0040 | 0.4331 | 0.0031 |
| | | | **Y** | F | 0.2442 | 0.0544 | 0.2090 | 0.0339 | 0.4219 | 0.0134 |
| | | | | G | 0.2079 | 0.0043 | 0.1861 | 0.0030 | 0.4315 | 0.0024 |
| | H2B | 6 | **X** | F | 0.2621 | 0.0584 | 0.2194 | 0.0334 | 0.4143 | 0.0137 |
| | | | | G | 0.2337 | 0.0149 | 0.2032 | 0.0095 | 0.4205 | 0.0066 |
| | | | **Y** | F | 0.2648 | 0.0538 | 0.2218 | 0.0304 | 0.4136 | 0.0134 |
| | | | | G | 0.2314 | 0.0102 | 0.2018 | 0.0067 | 0.4211 | 0.0050 |
| H3 | H3A | 98 | **X** | F | 0.3236 | 0.0646 | 0.2529 | 0.0320 | 0.3937 | 0.0208 |
| | | | | G | 0.2707 | 0.0101 | 0.2268 | 0.0064 | 0.4106 | 0.0032 |
| | | | **Y** | F | 0.3204 | 0.0794 | 0.2497 | 0.0388 | 0.3970 | 0.0208 |
| | | | | G | 0.2664 | 0.0124 | 0.2243 | 0.0077 | 0.4105 | 0.0034 |
| | H3B | 46 | **X** | F | 0.3514 | 0.0641 | 0.2691 | 0.0294 | 0.3940 | 0.0156 |
| | | | | G | 0.3480 | 0.0282 | 0.2720 | 0.0156 | 0.4019 | 0.0047 |
| | | | **Y** | F | 0.3419 | 0.0681 | 0.2639 | 0.0323 | 0.3940 | 0.0163 |
| | | | | G | 0.3270 | 0.0263 | 0.2599 | 0.0148 | 0.4008 | 0.0052 |

doi:10.1371/journal.pone.0166868.t001

discrimination. The classes are preserved using this classification superimposed with Complexity and Fisher information features; see Figs C and D in S4 File.

## Signature stability measure

We now assess the stability of the features the classification procedure will use as input. Two measures of instability are computed over the PDFs obtained for each time series: one global (the Jensen-Shannon divergence [29–31]) and another local (the Jensen-Fisher divergence [32, 33]).

We propose using, for each subject, the square root of the Jensen-Shannon divergence over his/her 25 genuine signatures (denoted by $\eta^{(k)}$) as a global index of instability

$$\eta_X^{(k)} = \sqrt{\mathrm{S}\left[\frac{1}{25}\sum_{j=1}^{25} P_{X;j}^{(k;G)}\right] - \frac{1}{25}\sum_{j=1}^{25}\mathrm{S}\left[P_{X;j}^{(k;G)}\right]}, \quad (1)$$

$$\eta_Y^{(k)} = \sqrt{\mathrm{S}\left[\frac{1}{25}\sum_{j=1}^{25} P_{Y;j}^{(k;G)}\right] - \frac{1}{25}\sum_{j=1}^{25}\mathrm{S}\left[P_{Y;j}^{(k;G)}\right]}, \quad (2)$$

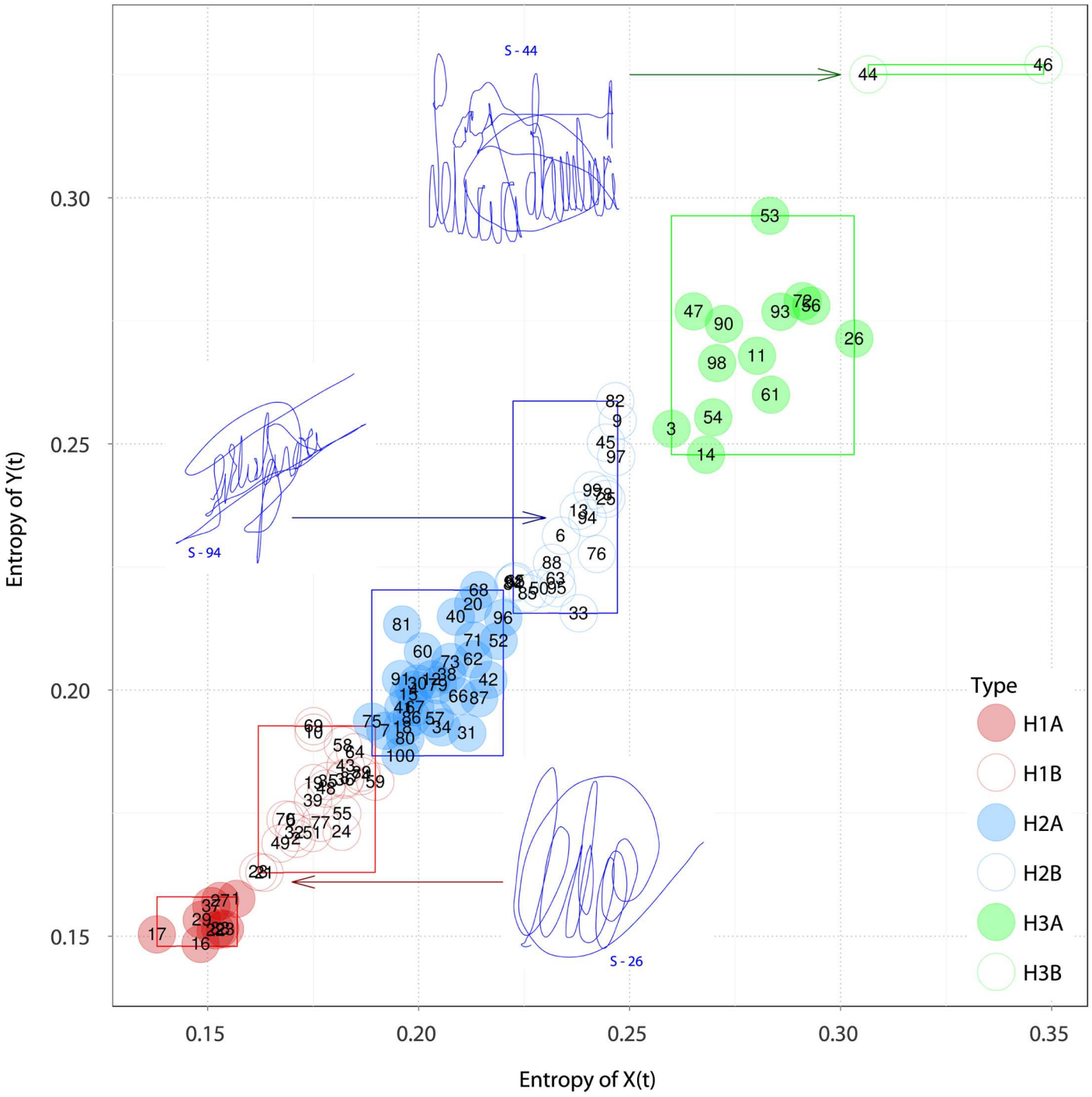


**Fig 6. Classification by the rule of the parallelepiped of genuine signatures using entropy (one signature example from each of the three groups is shown).** Each subject is identified by its ID.

doi:10.1371/journal.pone.0166868.g006

in which, S[•] represents the Shannon entropy, $P_{X;j}^{(k;G)}$ and $P_{Y;j}^{(k;G)}$ are the Bandt-Pompe's PDF associated to time series of coordinates $\tilde{x}$ and $\tilde{y}$ of the $j$ genuine signature ($\alpha = G, j = 1, \ldots, 25$) of subject $k$ ($k = 1, \ldots, 100$).

Analogously, we define a local instability index using the Fisher information measure, $\mathcal{F}[\bullet]$, and evaluating the Jensen-Fisher divergence. We then have

$$\xi_X^{(k)} = \sqrt{\frac{1}{25}\sum_{j=1}^{25}\mathcal{F}\left[P_{X;j}^{(k;G)}\right] - \mathcal{F}\left[\frac{1}{25}\sum_{j=1}^{25}P_{X;j}^{(k;G)}\right]}, \tag{3}$$

$$\xi_Y^{(k)} = \sqrt{\frac{1}{25}\sum_{j=1}^{25}\mathcal{F}\left[P_{Y;j}^{(k;G)}\right] - \mathcal{F}\left[\frac{1}{25}\sum_{j=1}^{25}P_{Y;j}^{(k;G)}\right]}. \tag{4}$$

Fig 7 shows the plots of mean with standard error bars of instability index calculated by each type of genuine signatures by subclasses as obtained from preclasification. The first observation is that the Jensen-Fisher local measure of instability (bottom) is the same in the horizontal (left) and right (left) time series, whereas it changes when measured by the Jensen-Shannon global measure (top).

The global measure of instability indicates that the most unstable group of genuine signatures is H3B, but only two samples are available in this class. Both H2 classes exhibit similar instabilities in both horizontal and vertical time series **X** and **Y**. The **X** and **Y** time series show a symmetrical behavior in class H1: **X** is more stable than **Y** in H1A, whereas **Y** is more stable than **X** in H1B. The least variable instability is observed in the H2 class.

All mean local stabilities, except that of H3B, are similar in the horizontal and vertical directions. The subclass H3B is, again, the most unstable, but it is more stable in the vertical direction.

Overall, the measured instability is small in all subclasses granting, thus, stable classification results based on these features.

## Quasi-offline signature verification

The problem we have at hand consists of identifying suspicious signatures given that we only have examples from genuine signatures. In practice, it is too expensive, too hard or even impossible to obtain a significant number of good quality forgery signatures for every possible individual in the data base. This, thus, configures a One-Class classification problem.

Support Vector Machines (SVMs) are suitable for solving machine learning problems even in large dimensional feature spaces [34–36]. We provide a brief description of SVMs and One-Class SVMs in the Supplementary Information S2 File along with a toy example with simulated data. We used the `libsvm` (version 2.0) tool, linked with the `R` software that implements SVM classification and regression, and One-Class SVMs (OC-SVM) [37] tools, with the default parameters.

We assess the consistency of our procedure in a reproducible manner by evaluating the performance of the proposed verification system for different training samples. Were selected random samples of size $n = 5, 10, 14, 18, 22$ of genuine signatures for each user. Table 2 presents the average value of all performance metrics using $\sigma^2 = 10$ (see Supplementary Information S2 File). The observed Accuracy (ACC) suggests that the larger the training sample is the better the performance is. The Area Under the ROC Curve (AUC) presents a similar tendency, and its average is larger than 0.88, indicating that our verification system produces excellent classification.

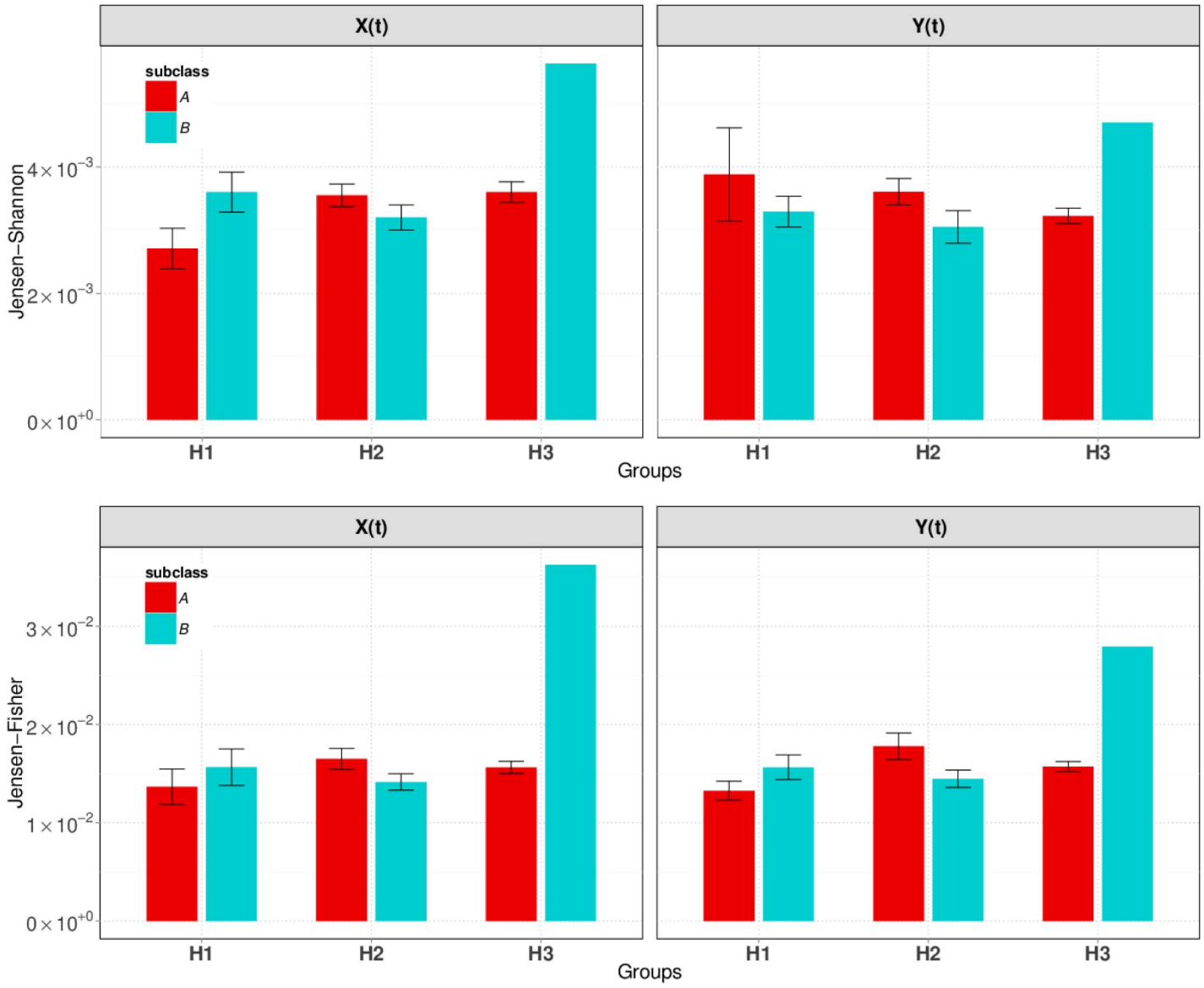


**Fig 7. Global Jensen-Shannon (top) and local Jensen-Fisher (bottom) measures of instability in genuine signatures.** Bars show the mean, and lines show the standard error over the subjects. The standard error of H3B is not plotted because there are only two subjects in this class.

doi:10.1371/journal.pone.0166868.g007

**Table 2. Performance of the system trained with varying number *n* of samples of genuine signatures; ↑ and ↓ denote measures of quality (the higher the better) and of error (the smaller the better), respectively.**

| *n* | ACC (↑) | AUC (↑) | EER(%) (↓) |
|---|---|---|---|
| 5 | 0.6940 | 0.8816 | 0.1890 |
| 10 | 0.7678 | 0.8940 | 0.1711 |
| 14 | 0.8144 | 0.8975 | 0.1634 |
| 18 | 0.8250 | 0.8866 | 0.1731 |
| 22 | 0.8389 | 0.8909 | 0.1632 |

doi:10.1371/journal.pone.0166868.t002

**Table 3. Performance of the classification of pre-classified samples varying the number *n* of samples of genuine signatures used for training; same coding as in Table 2.**

| Class | *n* | ACC (↑) | AUC (↑) | EER(%) (↓) |
|---|---|---|---|---|
| H1 | 5 | 0.6758 | 0.8692 | 0.1976 |
| | 10 | 0.7566 | 0.8828 | 0.1812 |
| | 14 | 0.8039 | 0.8857 | 0.1717 |
| | 18 | 0.8217 | 0.8894 | 0.1662 |
| | 22 | 0.8277 | 0.8788 | 0.1631 |
| H2 | 5 | 0.7059 | 0.8945 | 0.1784 |
| | 10 | 0.7819 | 0.9079 | 0.1548 |
| | 14 | 0.8284 | 0.9096 | 0.1509 |
| | 18 | 0.8327 | 0.8900 | 0.1734 |
| | 22 | 0.8515 | 0.8996 | 0.1608 |
| H3 | 5 | 0.6948 | 0.8653 | 0.2053 |
| | 10 | 0.7450 | 0.8720 | 0.2036 |
| | 14 | 0.7907 | 0.8832 | 0.1874 |
| | 18 | 0.8062 | 0.8686 | 0.1874 |
| | 22 | 0.8214 | 0.8889 | 0.1716 |

doi:10.1371/journal.pone.0166868.t003

As mentioned in the introduction, the two methodologies with best results are those based on Dynamic Time Warping (DTW) and Hidden Markov Models (HMM). In the following we compare our proposal with these two recent state-of-the-art methods using the Equal Error Rate, EER(%) over the same data base:

- Fierrez-Aguilar *et al.* [38], ERR(%) = 2.12 (five training signatures; Global (Parzen WC) and local (HMM) experts function);
- Fierrez-Aguilar *et al.* [22], ERR(%) = 0.74 (ten training signatures; HMM based algorithm);
- Pascual-Gaspar *et al.* [39], ERR(%) = 1.23 (five training signatures; DTW-bases algorithm, result with scenario-dependent optimal features.

The results of our proposal using five (ten, respectively) training samples, are ERR(%) = 0.19 (0.17, respectively). Our system, thus, provides better performance using similar number of training signatures (see Table 2 for more details).

In the following we analyze the performance of the proposed procedure applied selectively to the pre-classified samples. Table 3 presents the performance of the system when applied to genuine pre-classified signatures. For all classes, larger training samples lead to larger average ACC. The best average AUC are observed for the class H2, followed by H1 and H3. This indicates that H2 signatures are easily identifiable. Note that the mean values of ERR(%) for H2 are smaller than H1 and H3. The ERR(%) values in H3 indicate that identifying forgeries in this class is hard.

## Conclusions

We proposed a quasi-offline procedure for identifying skilled forgery of handwritten signatures using time causal information Theory quantifiers and One-Class Support Vector Machines. This is a competitive proposal from the computational viewpoint as it uses only the signatures coordinates, and it produces better results than state-of-the-art techniques. The improvement is obtained in a six-dimensional feature space, while other techniques employ forty or more features. As a consequence, the processing time, memory and storage required

are reduced and, at the same time, the procedure is less prone to the problems induced by the curse of dimensionality. Such improvements make our proposal apt for becoming stand-alone application in, e.g., mobile banking.

The technique also produces meaningful classification of the input data, as it is able to separate different types of signatures. To the best of our knowledge, this is the first time information theory quantifiers have been used for this problem.

The central contribution is the use of the Bandt and Pompe (BP) PDF symbolization which is invariant to a number of transformations of the input data. In fact, the original time series are pre-processed only to facilitate the signal sampling, and this scaling has no effect on the BP PDFs. This representation, which is sensitive to the time causality, is able to capture essential dynamical characteristics of the signatures that lead to excellent discrimination between skilled forgery and genuine handwritten signatures, despite the high variability the data possess. Additionally, obtaining the BP PDFs is computationally simple and efficient.

Only six information theory features are required for the classification, three from each horizontal and vertical direction: Shannon entropy, statistical complexity, and Fisher information. This contrasts many state-of-the-art works that require features in high-dimensional spaces, e.g. forty or even more. As said, our proposal does not require highly specialized hardware able to capture signature speed, pressure, orientation, etc.

The classification was performed by a One-Class Support Vector Machine trained with genuine signatures. The learned rule is consistent with respect to the number of training samples, and with as few as five examples it surpasses the performance of recent successful techniques. We assessed the performance of our proposal using the same data base employed in the current literature, with also the same measures of quality and error.

Future work includes the use of other variables already available in the MCYT data base (pressure, and azimuth and altitude angles), along with other features, e.g. clustering coefficient entropy, network clustering coefficient, permutation min-entropy [40–42], and clustering and classification techniques as, for instance, deep learning [43].

## Supporting Information

**S1 File. Supporting Information file that contains additional material about Information Theory Quantifiers.**
(PDF)

**S2 File. Supporting Information file that contains additional material about Support Vector Machines and One-Class Support Vector Machines.**
(PDF)

**S3 File. Supporting Information file that contains additional material about Exploratory Data Analysis.**
(PDF)

**S4 File. Supporting Information file that contains additional material about Signature Classification.**
(PDF)

## Acknowledgments

The authors are grateful to CONICET (Argentina), CNPq and FACEPE (Brazil) for partial funding of this research. The Biometrics Research Lab (ATVS), Universidad Autónoma de Madrid, provided the MCYT-100 signature corpus employed in this work.

## Author Contributions

**Conceptualization:** OAR ACF.

**Data curation:** OAR RO ACF.

**Formal analysis:** OAR RO ACF.

**Funding acquisition:** OAR RO ACF.

**Investigation:** OAR RO ACF.

**Methodology:** OAR RO ACF.

**Project administration:** OAR.

**Resources:** OAR RO ACF.

**Software:** OAR RO.

**Supervision:** OAR.

**Validation:** OAR RO ACF.

**Visualization:** OAR RO ACF.

**Writing – original draft:** OAR ACF.

**Writing – review & editing:** OAR ACF.

## References


1. Ortega-Garcia J, Bigun J, Reynolds D, Gonzalez-Rodriguez J. Authentication gets personal with biometrics. IEEE Signal Processing Magazine. 2004; 21(2): 50–62. doi: 10.1109/MSP.2004.1276113
2. Plamondon R, Lorette G. Automatic signature verification and writer identification: the state of the art. Pattern Recognition. 1989; 22(2): 107–131. Available from: http://www.sciencedirect.com/science/article/pii/0031320389900599 doi: 10.1016/0031-3203(89)90059-9
3. Leclerc F, Plamondon R. Automatic signature verification: The state of the art: 1989–1993. International Journal of Pattern Recognition and Artificial Intelligence. 1994; 8(3): 643–660. doi: 10.1142/S0218001494000346
4. Gupta G, McCabe A. A review of dynamic handwritten signature verification. Department of Computer Science, James Cook University, Australia; 1997. Available from: http://www.cs.jcu.edu.au/~alan/Work/HSV-Lit_rev.html
5. Impedovo D, Pirlo G. Automatic Signature Verification: The State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 2008; 38(5): 609–635. doi: 10.1109/TSMCC.2008.923866
6. El-Henawy IM, Rashad MZ, Nomir O, Ahmed K. Online signature verification: state of the art. International Journal of Computers and Technology. 2013; 4: 664–678.
7. Hilton O. Signatures, review and a new view. Journal of Forensic Sciences. 1992; 37: 125–129. doi: 10.1520/JFS13219J
8. Denier van der Gon JJ, Thuring JP. The guiding of human writing movement. Kybernetik. 1965; 2: 145–148.
9. Nalwa VS. Automatic on-line signature verification. Proceedings of the IEEE. 1997; 85(2): 215–239. doi: 10.1109/5.554220
10. Longstaff M, Heath R. A nonlinear analysis of temporal characteristic of handwriting. Human Movement Science. 1999; 18: 485–524. doi: 10.1016/S0167-9457(99)00028-7
11. Rosso OA, Larrondo HA, Martín MT, Plastino A, Fuentes MA. Distinguishing noise from chaos. Physical Review Letters. 2007; 99: 154102. doi: 10.1103/PhysRevLett.99.154102 PMID: 17995170
12. Rosso OA, Olivares F, Plastino A. Noise versus chaos in a causal Fisher-Shannon plane. Papers in Physics. 2015; 7: 070006. doi: 10.4279/pip.070006

13. Ortega-Garcia J, Fierrez-Aguilar J, Simon D, Gonzalez J, Faundez-Zanuy M, Espinosa V, et al. MCYT baseline corpus: a bimodal biometric database. IEE Proceedings Vision, Image and Signal Processing. 2003; 150: 395–401. doi: 10.1049/ip-vis:20031078

14. Hlaváčková-Schindler K, Paluš M, Vejmelka M, Bhattacharya J. Causality detection based on information-theoretic approaches in time series analysis. Physics Reports. 2007; 441: 1–46. doi: 10.1016/j.physrep.2006.12.004

15. Gao Z-K, Jin N-D. A directed weighted complex network for characterizing chaotic dynamics from time series. Nonlinear Analysis: Real World Applications. 2012; 13: 947–952. doi: 10.1016/j.nonrwa.2011.08.029

16. Gao Z-K, Fang P-C, Ding M-S, Jin N-D. Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow. Experimental Thermal and Fluid Science. 2015; 60: 157–164. doi: 10.1016/j.expthermflusci.2014.09.008

17. Ravetti MG, Carpi LC, Amin B., Frery AC, Rosso OA. Distinguishing noise from chaos: objective versus subjective criteria using horizontal visibly graph Plos One. 2014; 9: e108004. doi: 10.1371/journal.pone.0108004 PMID: 25247303

18. Schieber TA, Carpi LC, Frery AC, Rosso OA, Pardalos PM, Ravetti MG. Information Theory Perspective on Network Robustness Physics Letter A. 2016; 380: 359–364. doi: 10.1016/j.physleta.2015.10.055

19. Gonçalves BA, Carpi LC, Rosso OA, Ravetti MG. Time Series Characterization via Horizontal Visibility Graph and Information Theory. Physica A. 2016, in press.

20. Gao ZK, Yang Y-X, Zhai L-S, Ding M-S, Jin N-D. Characterizing slug to churn flow transition by using multivariate pseudo Wigner distribution and multivariate multiscale entropy. Chemical Engineering Journal. 2016; 291: 74–81. doi: 10.1016/j.cej.2016.01.039

21. Bandt C, Pompe B. Permutation Entropy: A Natural Complexity Measure for Time Series. Physical Review Letters. 2002, 88: 174102. doi: 10.1103/PhysRevLett.88.174102 PMID: 12005759

22. Fierrez J, Ortega-Garcia J, Ramos D, Gonzalez-Rodriguez J. HMM-based on-line signature verification: Feature extraction and signature modeling. Pattern Recognition Letters. 2007; 28(16): 2325–2334. Available from: http://www.sciencedirect.com/science/article/pii/S0167865507002395 doi: 10.1016/j.patrec.2007.07.012

23. Garcia-Salicetti S, Houmani N Ly-Van B, Dorizzi B, Alonso-Fernandez F, Fierrez J, Ortega-Garcia J, et al. Online handwritten signature verification. In: Petrovska-Delacrétaz D, Chollet G, Dorizzi B, editors. Guide to Biometric Reference Systems and Performance Evaluation. London: Springer-Verlag; 2009. p. 125–165.

24. Baron R, Plamondon R. Acceleration measurement with an instrumented pen for signature verification and handwriting analysis. IEEE Transactions on Instrumentation and Measurement. 1989; 38(6): 1132–1138. doi: 10.1109/19.46414

25. Boulétreau V, Vincent N, Sabourin R, Emptoz H. Handwriting and signature: one or two personality identifiers? In: Proceedings. Fourteenth International Conference on Pattern Recognition. vol. 2; 1998. p. 1758–1760.

26. Houmani N, Garcia-Salicetti. Quality measures for online handwritten signatures. In: Scharcanski J et al., editors. Signal and Image Processing for Biometrics. No. 292 in Lecture Notes in Electrical Engineering. Springer; 2014. p. 255–283. doi: 10.1007/978-3-642-54080-6_10

27. Vincent N, Boulétreau V, Empotz H, Sabourin R. How to use fractal dimensions to qualify writings and writers. Fractals. 2000; 8: 85–97. doi: 10.1142/S0218348X0000010X

28. Richards JA, Jia X. Remote Sensing Digital Image Analysis. 4th ed. Berlin: Springer; 2006. doi: 10.1007/3-540-29711-1

29. Lin J. Divergence measures based on the Shannon Entropy. IEEE Transactions on Information Theory. 1991; 37(1): 145–151. doi: 10.1109/18.61115

30. Grosse I, Bernaola-Galván P, Carpena P, Román-Roldán R, Oliver J, Stanley HE. Analysis of symbolic sequences using the Jensen-Shannon divergence. Physical Review E. 2002; 65: 041905. doi: 10.1103/PhysRevE.65.041905 PMID: 12005871

31. Ré MA, Azad RK. Generalization of entropy based divergence measures for symbolic sequence analysis. Plos ONE. 2014; 9(4): e93532. doi: 10.1371/journal.pone.0093532 PMID: 24728338

32. Sánchez-Moreno P, Dehesa JS, Yáñez RJ. Discrete densities and Fisher Information. Proceedings of the 14th International Conference on Difference Equations and Applications. Uğur-Bahçeşehir University Publishing Company, Istanbul, Turkey. Difference Equations and Applications, 2009; 291–298.

33. Sánchez-Moreno P, Zarzo A., Dehesa JS. Jensen divergence based on Fisher's information. Journal of Physics A: Mathematical and Theoretical. 2012, 45: 125305. doi: 10.1088/1751-8113/45/12/125305

34. Campbell C, Ying Y. Learning with Support Vector Machines. In: Brachman RJ, Dietterich T, editors. Synthesis Lectures on Artificial Intelligence and Machine Learning. No. 5 in Synthesis Lectures on Artificial Intelligence and Machine Learning. Santa Fe, CA: Morgan and Claypool; 2011. p. 1–95. doi: 10.2200/S00324ED1V01Y201102AIM010

35. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. Pittsburgh: ACM Press; 1992. p. 144–152.

36. Vapnik VN. The Nature of Statistical Learning Theory. Springer; 1995.

37. Chang CC, Lin CJ. LIBSVM: a library for support vector machines; 2001. Available from: http://www.csie.ntu.edu.tw/~cjlin/libsvm

38. Fierrez-Aguilar J, Nanni L, Lopez-Peñalba J, Ortega-Garcia J, Maltoni D. An On-Line Signature Verification System Based on Fusion of Local and Global Information. In: Kanade T, Jain A, Ratha N, editors. 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA). vol. 3546 of Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2005. p. 523–532. Available from: http://dx.doi.org/10.1007/11527923_54

39. Pascual-Gaspar JM, Cardeñoso Payo V, Vivaracho-Pascual CE. Practical On-Line Signature Verification. In: Tistarelli M, Nixon MS, editors. Proceedings Third International Conference Advances in Biometrics ICB. vol. 5558 of Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2009. p. 1180–1189. Available from: http://dx.doi.org/10.1007/978-3-642-01793-3_119

40. Gao ZK, Yang Y-X, Fang P-C, Zou Y, Xia C-Y, Du M. Multiscale complex network for analyzing experimental multivariate time series. Europhysics Letters. 2015; 109: 30005. doi: 10.1209/0295-5075/109/30005

41. Gao ZK, Yang Y-X, Fang P-C, Jin N-D, Xia C-Y, Hu L-D. Multi-frequency complex network from time series for uncovering oil-water flow structure. Scientific Reports. 2015; 5: 8222. doi: 10.1038/srep08222 PMID: 25649900

42. Zunino L, Olivares F, Rosso OA. Permutation min-entropy: an improved quantifier for unveiling subtle temporal correlations. Europhysics Letters. 2015; 109: 10005. doi: 10.1209/0295-5075/109/10005

43. Zhang L, Zhang L, Du B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. IEEE Geoscience and Remote Sensing Magazine. 2016; 4: 22–40. doi: 10.1109/MGRS.2016.2540798

# Information Theory quantifiers

Physics, as well as, other scientific disciplines like biology or finance, can be considered observational sciences, that is, they try to infer properties of an unfamiliar system from the analysis of measured time record of it behavior (time series). Dynamical systems are systems that evolve in time. In practice, one may only be able to measure a scalar time series $\mathcal{X}(t)$ which may be a function of variables $\mathcal{V} = \{v_1, v_2, \cdots, v_k\}$ describing the underlying dynamics (i.e. $d\mathcal{V}/dt = f(\mathcal{V})$. Then, the natural question is, from $\mathcal{X}(t)$ how much we can learn about the dynamics of the system. In a more formal way, given a system, be it natural or man-made, and given an observable of such system whose evolution can be tracked through time, a natural question arises: how much information is this observable encoding about the dynamics of the underlying system? The information content of a system is typically evaluated via a probability distribution function (PDF) $P$ describing the apportionment of some measurable or observable quantity, generally a time series $\mathcal{X}(t)$. Quantifying the information content of a given observable is therefore largely tantamount to characterizing its probability distribution. This is often done with the wide family of measures called Information Theory quantifiers [1]. We can define Information Theory quantifiers as measures able to characterize relevant properties of the PDF associated with these time series, and in this way we should judiciously extract information on the dynamical system under study.

### Shannon entropy, Fisher Information measure, and Statistical Complexity

Entropy is a basic quantity with multiple field-specific interpretations: for instance, it has been associated with disorder, state-space volume, and lack of information [2]. When dealing with information content, the Shannon entropy is often considered as the foundational and most natural one [3, 4].

Entropy, regarded as a measure of uncertainty, is the most paradigmatic example of these information quantifiers. Given a continuous probability distribution function (PDF) $\rho(x)$ with $x \in \Omega \subset \mathbb{R}$ and $\int_\Omega \rho(x)\, dx = 1$, its associated *Shannon Entropy* S [3, 4] is defined by

$$\mathrm{S}[\rho] \;=\; -\int_\Omega \rho(x) \ln\left[\rho(x)\right] dx. \tag{S1.1}$$

It is a global measure, that is, it is not too sensitive to strong changes in the distribution taking place on a small-sized region of $\Omega$. Such is not the case with *Fisher's Information Measure* (FIM) $\mathcal{F}$[5, 6], which constitutes a measure of the gradient content of the distribution $\rho$, thus being quite sensitive even to tiny localized perturbations. It reads

$$\mathcal{F}[\rho] \;=\; \int \frac{|\vec{\nabla}\rho(x)|^2}{\rho(x)}\, dx \;\;=\; 4\int |\vec{\nabla}\psi(x)|^2\, dx, \quad \text{where } \psi(x) \;=\; \sqrt{\rho(x)}\,. \tag{S1.2}$$

The Fisher Information Measure can be variously interpreted as a measure of the ability to estimate a parameter, as the amount of information that can be extracted from a set of measurements, and also as a measure of the state of disorder of a system or phenomenon[6], its most important property being the so-called Cramer-Rao bound. It is important to remark that the gradient operator significantly influences the contribution of minute local $\rho$-variations to the Fisher information value, accordingly, this quantifier is called "local" [6]. Note that the Shannon entropy decreases with the distribution skewness, while the Fisher information increases.

Local sensitivity is useful in scenarios whose description necessitates an appeal to a notion of "order". In the previous definition of FIM (Eq. (S1.2)) the division by $\rho(x)$ is not convenient

if $\rho(x) \to 0$ at certain points of the support $\Omega$. We avoid this if we work with real probability amplitudes, by means of the alternative expression that employs $\psi(x)$ [5, 6]. This form requires no divisions, and shows that $\mathcal{F}$ simply measures the gradient content in $\psi(x)$.

Let now $P = \{p_i; i = 1, \ldots, N\}$ with $\sum_{i=1}^{N} p_i = 1$, be a discrete probability distribution, with $N$ the number of possible states of the system under study. The Shannon's logarithmic information measure reads

$$\mathrm{S}[P] \;=\; -\sum_{i=1}^{N} p_i \ln [p_i] \;. \tag{S1.3}$$

This can be regarded to as a measure of the uncertainty associated (information) to the physical process described by $P$. For instance, if $\mathrm{S}[P] = \mathrm{S}_{\min} = 0$, we are in position to predict with complete certainty which of the possible outcomes $i$, whose probabilities are given by $p_i$, will actually take place. Our knowledge of the underlying process described by the probability distribution is maximal in this instance. In contrast, our knowledge is minimal for a uniform distribution $P_e = \{p_i = 1/N, \forall i = 1, \ldots, N\}$ since every outcome exhibits the same probability of occurrence, and the uncertainty is maximal, i.e., $\mathrm{S}[P_e] = \mathrm{S}_{\max} = \ln N$. In the discrete case, we define a "normalized" Shannon entropy, $0 \leq \mathcal{H} \leq 1$, as

$$\mathcal{H}[P] \;=\; \mathrm{S}[P]/\mathrm{S}_{\max} \;. \tag{S1.4}$$

The concomitant problem of loss of information due to the discretization has been thoroughly studied (see, for instance, [7, 8] and references therein) and, in particular, it entails the loss of Fisher's shift-invariance, which is of no importance for our present purposes. For the FIM we take the expression in terms of real probability amplitudes as starting point, then a discrete normalized FIM, $0 \leq \mathcal{F} \leq 1$, convenient for our present purposes, is given by

$$\mathcal{F}[P] \;=\; F_0 \sum_{i=1}^{N-1} \left[\sqrt{p_{i+1}} - \sqrt{p_i}\right]^2 \;. \tag{S1.5}$$

It has been extensively discussed that this discretization is the best behaved in a discrete environment [9, 10]. Here the normalization constant $F_0$ reads

$$F_0 \;=\; \begin{cases} 1, & \text{if } p_{i^*} = 1 \text{ for } i^* = 1 \text{ or } i^* = N \text{ and } p_i = 0, \forall i \neq i^*, \\ 1/2, & \text{otherwise .} \end{cases} \tag{S1.6}$$

Complexity denotes a state of affairs that one can easily appreciate when confronted with it; however, is rather difficult to define it quantitatively, probably due to the fact that there is no universal definition of complexity. In between the two special instances of perfect order and complete randomness, a wide range of possible degrees of physical structure exists that should be reflected in the features of the underlying probability distribution $P$. One would like to assume that the degree of correlational structures would be adequately captured by some functional $\mathcal{C}[P]$ in the same way that Shannon's entropy $\mathrm{S}[P]$ [3] "captures" randomness.

Clearly, the ordinal structures present in a process is not quantified by randomness measures, and consequently, measures of statistical or structural complexity are necessary for a better understanding (characterization) of the system dynamics represented by their time series [11]. The opposite extremes of perfect order and maximal randomness are very simple to describe, because they do not have any structure. The complexity should be zero in these cases. At a given distance from these extremes, a wide range of possible ordinal structures exists. Complexity can be characterized by a certain degree of organization, structure, memory, regularity, symmetry, and patterns

[12]. The complexity measure does much more than satisfy the boundary conditions of vanishing in the high- and low-entropy limits. In particular the maximum complexity occurs in the region between the system's perfectly ordered state and the perfectly disordered one. Complexity is allows us to detect essential details of the dynamics, and more importantly to characterize the correlational structure of the orderings present in the time series.

The perfect crystal and the isolated ideal gas are two typical examples of systems with minimum and maximum entropy, respectively. However, they are also examples of simple models and therefore of systems with zero complexity, as the structure of the perfect crystal is completely described by minimal information (i.e., distances and symmetries that define the elementary cell) and the probability distribution for the accessible states is centered around a prevailing state of perfect symmetry. On the other hand, all the accessible states of the ideal gas occur with the same probability and can be described by a "simple" uniform distribution.

Statistical complexity is often characterized by the paradoxical situation of a complicated dynamics generated from relatively simple systems. Obviously, if the system itself is already involved enough and is constituted by many different parts, it clearly may support a rather intricate dynamics, but perhaps without the emergence of typical characteristic patterns [13]. Therefore, a complex system does not necessarily generate a complex output. Statistical complexity is therefore related to patterned structures hidden in the dynamics, emerging from a system which itself can be much simpler than the dynamics it generates [13].

According to López-Ruiz, Mancini and Calbet [14], and using an oxymoron, an object, a procedure, or system is said to be complex when it does not exhibit patterns regarded as simple. It follows that a suitable complexity measure should vanish both for completely ordered and for completely random systems and cannot only rely on the concept of information (which is maximal and minimal for the above mentioned systems). A suitable measure of complexity can be defined as the product of a measure of information and a measure of disequilibrium, i.e. some kind of distance from the equiprobable distribution of the accessible states of a system. In this respect, Rosso and coworkers [16] introduced an effective *Statistical Complexity Measure* (SCM) $\mathcal{C}$, that is able to detect essential details of the dynamical processes underlying the dataset.

Based on the seminal notion advanced by López-Ruiz *et al.* [14], this statistical complexity measure[15, 16] is defined through the functional product form

$$\mathcal{C}[P] \;=\; \mathcal{Q}_J[P, P_e] \cdot \mathcal{H}[P] \tag{S1.7}$$

of the normalized Shannon entropy $\mathcal{H}$, see Eq. (S1.4), and the disequilibrium $\mathcal{Q}_J$ defined in terms of the Jensen-Shannon divergence $\mathcal{J}[P, P_e]$. That is,

$$\mathcal{Q}_J[P, P_e] \;=\; Q_0 \mathcal{J}[P, P_e] \;=\; Q_0\{\mathrm{S}[(P + P_e)/2] - \mathrm{S}[P]/2 - \mathrm{S}[P_e]/2\}, \tag{S1.8}$$

the above-mentioned Jensen-Shannon divergence and $Q_0$, a normalization constant such that $0 \leq \mathcal{Q}_J \leq 1$:

$$Q_0 \;=\; -2\left\{\frac{N+1}{N}\ln(N+1) - \ln(2N) + \ln N\right\}^{-1}, \tag{S1.9}$$

are equal to the inverse of the maximum possible value of $\mathcal{J}[P, P_e]$. This value is obtained when one of the components of $P$, say $p_m$, is equal to one and the remaining $p_j$ are zero.

The Jensen-Shannon divergence, which quantifies the difference between probability distributions, is especially useful to compare the symbolic composition between different sequences[17, 18, 19]. Note that the above introduced SCM depends on two different probability distributions: one

associated with the system under analysis, $P$, and the other the uniform distribution, $P_e$. Furthermore, it was shown that for a given value of $\mathcal{H}$, the range of possible $\mathcal{C}$ values varies between a minimum $\mathcal{C}_{min}$ and a maximum $\mathcal{C}_{max}$, restricting the possible values of the SCM [20]. Thus, it is clear that important additional information related to the correlational structure between the components of the physical system is provided by evaluating the statistical complexity measure.

If our system, with associated discrete PDF, lies in a very ordered state, will be represented by an extremely narrow PDF, that is almost all the $p_i$–values are almost zero except for a particular state $k \neq i$ with $p_k \cong 1$,then both the normalized Shannon entropy and statistical complexity are close to zero ($\mathcal{H} \approx 0$ and $\mathcal{C} \approx 0$), and the normalized Fisher's information measure is close to one ($\mathcal{F} \approx 1$). On the other hand, when the system under study is represented by a very disordered state, that is when all the $p_i$–values oscillate around the same value, we have $\mathcal{H} \approx 1$ while $\mathcal{C} \approx 0$ and $\mathcal{F} \approx 0$. One can state that the general FIM–behavior of the present discrete version (Eq. (S1.5)), is opposite to that of the Shannon entropy, except for periodic motions. The local sensitivity of FIM for discrete–PDFs is reflected in the fact that the specific "$i-$ordering" of the discrete values $p_i$ must be seriously taken into account in evaluating the sum in Eq. (S1.5). This point was extensively discussed by Rosso and co-workers [21, 22]. The summands can be regarded to as a kind of "distance" between two contiguous probabilities. Thus, a different ordering of the pertinent summands would lead to a different FIM-value, hereby its local nature. In the present work, we follow the Lehmer lexicographic order [23] in the generation of Bandt and Pompe PDF (see next section). Given the local character of FIM, when combined with a global quantifier as the normalized Shannon entropy, conforms the Shannon–Fisher plane, $\mathcal{H} \times \mathcal{F}$, introduced by Vignat and Bercher [24]. These authors showed that this plane is able to characterize the non-stationary behavior of a complex signal.

## The Bandt and Pompe approach to the PDF determination

The evaluation of the Information Theory derived quantifiers, like those previously introduced (Shannon entropy, Fisher information and statistical complexity), suppose some prior knowledge about the system; specifically, a probability distribution associated to the time series under analysis should be provided beforehand. The determination of the most adequate PDF is a fundamental problem because the PDF $P$ and the sample space $\Omega$ are inextricably linked.

Usual methodologies assign to each time point of the series $\mathcal{X}(t)$ a symbol from a finite alphabet $\mathfrak{A}$, thus creating a *symbolic sequence* that can be regarded to as a *non causal coarse grained* description of the time series under consideration. As a consequence, order relations and the time scales of the dynamics are lost. The usual histogram technique corresponds to this kind of assignment. *Causal information* may be duly incorporated if information about the past dynamics of the system is included in the symbolic sequence, i.e., symbols of alphabet $\mathfrak{A}$ are assigned to a portion of the phase-space or trajectory.

Many methods have been proposed for a proper selection of the probability space $(\Omega, P)$. Among others, of type non causal coarse grained, we can mention frequency counting [25], procedures based on amplitude statistics [26], binary symbolic dynamics [27], Fourier analysis [28], or wavelet transform [29]. The suitability of each of the proposed methodologies depends on the peculiarity of data, such as stationarity, length of the series, the variation of the parameters, the level of noise contamination, etc. In all these cases, global aspects of the dynamics can be somehow captured, but the different approaches are not equivalent in their ability to discern all relevant physical details. Bandt and Pompe (BP)[30] introduced a simple and robust symbolic methodology that takes into account time causality of the time series (causal coarse grained methodology)

by comparing neighboring values in a time series. The symbolic data are: *(i)* created by ranking the values of the series; and *(ii)* defined by reordering the embedded data in ascending order, which is tantamount to a phase space reconstruction with embedding dimension (pattern length) $D$ and time lag $\tau$. In this way, it is possible to quantify the diversity of the ordering symbols (patterns) derived from a scalar time series.

Note that the appropriate symbol sequence arises naturally from the time series, and no model-based assumptions are needed. In fact, the necessary "partitions" are devised by comparing the order of neighboring relative values rather than by apportioning amplitudes according to different levels. This technique, as opposed to most of those in current practice, takes into account the temporal structure of the time series generated by the physical process under study. As such, it allows us to uncover important details concerning the ordinal structure of the time series [31, 32, 33, 34] and can also yield information about temporal correlation [35, 36].

It is clear that this type of analysis of a time series entails losing details of the original series' amplitude information. Nevertheless, by just referring to the series' intrinsic structure, a meaningful difficulty reduction has indeed been achieved by BP with regard to the description of complex systems. The symbolic representation of time series by recourse to a comparison of consecutive ($\tau = 1$) or nonconsecutive ($\tau > 1$) values allows for an accurate empirical reconstruction of the underlying phase-space, even in the presence of weak (observational and dynamic) noise [30]. Furthermore, the ordinal patterns associated with the PDF are invariant with respect to nonlinear monotonous transformations. Accordingly, nonlinear drifts or scaling artificially introduced by a measurement device will not modify the estimation of quantifiers, a nice property if one deals with experimental data (see, e.g., [37]). These advantages make the BP methodology more convenient than conventional methods based on range partitioning, i.e., a PDF based on histograms.

To use the BP methodology[30] for evaluating the PDF, $P$, associated with the time series (dynamical system) under study, one starts by considering partitions of the $D$-dimensional space that will hopefully "reveal" relevant details of the ordinal structure of a given one-dimensional time series $\mathcal{X}(t) = \{x_t; t = 1, \ldots, M\}$ with embedding dimension $D > 1$ ($D \in \mathbb{N}$) and time lag $\tau$ ($\tau \in \mathbb{N}$). We are interested in "ordinal patterns" of order (length) $D$ generated by

$$(s) \mapsto \left(x_{s-(D-1)\tau}, x_{s-(D-2)\tau}, \ldots, x_{s-\tau}, x_s\right) \ , \tag{S1.10}$$

which assign to each time $s$ the $D$-dimensional vector of values at times $s-(D-1)\tau, \ldots, s-\tau, s$. Clearly, the greater $D$, the more information on the past is incorporated into our vectors. By "ordinal pattern" related to the time $(s)$, we mean the permutation $\pi = (r_0, r_1, \ldots, r_{D-1})$ of $[0, 1, \ldots, D-1]$ defined by

$$x_{s-r_{D-1}\tau} \leq \ x_{s-r_{D-2}\tau} \leq \cdots \leq \ x_{s-r_1\tau} \leq x_{s-r_0\tau}. \tag{S1.11}$$

In this way the vector defined by Eq. (S1.10) is converted into a unique symbol $\pi$. We set $r_i < r_{i-1}$ if $x_{s-r_i} = x_{s-r_{i-1}}$ for uniqueness, although ties in samples from continuous distributions have null probability.

In order to illustrate BP method, we will consider a simple example: a time series with seven ($M = 7$) values $\mathcal{X} = \{4, 7, 9, 10, 6, 11, 3\}$ and we evaluate the BP-PDF for $D = 3$ and $\tau = 1$. In this case the state space is divided into 3! partitions and 6 mutually exclusive permutation symbols are considered. The triplet $(4, 7, 9)$ and $(7, 9, 10)$ represent the permutation pattern [012] since they are in increasing order. On the other hand, $(9, 10, 6)$ and $(6, 11, 3)$ correspond to the permutation pattern [201] since $x_{s+2} < x_s < x_{s+1}$, while $(10, 6, 11)$ has the permutation

pattern [102] with $x_{s+1} < x_s < x_{s+2}$. Then, the associated probabilities to the 6 patterns are: $p([012]) = p([201]) = 2/5$; $p([102]) = 1/5$; $p([021]) = p([120]) = p([210]) = 0$.

Fig A illustrates the construction principle of the ordinal patterns of length $D = 2$, 3 and 4 with $\tau = 1$[38]. Consider the sequence of observations $\{x_0, x_1, x_2, x_3\}$. For $D = 2$, there are only two possible directions from $x_0$ to $x_1$: up and down. For $D = 3$, starting from $x_1$ (up) the third part of the pattern can be above $x_1$, below $x_0$, or between $x_0$ and $x_1$. A similar situation can be found starting from $x_1$ (down). For $D = 4$, for each one of the six possible positions for $x_2$, there are four possible localizations for $x_3$, yielding $D! = 4! = 24$ different possible ordinal patterns. In Fig A, full circles and continuous lines represent the sequence values $x_0 < x_1 > x_2 > x_3$, which leads to the pattern $\pi = [0321]$. A graphical representation of all possible patterns corresponding to $D = 3, 4$ and 5 can be found in Fig 2 of Parlitz *et al.*[38].

For all the $D!$ possible orderings (permutations) $\pi_i$ when embedding dimension is $D$, and time-lag $\tau$, their relative frequencies can be naturally computed according to the number of times this particular order sequence is found in the time series, divided by the total number of sequences,

$$p(\pi_i) = \frac{\#\{s | s \leq N - (D-1)\tau; (s) \text{ is of type } \pi_i\}}{N - (D-1)\tau}, \tag{S1.12}$$

where # denotes cardinality. Thus, an ordinal pattern probability distribution $P = \{p(\pi_i), i = 1, \ldots, D!\}$ is obtained from the time series.

The embedding dimension $D$ plays an important role in the evaluation of the appropriate probability distribution, because $D$ determines the number of accessible states $D!$ and also conditions the minimum acceptable length $M \gg D!$ of the time series that one needs in order to work with reliable statistics [39]. In the present work, we follow the Lehmer lexicographic order [23] in the generation of Bandt and Pompe PDF.

Regarding the selection of the parameters, Bandt and Pompe suggested working with $4 \leq D \leq 6$, and specifically considered a time lag $\tau = 1$ in their cornerstone paper [30]. Nevertheless, it is clear that other values of $\tau$ could provide additional information. It has been recently shown that this parameter is strongly related, if it is relevant, to the intrinsic time scales of the system under analysis [40, 41, 42].

Additional advantages of the method reside in *i)* its simplicity (it requires few parameters: the pattern length/embedding dimension $D$ and the time lag $\tau$), and *ii)* the extremely fast nature of the calculation process. The BP methodology can be applied not only to time series representative of low dimensional dynamical systems, but also to any type of time series (regular, chaotic, noisy, or reality based). In fact, the existence of an attractor in the $D$-dimensional phase space in not assumed. The only condition for the applicability of the BP method is a very weak stationary assumption: for $k \leq D$, the probability for $x_t < x_{t+k}$ should not depend on $t$. For a review of BP's methodology and its applications to physics, biomedical and econophysics signals see Zanin *et al.* [43]. Moreover, Rosso *et al.* [31] show that the above mentioned quantifiers produce better descriptions of the process associated dynamics when the PDF is computed using BP rather than using the usual histogram methodology.

The BP proposal for associating probability distributions to time series (of an underlying symbolic nature) constitutes a significant advance in the study of nonlinear dynamical systems [30]. The method provides univocal prescription for ordinary, global entropic quantifiers of the Shannon-kind. However, as was shown by Rosso and coworkers [21, 22], ambiguities arise in applying the BP technique with reference to the permutation of ordinal patterns. This happens if one wishes to employ the BP-probability density to construct local entropic quantifiers, like the Fisher information measure, which would characterize time series generated by nonlinear

dynamical systems.

The local sensitivity of the Fisher information measure for discrete PDFs is reflected in the fact that the specific "$i$-ordering" of the discrete values $p_i$ must be seriously taken into account in evaluating Eq. (S1.5). The numerator can be regarded to as a kind of "distance" between two contiguous probabilities. Thus, a different ordering of the summands will lead, in most cases, to a different Fisher information value. In fact, if we have a discrete PDF given by $P = \{p_i, i = 1, \ldots, N\}$, we will have $N!$ possibilities for the $i$-ordering.

The question is, which is the arrangement that one could regard as the "proper" ordering? The answer is straightforward in some cases, like the one that pertains to histogram-based PDFs. For extracting a time-series PDF $P$ via an histogram procedure, one first divides the interval $[a, b]$ (with $a$ and $b$ the minimum and maximum amplitude values in the time series) into a finite number $N_{bin}$ ($N \equiv N_{bin}$ in Eqs. (S1.3) and (S1.5)) of non overlapping equal sized consecutive subintervals $A_k : [a, b] = \bigcup_{k=1}^{N_{bin}} A_k$ and $A_n \bigcap A_m = \emptyset \ \forall n \neq m$. Then, recourse to the usual histogram method, based on counting the relative frequencies of the time series' values within each subinterval, is made. Of course, in this approach the temporal order in which the time-series values emerge plays no role at all. The only pieces of information we have here are the $x_t$-values that allow one to assign inclusion within a given bin, ignoring just where they are located (this is, the subindex $i$). Note that the division procedure of the interval $[a, b]$ provides the natural order sequence for the evaluation of the PDF gradient involved in Fisher's information measure.

From now on, we assume that the all Information Theory quantifiers will be evaluated with BP-PDF's, by this reason usually they are called permutation quantifiers or time causal Information Theory quantifiers. In our current paper, we chose the lexicographic ordering given by the algorithm of Lehmer [23], among other possibilities, due to its better distinction of different dynamics in the Shannon–Fisher plane, $\mathcal{H} \times \mathcal{F}$ (see [21, 22]).

## Causal information planes

In statistical mechanics one is often interested in isolated systems characterized by an initial, arbitrary, and discrete probability distribution. Evolution towards equilibrium is to be described, as the overriding goal. At equilibrium, we can suppose, without loss of generality, that this state is given by the equiprobable distribution $P_e = \{p_i = 1/N, \forall i = 1, \ldots, N\}$. The temporal evolution of the above introduced Information Theory quantifiers, Shannon entropy $\mathcal{H}$, statistical complexity $\mathcal{C}$ and Fisher information measure $\mathcal{F}$, can be analyzed using a two-dimensional (2D) diagrams of the corresponding quantifiers versus time $t$. However, the second law of thermodynamics states that, for isolated systems, entropy grows monotonically with time ($d\mathcal{H}/dt \geq 0$) [44]. This implies that entropy $\mathcal{H}$ can be regarded as an arrow of time, so that an equivalent way to study the temporal evolution of these quantifiers is using the normalized entropy $\mathcal{H}$ as substitute for the time-axis.

Two causal information planes are defined (the term causality remembers the fact that temporal correlations between successive samples are taken into account through the Bandt and Pompe PDF recipe used to estimate both Information Theory quantifiers): *a)* The *causality entropy–complexity plane*, $\mathcal{H} \times \mathcal{C}$, is based only on global characteristics of the associated time series PDF (both quantities are defined in terms of Shannon entropies); while *b)* the *causality Shannon-Fisher plane*, $\mathcal{H} \times \mathcal{F}$, is based on global and local characteristics of the PDF. In the case of $\mathcal{H} \times \mathcal{C}$ the variation range is $[0, 1] \times [\mathcal{C}_{min}, \mathcal{C}_{max}]$ (with $\mathcal{C}_{min}$ and $\mathcal{C}_{max}$ the minimum and maximum statistical complexity values, respectively, for a given $\mathcal{H}$ value [20]), while in the causality plane $\mathcal{H} \times \mathcal{F}$ the range is $[0, 1] \times [0, 1]$.

These two diagnostic tools were shown to be particularly efficient to distinguish between the deterministic chaotic and stochastic nature of a time series since the permutation quantifiers have distinctive behaviors for different types of motion. According to the findings obtained by Rosso *et al.* [31, 22, 45], chaotic maps have intermediate entropy $\mathcal{H}$ and Fisher $\mathcal{F}$ values, while complexity $\mathcal{C}$ reaches larger values, very close to those of the limit. For regular processes, entropy and complexity have small values, close to zero, while the Fisher is close to one. Finally, totally uncorrelated stochastic processes are located in the planar location associated with $\mathcal{H}$ near one and, $\mathcal{C}$, $\mathcal{F}$ near to zero, respectively. It has also been found that $1/f^{\alpha}$ correlated stochastic processes with $1 \leq \alpha \leq 3$ are characterized by intermediate permutation entropy and intermediate statistical complexity values [31], as well as, intermediate low Fisher information [31, 22, 45]. Moreover, note that in both causal information planes the localization of these stochastic behavior look like a separation border with respect to chaotic, which are localized above of it. In addition, these two causal information planes have been profitably used to visualization and characterization of different dynamical regimes when the system parameters vary [21, 22, 46, 47, 48, 49, 50, 51, 52, 53, 54]; to study time dynamic evolution [39, 55, 56]; identifying periodicities in natural time series [57]; identification of deterministic dynamics contaminated with noise [32, 33]; estimating intrinsic time scales and delayed systems [40, 41, 42, 58]; characterization of pseudo-random number generators [59, 60] measure of complexity of two-dimensional patterns [61] among other biomedical and econophysics applications (see [43] and references therein).

## References


[1] Gray, RM (1990) Entropy and Information Theory. Springer, Berlin-Heidelberg, Germany.

[2] Brissaud, JB (2005) The meaning of entropy. Entropy 7: 68–96.

[3] Shannon CE (1948) A mathematical theory of communication. Bell Syst. Technol. J. 27: 379–423, 623–56.

[4] Shannon, CE, Weaver, W (1949) The Mathematical Theory of Communication; University of Illinois Press: Champaign, IL, USA.

[5] Fisher, RA (1922) On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond. Ser. A 222: 309–368.

[6] Frieden RB (2004) Science from Fisher information: A Unification. Cambridge University Press, Cambridge, UK.

[7] Zografos K, Ferentinos K, Papaioannou T (1986) Discrete approximations to the Csiszár, Renyi, and Fisher measures of information. Canad. J. Stat. 14: 355–366 .

[8] Pardo L, Morales D, Ferentinos K, Zografos K (1994) Discretization problems on generalized entropies and R-divergences. Kybernetika 30: 445–460 .

[9] Sánchez-Moreno P, Dehesa JS, Yáñez RJ (2009) Discrete densities and Fisher Information. Proceedings of the 14th International Conference on Difference Equations and Applications. Uğur-Bahçeşehir University Publishing Company, Istanbul, Turkey Difference Equations and Applications: 291–298.

[10] Sánchez-Moreno P, Zarzo A, Dehesa JS (2012) Jensen divergence based on Fisher's information. Journal of Physics A: Mathematical and Theoretical. 45: 125305.

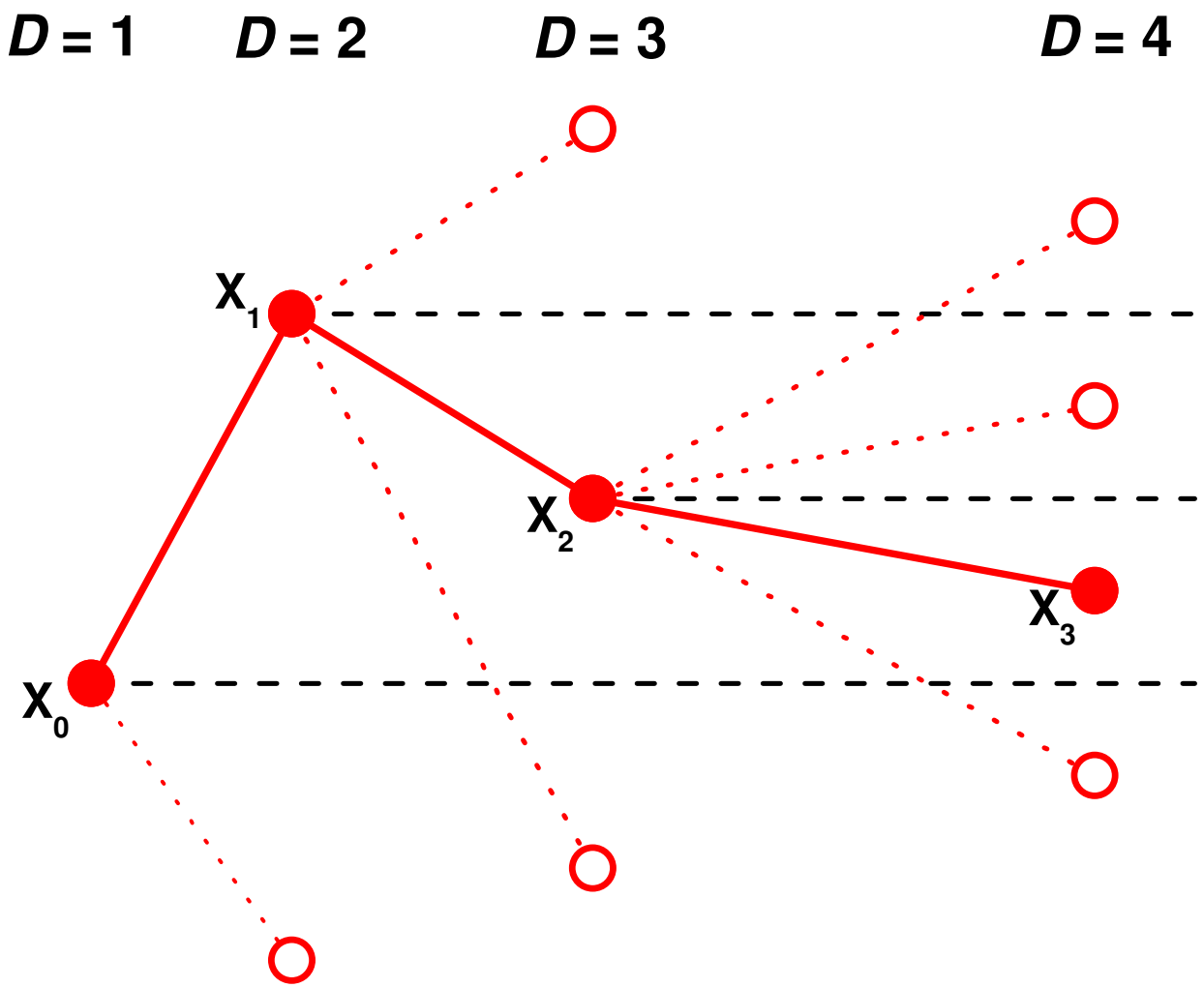


Figure A: Illustration of the construction principle for ordinal patterns of length $D$ [38]. If $D = 4$ and $\tau = 1$, full circles and continuous lines represent the sequence of values $x_0 < x_1 > x_2 > x_3$ which lead to the pattern $\pi = [0321]$.

[11] Feldman DP, Crutchfield JP (1998) Measures of Statistical Complexity: Why? Phys. Lett. A 238: 244– 252.

[12] Feldman DP, McTague CS, Crutchfield JP (2008) The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing. Chaos 18: 043106.

[13] Kantz H, Kurths J, Meyer-Kress G (1998) Nonlinear Analysis of Physiological Data, Springer, Berlin.

[14] López-Ruiz R, Mancini HL, Calbet X (1995) A statistical measure of complexity. Phys. Lett. A 209: 321–326.

[15] Martín, MT, Plastino, A, Rosso, OA (2003) Statistical complexity and disequilibrium, Phys. Lett. A 311: 126–132.

[16] Lamberti PW, Martín MT, Plastino A, Rosso OA (2004) Intensive entropic non-triviality measure. Physica A 334: 119–131.

[17] Lin J (1991) Divergence measures based on the Shannon Entropy IEEE Transactions on Information Theory 37(1): 145–151.

[18] Grosse I, Bernaola-Galván P, Carpena P, Román-Roldán R, Oliver J, Stanley HE (2002) Analysis of symbolic sequences using the Jensen-Shannon divergence. Phys. Rev. E 65: 041905.

[19] Ré MA, Azad RK (2014) Generalization of entropy based divergence measures for symbolic sequence analysis. Plos ONE 9(4): e93532.

[20] Martín MT, Plastino A, Rosso OA (2006) Generalized statistical complexity measures: Geometrical and analytical properties. Physica A 369: 439–462.

[21] Olivares F, Plastino A, Rosso OA (2012) Ambiguities in the Bandt-Pompe's methodology for local entropic quantifiers. Physica A 391: 2518–2526.

[22] Olivares F, Plastino A, Rosso OA (2012) Contrasting chaos with noise via local versus global information quantifiers, Phys. Lett. A 376: 1577–1583.

[23] Schwarz, K (2011) The Archive of Interesting Code, `http://www.keithschwarz.com/interesting/code/?dir=factoradic-permutation`

[24] Vignat C, Bercher JF (2003) Analysis of signals in the Fisher–Shannon information plane. Phys. Lett. A 312: 27–33.

[25] Rosso OA, Craig H, Moscato P (2009) Shakespeare and other english renaissance authors as characterized by Information Theory complexity quantifiers. Physica A 388: 916–926.

[26] De Micco L, González CM, Larrondo HA, Martín MT, Plastino A, Rosso OA (2008) Randomizing nonlinear maps via symbolic dynamics. Physica A 387: 3373–3383.

[27] Mischaikow K, Mrozek M, Reiss J, Szymczak A (1999) Construction of symbolic dynamics from experimental time series. Phys. Rev. Lett. 82 1114–1147.

[28] Powell GE, Percival I (1979) A spectral entropy method for distinguishing regular and irregular motion of hamiltonian systems. J. Phys. A: Math. Gen. 12 2053–2071.

[29] Rosso OA, Blanco S, Jordanova J, Kolev V, Figliola A, Schürmann M, Başar E (2001) Wavelet entropy: a new tool for analysis of short duration brain electrical signals. J. Neurosc. Meth. 105 65–75.

[30] Bandt C, Pompe B (2002) Permutation Entropy: A Natural Complexity Measure for Time Series. Phys. Rev. Lett. 88: 174102.

[31] Rosso OA, Larrondo HA, Martín MT, Plastino A, Fuentes MA (2007) Distinguishing noise from chaos. Phys. Rev. Lett. 99: 154102.

[32] Rosso OA, Carpi LC, Saco PM, Gómez Ravetti M, Plastino A, Larrondo H (2012) Causality and the Entropy-Complexity Plane: Robustness and Missing Ordinal Patterns Physica A 391 42–55.

[33] Rosso OA, Carpi LC, Saco PM, Gómez Ravetti M, Larrondo H, Plastino A(2012) The Amigó paradigm of forbidden/missing patterns: a detailed analysis Eur. Phys. J. B 85: 419–430.

[34] Rosso OA, Olivares F, Zunino L, De Micco L, Aquino ALL, Plastino A, Larrondo HA (2013) Characterization of chaotic maps using the permutation Bandt-Pompe probability distribution. Eur. Phys. J. B 86: 116–129.

[35] Rosso OA, Masoller C (2009) Detecting and quantifying stochastic and coherence resonances via information-theory complexity measurements. Phys. Rev. E 79: 040106(R).

[36] Rosso OA, Masoller C (2009) Detecting and quantifying temporal correlations in stochastic resonance via information theory measures. Eur. Phys. J. B 69: 37–43.

[37] Saco PM, Carpi LC, Figliola A, Serrano E, Rosso OA (2010) Entropy analysis of the dynamics of El Niño/Southern Oscillation during the Holocene. Physica A 389: 5022–5027.

[38] Parlitz U, Berg S, Luther S, Schirdewan A, Kurths J, Wessel N (2012) Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics. Comput. Biol. Med. 42: 319–327.

[39] Kowalski AM, Martín MT, Plastino A, Rosso OA (2007) Bandt-Pompe approach to the classical-quantum transition. Physica D 233: 21–31.

[40] Zunino L, Soriano MC, Fischer I, Rosso OA, Mirasso CR (2010) Permutation information-theory approach to unveil delay dynamics from time-series analysis. Phys. Rev. E 82: 046212.

[41] Soriano MC, Zunino L, Rosso OA, Fischer I, Mirasso CR (2011) Time scales of a chaotic semiconductor laser with optical feedback under the lens of a permutation information analysis. IEEE J. Quantum Electron. 47: 252–261.

[42] Zunino L, Soriano MC, Rosso OA (2012) Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. Phys. Rev. E 86: 046210.

[43] Zanin M, Zunino L, Rosso OA, Papo D (2012) Permutation entropy and its main biomedical and econophysics applications: A review. Entropy 14: 1553–1577.

[44] Plastino AR, Plastino A (1996) Symmetries of the Fokker-Plank equation and Fisher-Frieden arrow of time. Phys. Rev. E 54: 4423 – 4326.

[45] Rosso OA, Olivares F, Plastino A (2015) Noise versus chaos in a causal Fisher-Shannon plane. Papers in Physics 7 070006.

[46] Rosso OA, De Micco L, Plastino A, Larrondo H (2010) Info-quantifiers' map-characterization revisited. Physica A 389: 249–262.

[47] Kowalski AM, Martín MT, Plastino A, Rosso OA (2011) Fisher-information description of the classical-quantal transition. Physica A 390:2435–2441.

[48] De Micco L, Fernández JG, Larrondo HA, Plastino A, Rosso OA (2012) Sampling period, statistical complexity, and chaotic attractors. Physica A 391: 2564–2575.

[49] Lange H, Rosso OA, Hauhs M (2013) Ordinal pattern and statistical complexity analysis of daily stream flow time series Eur. Phys. J. Special Topics 222: 535–552.

[50] Serinaldi F, Zunino L, Rosso OA (2014) Complexity-entropy analysis of daily stream flow time series in the continental United States Stochastic Environmental Research and Risk Assessment 28: 1685–1708.

[51] Montani F, Deleglise EB, Rosso OA (2014) Efficiency characterization of a large neuronal network: a causal information approach Physica A 401:58–70.

[52] Montani F, Rosso OA (2014) Entropy-Complexity Characterization of Brain Development in Chickens. Entropy 16: 4677–4692.

[53] Montani F, Rosso OA, Matias F, Bressler SL, Mirasso CR (2015) A symbolic information approach to determine anticipated and delayed synchronization in neuronal circuit models. Phil. Trans. R. Soc. A 373: 20150110.

[54] Montani F, Baravalle R, Montangie L, Rosso OA (2015) Causal information quantification of prominent dynamical features of biological neurons. Phil. Trans. R. Soc. A 373: 20150109.

[55] Bariviera A, Guercio MB, Martinez LB, Rosso OA (2015) The (in)visible hand in the Libor market: an Information Theory approach. Eur. Phys. J. B 88: 208.

[56] Bariviera A, Guercio MB, Martinez LB, Rosso OA (2015) A permutation Information Theory tour through different interest rate maturities: the LIBOR case. Phil. Trans. R. Soc. A 373: 20150119.

[57] Bandt C (2005) Ordinal time series analysis. Ecol. Modell. 182: 229–238.

[58] Aquino ALL, Cavalcante TSG, Almeida ES, Frery A, Rosso OA (2015) Characterization of vehicle behavior with information theory. Eur. Phys. J. B 85: 257.

[59] De Micco L, González CM, Larrondo HA, Martín MT, Plastino A, Rosso OA (2008) Randomizing nonlinear maps via symbolic dynamics. Physica A 387:3373–3383

[60] De Micco L, Larrondo HA, Plastino A, Rosso OA (2009) Quantifiers for randomness of chaotic pseudo-random number generators. Phil. Trans. R. Soc. A 367: 3281–3296.

[61] Ribeiro HV, Zunino L, Lenzi EK, Santoro PA, Mendes RS (2012) Complexity-Entropy Causality Plane as a Complexity Measure for Two-Dimensional Patterns. Plos One 7: e40689.

## Support Vector Machines and One-Class Support Vector Machines

Support Vector Machines (SVMs) were introduced by Vapnik and co-workers [1, 2, 3, 4], and extended by a number of other researchers. Their remarkably robust performance with respect to sparse and noisy data makes them the choice in several applications. A SVM is primarily a method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. SVMs perform both regression and classification tasks and can handle multiple continuous and categorical variables. To construct an optimal hyperplane, a SVM employs an iterative training algorithm, which is used to minimize an error function.

One-Class Support Vector Machines (OC-SVMs) are a natural extension of SVMs [5, 6]. In order to identify suspicious observations, an OC-SVM estimates a distribution that encompasses most of the observations, and then labels as "suspicious" those that lie far from it with respect to a suitable metric. An OC-SVM solution is built estimating a probability distribution function which makes most of the observed data more likely than the rest, and a decision rule that separates these observation by the largest possible margin. The computational complexity of the learning phase is intensive because the training of an OC-SVM involves a quadratic programming problem [2], but once the decision function is determined, it can be used to predict the class label of new test data effortlessly.

In our case, the observations are six-dimensional vectors: Entropy, Complexity and Fisher Information in each of the two directions, horizontal and vertical, and we train the OC-SVM with genuine signatures. Let $\mathcal{Z} = \{z_1, z_2, \ldots, z_N\}$ be the six-dimensional training examples of genuine signatures. Let $\Phi\colon \mathcal{Z} \to \mathcal{G}$ be a kernel map which transforms the training examples to another space. Then, to separate the data set from the origin, one needs to solve the following quadratic programming problem:

$$\min_{\mathbf{w}\in\mathcal{G},\xi_i,b\in\mathbb{R}} \qquad \left\{\frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - b\right\} \tag{S2.1}$$

subject to

$$\nu \in (0,1],\ \xi_i \geq 0,\ \forall i = 1,\ldots,N,\ \text{and} \tag{S2.2}$$

$$(\mathbf{w}\cdot\Phi(z_i)) \geq b - \xi_i,\ \forall i = 1,\ldots,N, \tag{S2.3}$$

where $\xi_i$ are nonzero slack variables which allow the procedure to incur in errors. The parameter $\nu$ characterizes the solution as *a)* it sets an upper bound on the fraction of outliers (training examples regarded out-of-class) and, *b)* it is a lower bound on the number of training examples used as Support Vectors. We used $\nu = 0.1$ in our proposal.

Using Lagrange techniques and a kernel function $K(z, z_i) = \Phi(z)^T\Phi(z_i)$, for the dot-product calculations, the decision function $f(z)$ becomes:

$$f(z) = \text{sign}\left\{(\mathbf{w}\cdot\Phi(z)) - b\right\} = \text{sign}\left\{\sum_{i=1}^{N}\ \alpha_i\ K(z, z_i) - b\right\}. \tag{S2.4}$$

This method thus creates a hyperplane characterized by $\mathbf{w}$ and $b$ which has maximal distance from the origin in the feature space $\mathcal{G}$ and separates all the data points from the origin. Here $\alpha_i$ are the Lagrange multipliers; every $\alpha_i > 0$ is weighted in the decision function and thus "supports" the machine; hence the name Support Vector Machine. Since SVMs are considered to be sparse, there will be relatively few Lagrange multipliers with a nonzero value.

Our choice for the kernel is the Gaussian Radial Base function:

$$K(z_i, z_j) = \exp\Big(-\frac{1}{2\sigma^2}\|z_i - z_j\|^2\Big), \tag{S2.5}$$

where $\sigma \in \mathbb{R}$ is a kernel parameter and $\|z_i - z_j\|^2$ is the dissimilarity measure; we used Euclidean distance.

The parameter $\sigma^2 = 10$ was selected by 5-fold-cross validation, that its, the dataset is divided into five disjoint subsets, and the method is repeated five times. Each time, one of the subsets is used as the test set and the other four subsets are put together to form the training set. Then the average error across all trials is computed. Every observation belongs to a test set exactly once, and belongs to a training set four times. Accuracy (ACC), Area Under the ROC Curve (AUC) and Equal Error Rate (EER) are used as performance measures [7].

In the context of signature verification one-class classification problems, a false positive occurs when a genuine signature is erroneously classified as being atypical. The probability of false positive misclassification is the false positive rate, which is controlled by the parameters $\nu$ in the aforementioned OC-SVM formulation. The parameter $\nu$ can be fixed a *priori* and it corresponds to the percentage of observations of the typical data which will be assigned as the Type I Error.

The `R` interface to `libsvm` in package `e1071` is designed to be as intuitive as possible. In the following we generate a toy dataset in $\mathbb{R}^2$, and show how to train and test an SVM classifier.

```
n <- 150 # number of data points
p <- 2 # dimension
sigma <- 1 # variance of the distribution
meanpos <- 0 # centre of the distribution of true signatures
meanneg <- 3 # centre of the distribution of false signatures
npos <- round(n/2) # number of true signatures
nneg <- n-npos # number of false signatures
# Generate the true and false signatures
xpos <- matrix(rnorm(npos*p, mean=meanpos, sd=sigma),npos,p)
xneg <- matrix(rnorm(nneg*p, mean=meanneg, sd=sigma),npos,p)
x <- rbind(xpos, xneg)
# Generate the labels of signatures
y <- matrix(c(rep(1, npos),rep(-1, nneg)))
```

Now we split the data into a training set (80%) and a test set (20%)

```
## Prepare a training and a test set ##
ntrain <- round(n*0.8) # number of training examples
tindex <- sample(n,ntrain) # indices of training samples
xtrain <- x[tindex,]; xtest <- x[-tindex,]
ytrain <- y[tindex];  ytest <- y[-tindex]
istrain=rep(0,n); istrain[tindex]=1
```

Training a SVM (one-class) using the radial basis function kernel with fixed hyper-parameters $\nu = 0.1$ and $\sigma = 0.05$:

```
library(e1071)# Functions for support vector machines
library(rpart) # Recursive Partitioning and Regression Trees
```

```
svm.model <- svm(Type ~ ., data = trainset, type='one-classification',
nu=0.10, scale=TRUE, kernel="radial", gamma=0.05)
```

Prediction of class signatures

```
svm.pred <- predict(svm.model, testset[,-10])
```

For more details, see the Reproducible research material available at `http://www.de.ufpe.br/~raydonal/ReproducibleResearch/Signatures/Signatures.html`

## References


[1] Campbell C, Ying Y. Learning with Support Vector Machines. In: Brachman RJ, Dietterich T, editors. Synthesis Lectures on Artificial Intelligence and Machine Learning. No. 5 in Synthesis Lectures on Artificial Intelligence and Machine Learning. Santa Fe, CA: Morgan and Claypool; 2011. p. 1–95.

[2] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. Pittsburgh: ACM Press; 1992. p. 144–152.

[3] Vapnik VN. The Nature of Statistical Learning Theory. Springer; 1995.

[4] Vapnik VN. Statistical Learning Theory. Willey; 1998.

[5] Schölkopf B, Platt JC, Shawe-Taylor JC, Smola AJ, Williamson RC. Estimating the Support of a High-Dimensional Distribution. Neural Computation. 2001; 13(7): 1443–1471.

[6] Schölkopf B, Smola AJ. Learning with Kernels. Cambridge: MIT Press; 2002.

[7] Provost F, Kohavi R. On Applied Research in Machine Learning. Machine Learning. 1998; 30(2/3): 127–132.

## Exploratory data analysis

Figs A, B, C, and D present the supplementary exploratory data analysis of those features that were not presented in the body of the article.

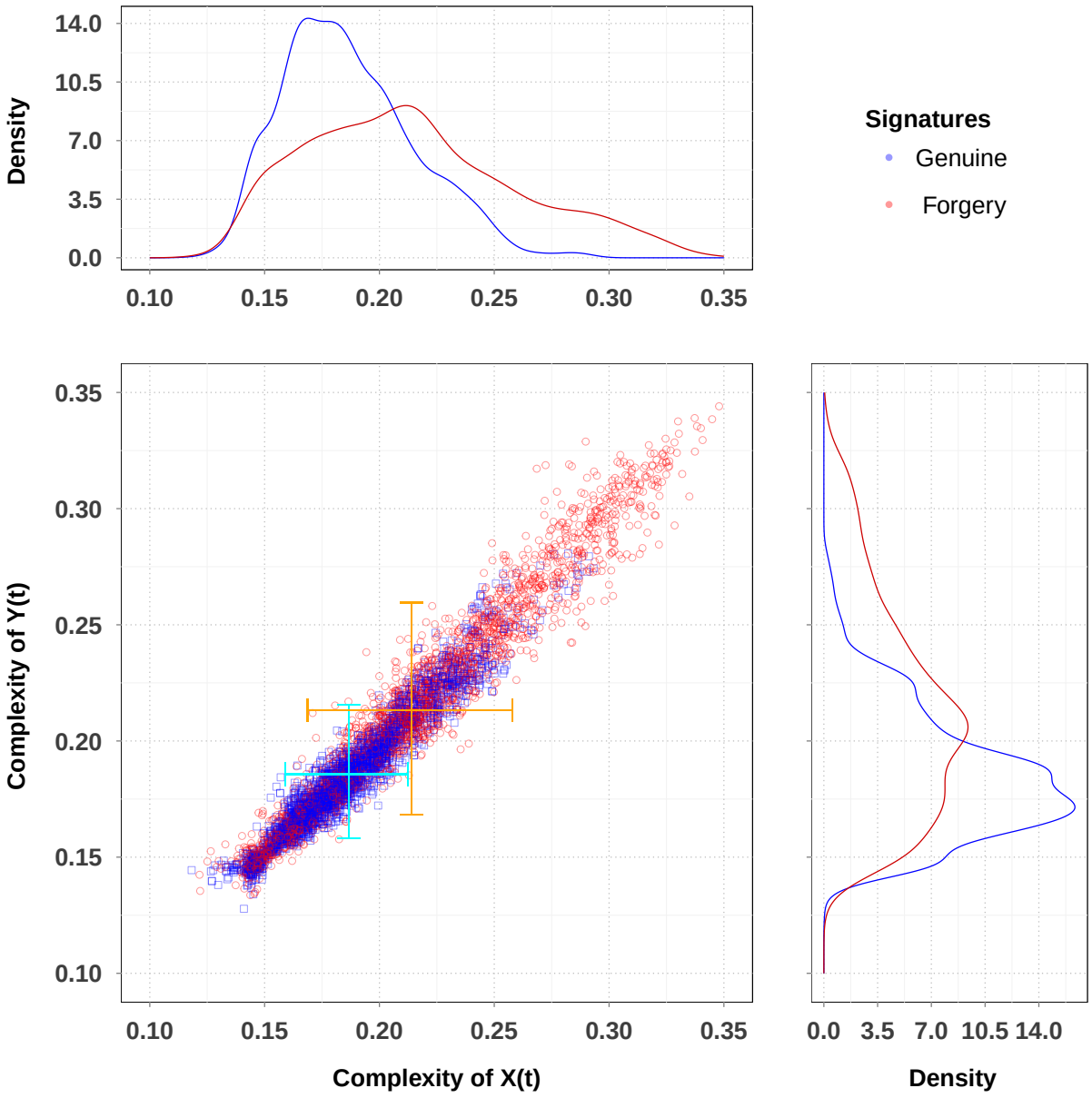


Figure A: Scatter plot with marginal kernel density estimates of complexity quantifiers in both trajectory coordinates time series **X** and **Y**. Genuine (blue) and skilled forgery signatures (red points), 100 subjects. Marginal kernel densities depict the distribution of statistical complexity along both axes.

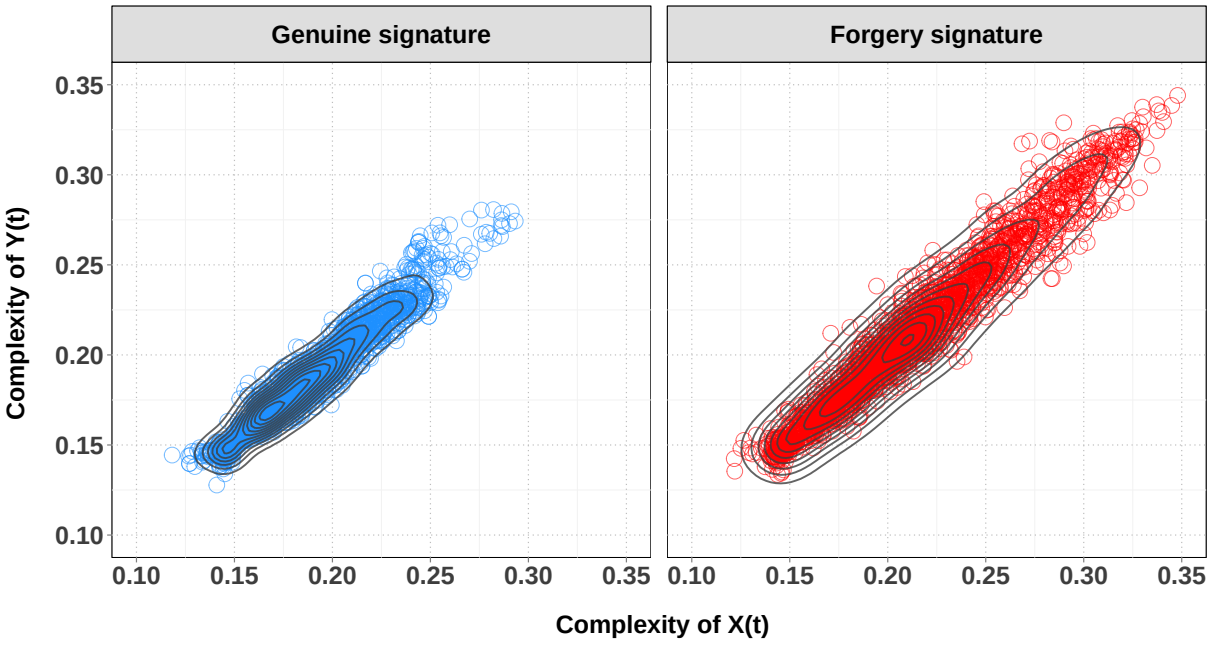


Figure B: Contour plot superimposed on the scatterplot of statistical complexity for genuine (right panel) and skilled forgery signatures (left panel).

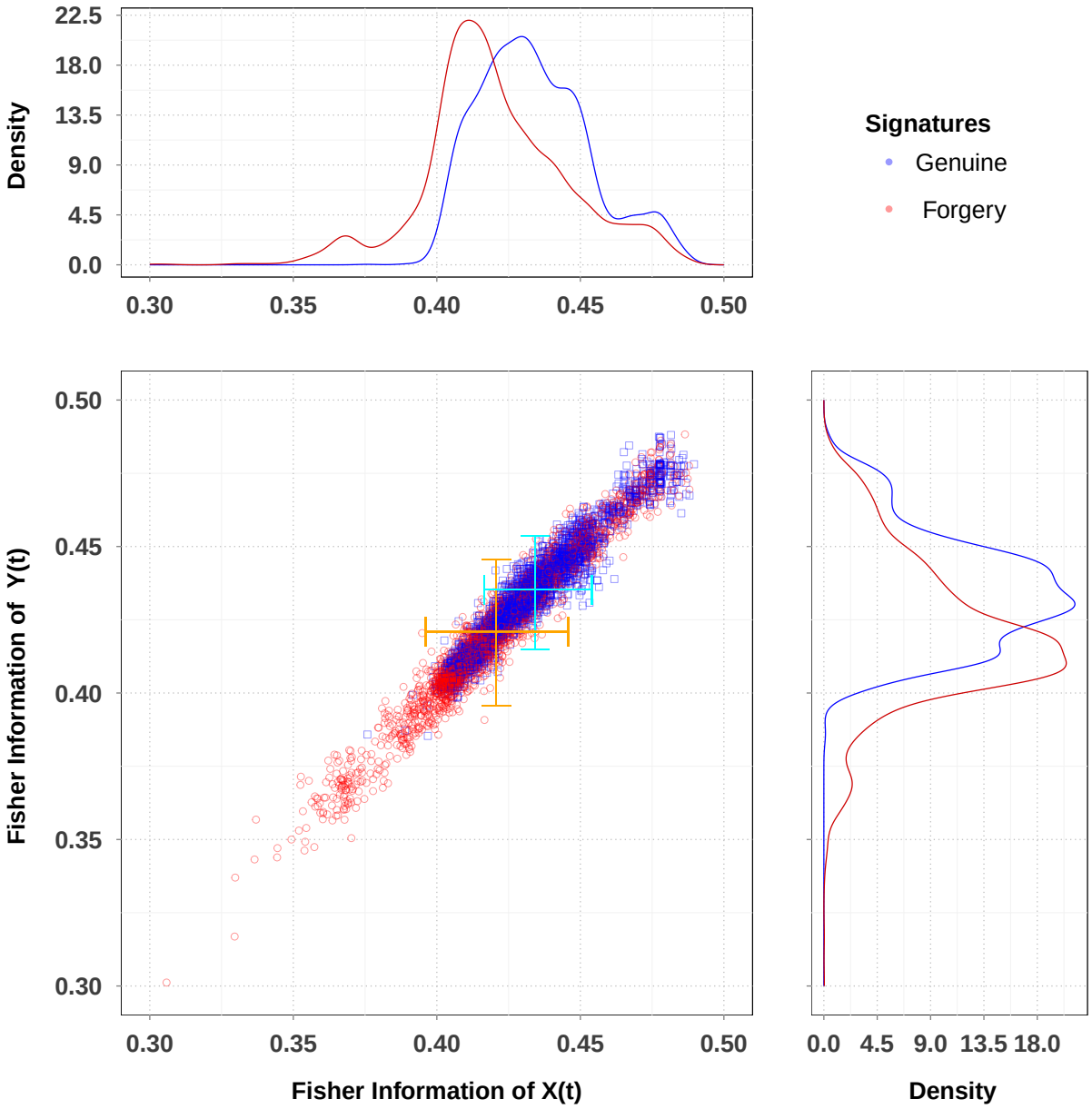


Figure C: Scatter plot with marginal kernel density estimates of Fisher information quantifiers in both trajectory coordinates time series $\mathbf{X}$ and $\mathbf{Y}$. Genuine (blue) and skilled forgery signatures (red points), 100 subjects. Marginal kernel densities depict the distribution of Fisher information along both axes.

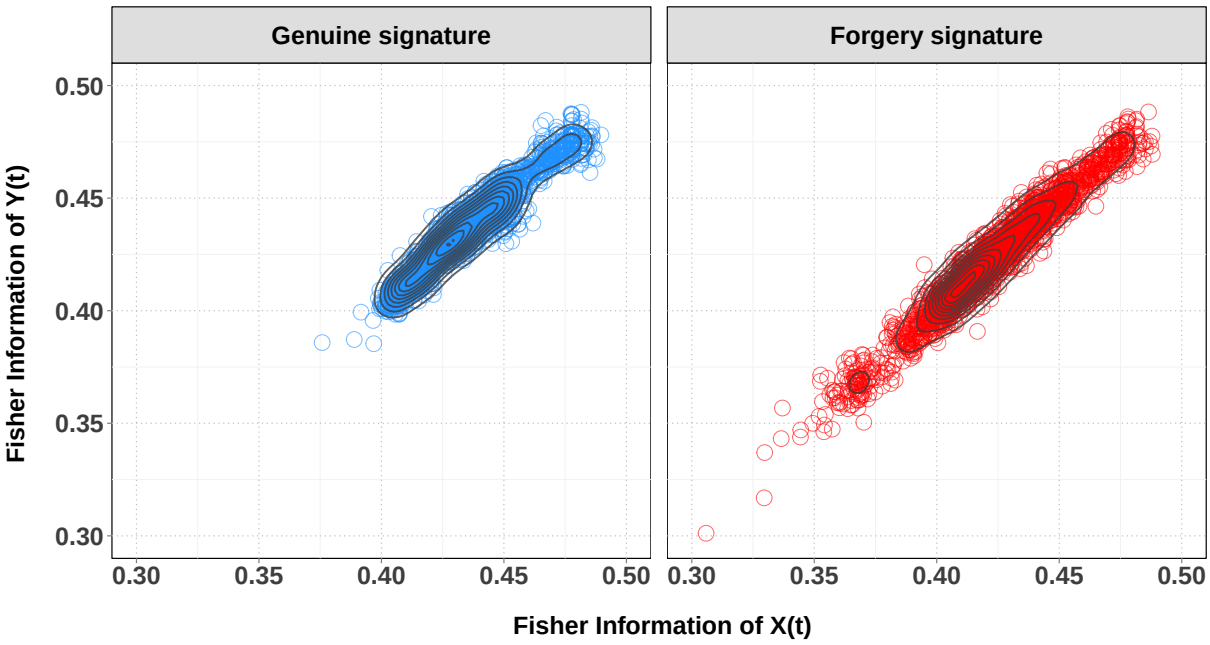


Figure D: Contour plot superimposed on the scatterplot of Fisher information quantifiers for genuine (right panel) and skilled forgery signatures (left panel).

## Signature classification

Figs A and B present the circular dendrograms of the two additional features: statistical complexity and Fisher information.

Figs C and D show the classification by the rule of the parallelepiped of genuine signatures using statistical complexity and Fisher information obtained as result of the application of entropy classification.


Figure A: Neighbor-joining, rooted, circular dendrogram clustering of genuine signatures by statistical complexity: H1, H2, and H3, in red, blue, and green, respectively.

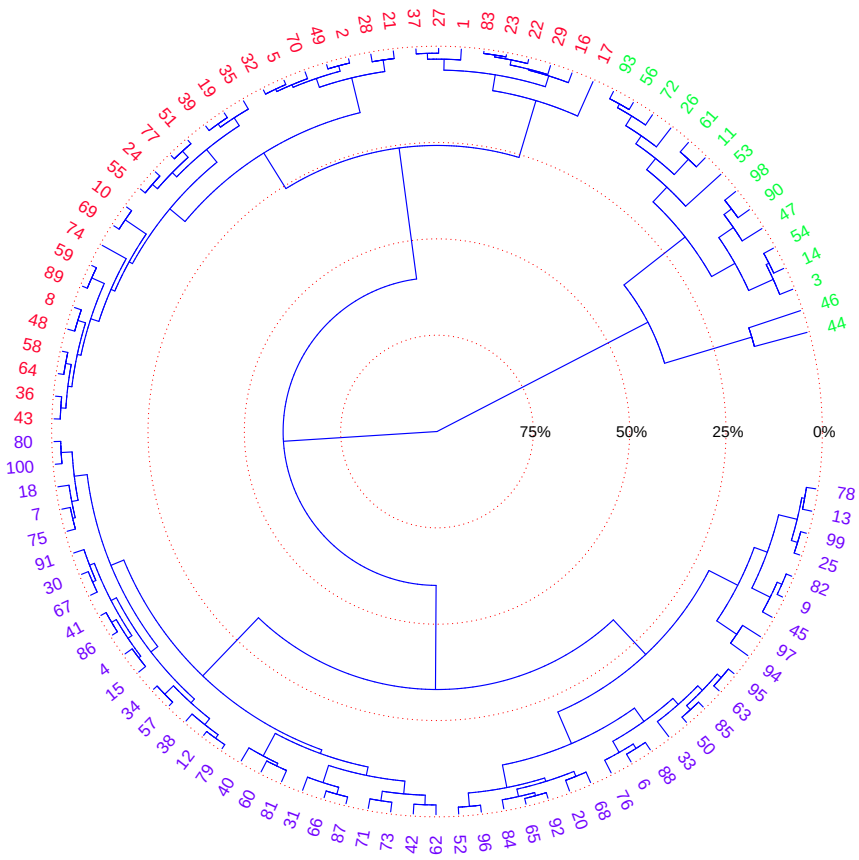


Figure B: Neighbor-joining, rooted, circular dendrogram clustering of genuine signatures by Fisher information: H1, H2, and H3, in red, blue, and green, respectively.

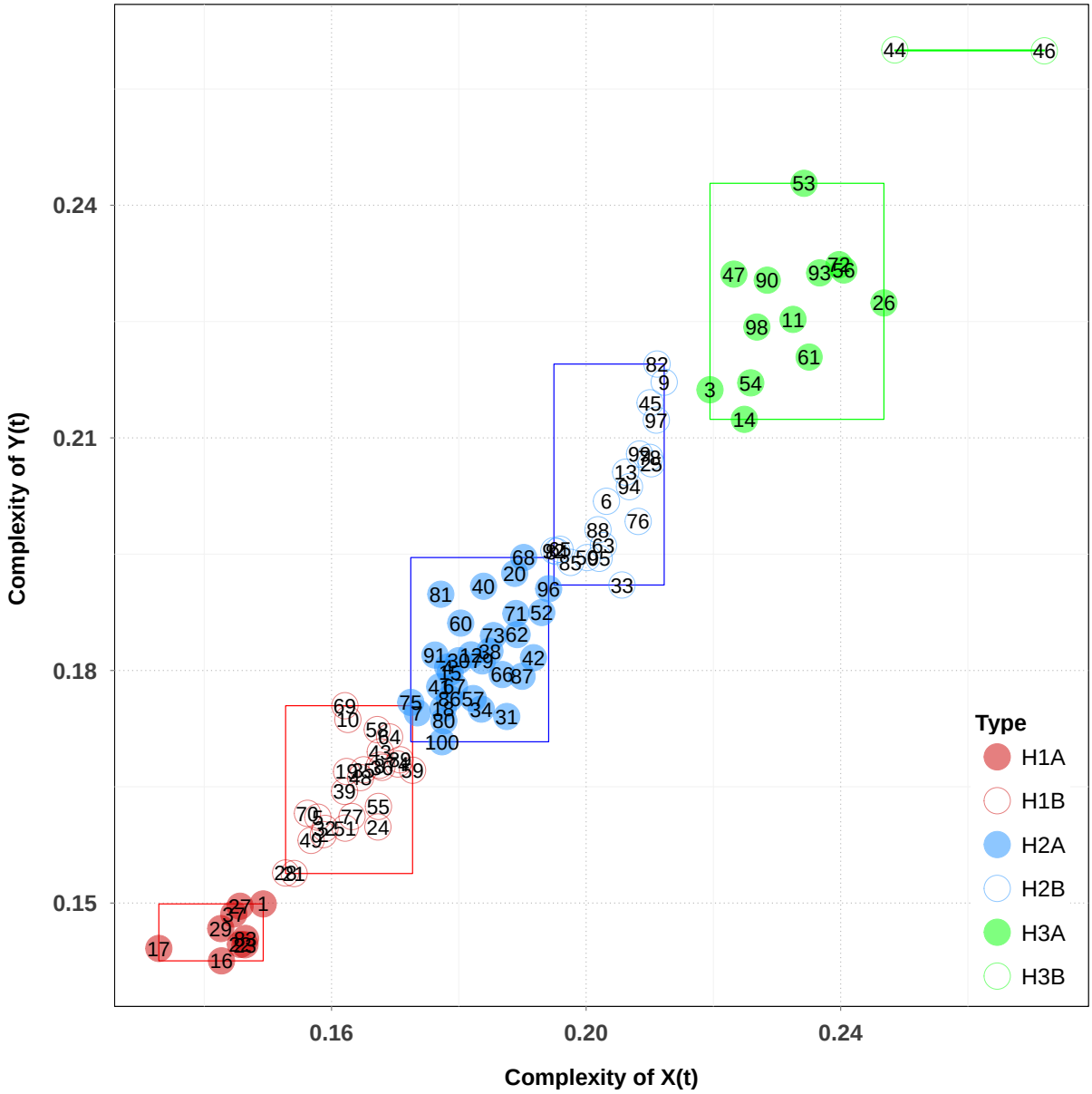


Figure C: Classification by the rule of the parallelepiped of genuine signatures using statistical complexity obtained as result of the application of entropy classification. Each subject is identified by its ID.

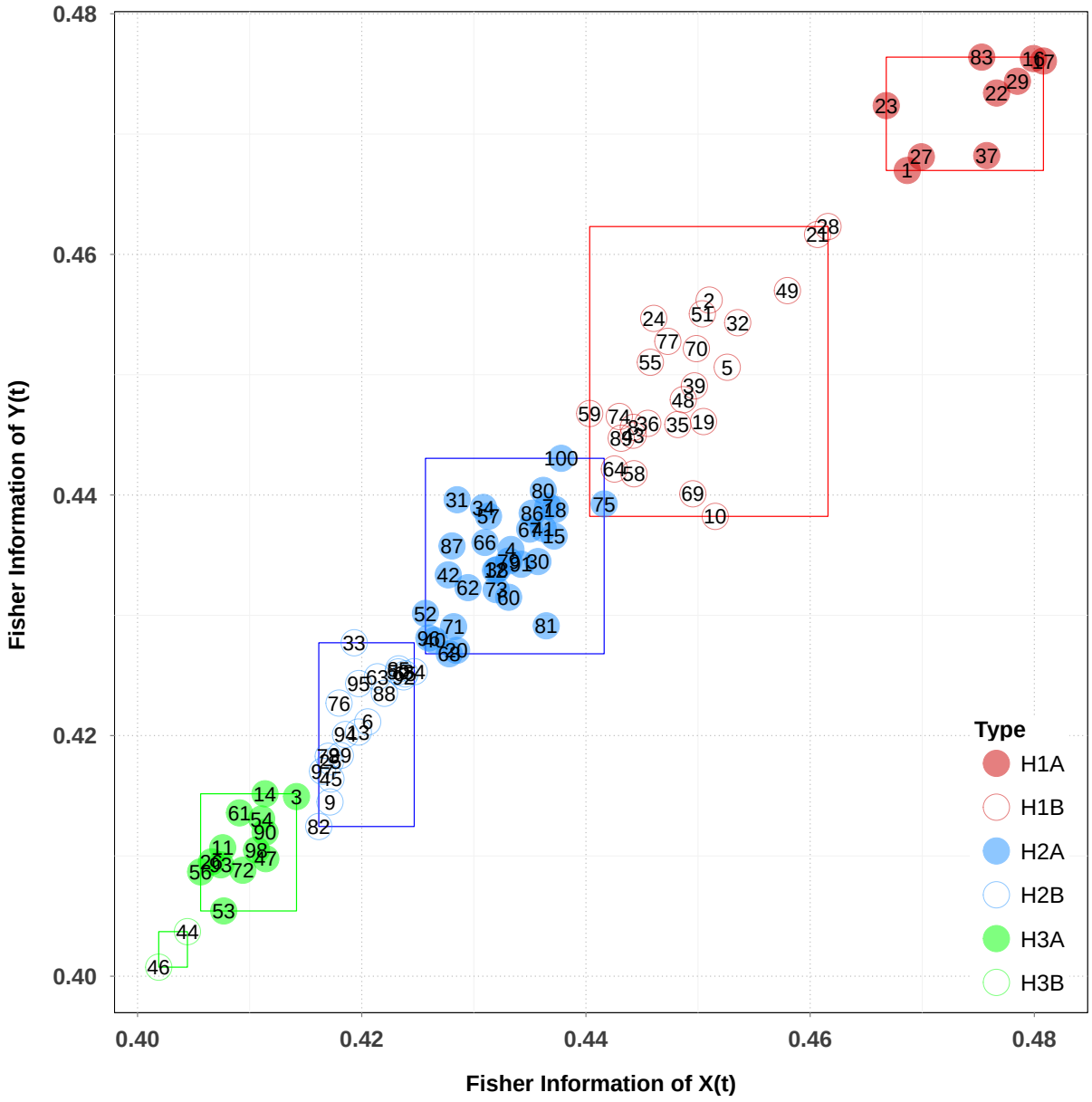


Figure D: Classification by the rule of the parallelepiped of genuine signatures using Fisher information obtained as result of the application of entropy classification. Each subject is identified by its ID.