Ordering Principal Components of Multivariate Fractional Brownian Motion for Solving Inverse Problems

Marisa Mohr*† and Ralf Möller*
* University of Luebeck, Institute of Information Systems, Luebeck, Germany
† inovex GmbH, Hamburg, Germany
E-mail: {mohr,moeller}@ifis.uni-luebeck.de

***Abstract*—The problem of recovering information about a generating time-dependent dynamical system from measurement data is investigated as an inverse problem. Fractional Brownian motion is often used for modelling real-world applications that deal with specific properties such as long-range dependence or self-similarity. The self-similarity parameter, also known as Hurst parameter H, is directly related to the distribution of ordinal patterns in fractional Brownian motion. Thus the corresponding information entropy measure, known as permutation entropy, can extract significant information about the generating system. As real-world applications often involve multivariate correlated measurements, multivariate variants of permutation entropy have to be considered. While pooled permutation entropy is able to estimate H, it fails to detect cross-correlations between variables. In this work, we use multivariate permutation entropy based on principal component analysis (MPE-PCA) to investigate self-similarity and cross-correlations. We examine the variation of permutation entropy of principal components with variations of H and cross-correlations and find those principal components behave equally to their origin under certain conditions, i.e., as H increases, the MPE-PCA of the principal components decreases. Furthermore, MPE-PCA discovers cross-correlations in multivariate fractional Brownian motion.**

***Index Terms*—Fractional Brownian Motion, Ordinal Pattern, Permutation Entropy, Principal Component Analysis**

I. Introduction

Time series data is part of many real-world applications, e.g., weather records, patient health evolution metrics, (industrial) sensor data, stock prices, website clicks, server metrics, or network data – just to name a few. For many real-world applications, modelling of time-dependent dynamical systems requires specific properties such as *long-range dependence* or *self-similarity*. For example, long-term memory is fundamental in internet traffic [1] or for financial data [2]. If return series exhibit a long-range dependence, this indicates that observed returns over time are not independent. When returns are not independent, past returns may help predict future returns [3]. Time-dependent dynamical systems with long-range dependence or self-similarity are commonly described by *fractional Brownian motion* (fBm), which has foundations based on a solid theory and proved to be successfully applied in many real-world challenges, e.g., proven in [4], [5], [6].

Solving the inverse problem of fBm means to infer from observable measurements the parameters of a generating dynamical system, e.g., the self-similarity of fBm, which is generally expressed by the so-called Hurst parameter H [7]. Over the years, several heuristic methods, such as re-scaled (R/S) range analysis [8] and detrended fluctuation analysis (DFA) [9] or classical statistical inversion methods, [7] have been proposed for this purpose. A more recently discussed approach from the field of machine learning is the application of principal component analysis for Hurst parameter estimation [10]. In addition, efficient mappings from observations to a set of scalar-valued features that capture specific properties can provide information about the underlying system. Specifically, entropies, a measure from information theory, are promising through an encoding that preserves information content [11]. Permutation entropy (PE) is a robust, scalar-valued measure that determines the degree of complexity of time series by analysing the distribution of ordinal patterns [12]. PE on fBm is well understood: The distribution of ordinal patterns of specific lengths yield interesting parameter functions, e.g., they are directly related to the Hurst parameter H and therefore helpful to solve inverse problems [13], [14], [15].

Nevertheless, in many fields of applications, multivariate measurements are performed. Examples for multivariate fractional Brownian motion (mfBm) can be found in economic time series [16], or functional Magnetic Resonance Imaging of several brain regions [17]. Pooled permutation entropy (PPE), a multivariate extension of PE, is also suitable to study the self-similarity of in the multivariate case, but it fails to distinguish mfBms with the same Hurst parameters but different cross-correlations between variables [18], [15]. In [19] we introduce *multivariate permutation entropy based on principal component analysis* (MPE-PCA), an alternative multivariate extension of PE that includes correlations of variables and show its relevance and efficiency on various multivariate real-world time series data sets in classification. In this work, we generalise MPE-PCA and apply it to mfBm, examining the variation of MPE-PCA of pricipal components (PCs) with variations of H and cross-correlations. We show that PCs of fBm behave like their origin under certain conditions, i.e., as H increases, MPE-PCA decreases. Furthermore, unlike PPE, MPE-PCA detects when large cross-correlations at high Hurst parameter value H cause the behaviour of all variables to converge and is thus suitable for discovering cross-correlations.

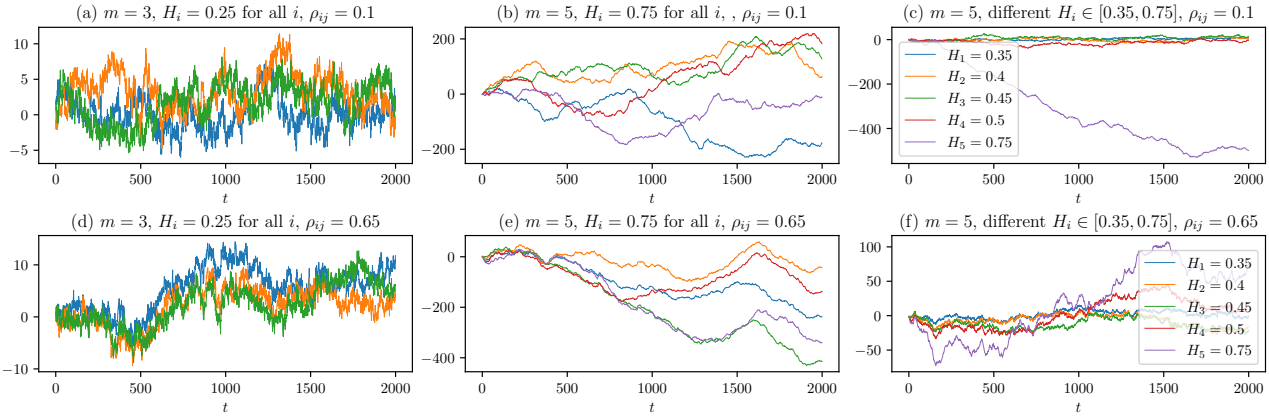

Fig. 1. Six realisations of mfBm of length $T = 2000$ with different variable dimensions, different Hurst parameters, and $\eta_{ij} = 0.1/(1 - H_i - H_j)$, where (a)-(c) have a low correlation coefficient $\rho_{ij} = 0.1$ and (d)-(e) have a high correlation coefficient $\rho_{ij} = 0.65$.

II. Preliminaries

We shortly formalise multivariate fractional Brownian motion (mfBm) and subsequently key concepts of ordinal pattern representations that are the basis of permutation entropy (PE).

A. Multivariate Fractional Brownian Motion

In this work, we focus on a class of special stochastic processes. A stochastic process or more generally a mathematical object that is similar to itself at all scales is called a fractal. When you zoom in on a fractal, it resembles or looks exactly like the original shape. The mathematical property is called self-similarity and is expressed by the so-called Hurst parameter H. The fractional Brownian motion (fBm) is the unique mean-zero Gaussian process, which is zero at origin and has stationary and self-similar increments. In case $H = 1/2$, fBm corresponds to the ordinary Brownian motion. In case $H > 1/2$, the process has a persistence property and positively correlated increments, i.e., an upward jump is more likely followed by another upward jump and vice versa, and the process exhibits long-range dependence. For $H \to 1$, the process becomes smoother, less irregular and more trendy. In case $H < 1/2$, the process has negatively correlated increments and an anti-persistence property.

Definition 1 ([20])**.** An m-multivariate process $((X^i(t))_{i=1}^m)_{t\in\mathbb{R}}$ is called *multivariate fractional Brownian motion (mfBm)* with Hurst parameter $H = (H_1, \ldots, H_m)$, $H_i \in (0, 1)$ for $i = 1, ..., m$, and denoted as $\mathbf{B}_H^m(t)$, if it is

1) Gaussian distributed,
2) self-similar with Hurst parameter H, i.e., there exists a vector $H = (H_1, \ldots, H_m)$ where $H_i \in (0, 1)$ for $i = 1, \ldots, m$ such that for any $a > 0$ it holds

$$(X^1(at), ..., X^m(at))_t \sim (a^{H_1}X^1(t), ..., a^{H_m}X^m(t))_t,$$

where $\sim$ denotes the equality of finite-dimensional distributions, and it has
3) stationary increments, i.e., $X^i(t) - X^i(s) \sim X^i(t - s)$.

Figure 1 shows six mfBms with different Hurst parameters. Stationarity refers to the fact that the distribution of the process does not change in time, which has significant consequences. In particular, the $B^i_{H_i}(t)$ are identically distributed, i.e., the expectation values and variances of components do not depend on time t. Furthermore, the distribution of $(B^i_{H_i}(t), B^i_{H_i}(s))$ depends only on $t - s$, so the correlations of the components also depend only on $t - s$. Therefore, mfBm can be characterised by its covariances and cross-covariances of its variables, i.e., by the parameters $\sigma_i > 0$, $\rho_{ij} \in (-1, 1)$ and $\eta_{ij} \in \mathbb{R}$ for $i, j = 1, ..., m$, which allow two components to be more or less correlated and the process to be reversible in time or not. Parameter $\sigma_i > 0$ is the standard deviation of the i-th variable at time $t = 1$. Parameter $\rho_{ij} = \rho_{ji}$ is the correlation coefficient between the variables i and j at time $t = 1$. Parameters $\eta_{ij} = -\eta_{ji}$ are antisymmetric and linked with the time-reversibility of mfBm.

Lemma 1 (Covariance Function of mfBm [21])**.** The mfBm $\mathbf{B}_H^m(t)$ marginally is an fBm $B^i_{H_i}(t)$, such that the covariance function of the i-th variable $B^i_{H_i}$ of mfBm is

$$Cov(B^i_{H_i}(s), B^i_{H_i}(t)) = \frac{\sigma_i^2}{2}(|s|^{2H_i} + |t|^{2H_i} - |t-s|^{2H_i}), \quad (1)$$

where $\sigma_i^2 = Var(B^i_{H_i}(1))$. The cross-covariances of mfBm for all $(i, j) \in \{1, ..., m\}^2$ and $i \neq j$ are given by

$$Cov(B^i_{H_i}(s), B^j_{H_j}(t)) = \frac{\sigma_i\sigma_j}{2}(w_{ij}(-s) + w_{ij}(t) - w_{ij}(t-s)) \quad (2)$$

where the function $w_{ij}(h)$ is defined as

$$w_{ij}(h) = \begin{cases} (\rho_{ij} - \eta_{ij}\,\mathrm{sign}(h))|h|^{H_i+H_j} & \text{if } H_i + H_j \neq 1, \\ \rho_{ij}|h| + \eta_{ij}h\log|h| & \text{if } H_i + H_j = 1. \end{cases} \quad (3)$$

Moreover, a setting of $\rho_{ij} = 1$ and $\eta_{ij} = 0$ in Eqs. (2) or (3) is matching with the definition in the univariate case, that is (1). For $m = 1$, Definition 1 matches the definition in the univariate case.

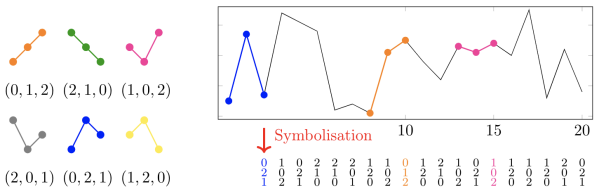

Fig. 2. All possible ordinal patterns of order $d = 3$ (left) and ordinal pattern determination of order $d = 3$ and time delay $\tau = 1$ for a univariate time series at any time point $t \in [\tau(d-1), T]$ (right).

B. Ordinal Pattern Representations

In this section, we introduce ordinal pattern symbolisation and the concept of permutation entropy based on them, which we use to investigate the qualitative behaviour of mfBms.

Ordinal Pattern Symbolisation. To calculate the entropy of a time series, the sequence of real-valued measured values must be encoded into a sequence of symbols. As far as current research is concerned, there are two general approaches for time series symbolisation. On the one hand, classical symbolisation approaches use threshold values and data range partitioning for symbol assignment, such as the well-known Symbolic Aggregate approXimation (SAX) representation introduced by Chiu et al. [22]. On the other hand, ordinal pattern symbolisation, an approach based on the idea of Bandt and Pompe [12], uses the total order between two or more neighbours encoded by its permutations for symbolisation. Benefits of the ordinal pattern approach are discussed, for example, in [19], [23], [24], [25]. The formalism is introduced as follows.

Definition 2. A vector $(x_1, ..., x_d) \in \mathbb{R}^d$ has *ordinal pattern* $(r_1, ..., r_d) \in \mathbb{N}^d$ of order $d \in \mathbb{N}$ if $x_{r_1} \geq ... \geq x_{r_d}$ and $r_{l-1} > r_l$ in the case $x_{r_{l-1}} = x_{r_l}$.

Note that equality of two values within a pattern is not allowed. In this case, for example, the newer value is replaced with a smaller value. Fig. 2 (left) shows all possible ordinal patterns of order $d = 3$ of a vector (x_1, x_2, x_3). To symbolise a time series $(x_1, x_2, ..., x_T) \in \mathbb{R}^T$ each point in time $t \in \{d, ..., T\}$ is assigned its ordinal pattern of order d. The order d is chosen to be much smaller than the total length T of the time series to look at smaller windows within the series. In order to assess the overarching trend, the delayed behaviour is of interest. The time delay $\tau \in \mathbb{N}$ is the delay between successive points in the symbol sequences. Different delays show different details of the structure of a time series. Fig. 2 (right) visualises the ordinal pattern determination of order $d = 3$ and time delay $\tau = 1$ in a time series at any time point $t \in [\tau(d-1), T]$.

Ordinal Pattern Distributions. Not the ordinal patterns themselves, but their distributions in different parts of a *univariate time series* $(x_t)_{t=1}^T$ are of interest. Thus, each ordinal pattern is identified with exactly one of the symbols $j = 1, 2, ..., d!$. Then, permutation entropy is defined as (Shannon) entropy of the distribution of ordinal pattern symbols.

Definition 3 ([12])**.** The *permutation entropy* (PE) of order $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ of a univariate time series $\mathbf{x} = (x_t)_{t=1}^T$, $T \in \mathbb{N}$ is defined as

$$\mathrm{PE}_{d,\tau}(\mathbf{x}) = -\sum_{j=1}^{d!} p_j^{\tau,d} \ln p_j^{\tau,d}, \tag{4}$$

where $p_j^{\tau,d}$ is the frequency of ordinal pattern j in the time series.

For time series with maximum random ordinal pattern symbols (resulting in a uniform pattern distribution), PE is $ln(d!)$. For time series with a regular pattern, e.g., in the case of monotony, PE is equal to zero [11]. As an extension of PE, Morabito et al. [26] introduce multi-scale permutation entropy (MSPE) capturing the complexity of time series on different time scales. In addition, Fadlallah et al. [27] introduce weighted permutation entropy (WPE), taking into account patterns that differ in amplitudes.

Multivariate Permutation Entropy. As in many fields of applications, multivariate measurements are performed, Keller and Lauffer [28] provide a canonical definition of multivariate permutation entropy, called PPE. This approach aims to use marginal frequencies of $d!$ ordinal patterns regarding all m variables as input for entropy calculation. For the determination of PPE, an auxiliary matrix has to be established first:

1) For each variable $i = 1, .., m$ and for each ordinal pattern $j = 1, ..., d!$, count all time steps $s \in [\tau(d-1)+1, T]$, for which the variable-time pair (i, s) has the ordinal pattern j.
2) Divide the counts by $m \cdot \delta$, where $\delta := T - \tau(d-1)$ is the total count of ordinal patterns each variable has.
3) Store the results, i.e., frequencies $p_{ij}^{\tau,d}$ in a so-called *pooling matrix* $P \in (0, 1)^{m \times d!}$, which reflects the distribution of ordinal patterns in the multivariate time series across its m variables.

Definition 4 ([28])**.** The *pooled permutation entropy* (PPE) of a multivariate time series $\mathbf{X} = ((x_t^i)_{i=1}^m)_{t=1}^T$ is defined as the PE of the marginal frequencies $p_{\cdot j}^{\tau,d} = \sum_{i=1}^m p_{ij}^{\tau,d}$ for $j = 1, ..., d!$ describing the distribution of the ordinal pattern is defined by

$$\mathrm{PPE}_{d,\tau}(\mathbf{X}) = -\sum_{j}^{d!} p_{\cdot j}^{\tau,d} \ln p_{\cdot j}^{\tau,d}. \tag{5}$$

For example, PPE is successfully used in analysing electroencephalography (EEG) signals, as cross-channel regularities between spatially distant variables, i.e., on different hemispheres or in different areas, can be extracted by long-range spatial non-linear correlations [28]. As a canonical extension for MSPE, Morabito et al. [26] provide multivariate multi-scale permutation entropy (MMSPE) and for WPE we provide multivariate weighted permutation entropy (MWPE) [15]. Both are based on PPE and are thus to be understood as

extensions (and not alternatives) that investigate the special aspects of scales and amplitudes.

C. Permutation Entropy on Fractional Brownian Motion

Bandt and Shiha [13] significantly contribute to the understanding of the underlying behaviour of PE on fBm. They investigate the distribution of ordinal patterns in fBm of different orders and, if possible, provide closed formulas for calculation of pattern distributions as follows:

1) $d = 2$: The ordinal patterns are equally distributed, more specifically
$$p^\tau_{12} = p^\tau_{21} = 1/2, \tag{6}$$
such that $\mathrm{PE}_{2,\tau}(B_H(t)) = -\ln(1/2)$ for all τ.
2) $d = 3$: The distribution of ordinal patterns for all τ is given by
$$p^\tau_{123} = \frac{1}{\pi}\arcsin 2^{H-1} =: u. \tag{7}$$
For a Gaussian process with stationary increments that includes fBm, it is
$$p^\tau_j = \begin{cases} u & \text{if } j = (123), (321), \\ (1-2u)/4 & \text{otherwise,} \end{cases} \tag{8}$$
for all τ, i.e., $\mathrm{PE}_{3,\tau}(B_H(t))$ is monotonically dependent on the Hurst parameter H.
3) $d = 4$: The distribution of ordinal patterns can also be expressed in a closed formula that can be taken from [13].
4) $d > 5$: There are no closed formulas.

In previous work, we contribute to the understanding of the behaviour in the multivariate case, i.e., PPE on mfBm. Corresponding proofs can be found in [18].

Theorem 1. *For order $d = 2$ and $H \in (0, 1)^m$, it holds*
$$\mathrm{PPE}(\mathbf{B}_H(t)) = -\ln(1/2). \tag{9}$$
for all delays $\tau \in \mathbb{N}$ and and variable dimensions m.

Theorem 2. *For order $d = 3$, $\mathrm{PPE}(\mathbf{B}_H(t))$ is independent of all delays $\tau \in \mathbb{N}$, but monotonically dependent on the number of variables m and the Hurst parameter $H \in (0, 1)^m$.*

III. Related Work

As it is not possible to establish a total order between two time points that are vector-valued, i.e., $x_t \in \mathbb{R}^m$ and $x_{t+1} \in \mathbb{R}^m$, there is no trivial generalisation of the PE algorithm to the multivariate case. While the canonical extensions PPE, MSPE and MWPE measure the complexity of the individual variables before pooling the information, there are numerous studies on additional multivariate variants based on alternative strategies. Multivariate Permutation Entropy (MvPE) introduced by He et al. [29] analysis the complexity of the phase space, i.e., ordinal patterns are established over all variables at a fixed time. Building on a theoretical foundation provided in [30], [31], we provide in a previous work an extension of the univariate ordinal pattern to multidimensionality by matrix assignment as symbols storing both temporal and phase space information. However, in an experimental evaluation, we show that the application of this approach to real data is limited, as the number of possible matrix assignments explodes due to the combinatorial possibilities of ordinal patterns [19].

Mapping high dimensional data into low dimensional representation can simplify the following learning task in terms of data compression (reducing processing time and storage space), visualisation (2D or 3D data are easier to visualise and interpret), or performance improvement (avoiding the curse of dimensionality, under-constrained problems and colinearity to improve the performance of the machine learning model). To reduce the number of variables m to only one dimension, Rayan, Mohammad and Ali [32] propose the application of various distance measures, in particular, Euclidian distance with reference point $(x^i_{t=0})^m_{i=1}$, Manhattan distance with reference point $(x^i_{t=0})^m_{i=1}$, and Euclidian distance with reference point $\mathbf{0}$. Consequently, after the dimensionality reduction, Definition 3 can be used directly for the calculation of PE. To account for correlations, in previous work, we propose to reduce the dimension using principal component analysis and show that the application of the so-called MPE-PCA increases the accuracy of many predictions for classification on different real-world data sets [19]. We exploit the use of correlations in this work to detect cross-correlations between variables of mfBm, which other methods such as PPE fail to do.

IV. Ordering the Principal Components of Fractional Brownian Motion

In this section, we recapitulate *multivariate permutation entropy based on principal component analysis* (MPE-PCA) and generalise it from the original work [19]. We then study the behaviour of MPE-PCA on mfBm with variations of the Hurst parameter H and cross-correlations to understand the qualitative behaviour of mfBms, e.g. for solving inverse problems.

A. Multivariate Permutation Entropy Based on Principal Component Analysis

Given a multidimensional time series $X = ((x^i_t)^m_{i=1})^T_{t=1}$, where m is the dimensionality of the time series. MPE-PCA can be intuitively understood as a transformation of data into a decorrelated representation, where the total variance describes some properties of a time series to be obtained by the transformation. pricipal component analysis (PCA) converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables by an orthogonal transformation. For m-dimensional data X, there are basically m basis vectors that are orthogonal. The variance of data points along each basis vector is the total variance of the data. In particular, applying PCA to data $X \in \mathbb{R}^{m\times T}$ means finding a linear mapping $V \in \mathbb{R}^{m\times m}$ onto a new decorrelated representation
$$Z = V^T X \in \mathbb{R}^{m\times T}, \tag{10}$$
such that the variance of the projected data
$$\mathrm{Var}(Z) = \frac{1}{n-1}\sum_i \|V^T x_i\|^2 \tag{11}$$

Algorithm 1: Computation of PCA

Input: Multivariate Time Series $X^{m\times T}$, where X is centered, i.e., mean zero

Function `PCA(`X`)`**:**
- $\Sigma^{m\times m} \leftarrow$ compute covariance matrix of X
- $V\Lambda V^{-1} \leftarrow$ compute eigendecomposition of Σ with $V = (v_1, \ldots, v_m)$ and $\lambda_1 \leq \lambda_m$
- $Z^{m\times T} \leftarrow$ compute orthogonal transformation `// see Eq. (10)`
- **return** Z

is maximal. Then, to find the direction v_1 of maximum variance, we need to solve

$$\max_{v_1} \frac{1}{n-1}\|v_1^T X\|^2 \qquad \text{s.t. } v_1^T v_1 = 1. \tag{12}$$

Rewriting the objective, we have

$$\frac{1}{n-1}\|v_1^T X\|^2 = \frac{1}{n-1} v_1^T X X^T v_1 = v_1^T \Sigma v_1, \tag{13}$$

where Σ is the covariance matrix of X. Maximisation under constraint $v_1^T v_1 = 1$ means solving the Lagrangian

$$f(v_1) = v_1^T \Sigma v_1 - \lambda_1 (v_1^T v_1 - 1) \tag{14}$$

with its derivative

$$\mathrm{d}f(v_1) = 2(v_1^T \Sigma - \lambda_1 v_1^T)\mathrm{d}v_1. \tag{15}$$

Setting the derivative to zero implies

$$\Sigma v_1 = \lambda_1 v_1. \tag{16}$$

From Eq. (,16) we see that v_1 must be an eigenvector of Σ for the largest eigenvalue. In general, PCA is thus based on eigenvalue analysis, i.e., an eigenvector of the covariance matrix $\mathrm{Cov}(X) = \Sigma$ corresponds to a basis vector. An appropriate algorithm to perform PCA is outlined in Algorithm 1 and can be found in several statistical textbooks, e.g., in [33].

For ordering d vectors $x_{t-(d-1)\tau}, ..., x_{t-\tau}, x_t$ with $x_i \in \mathbb{R}^m$ based on their values, we propose to use the decorrelated i-th principal component $Z^i = (z_{i,t-(d-1)\tau}, ..., z_{i,t-\tau}, z_{i,t})$ with $z_{i,l} \in \mathbb{R}$. The rest of the PE calculation is the same as in Definition 3 or Definition 4.

Definition 5 ([19])**.** The *multivariate permutation entropy based on principal component analysis* (MPE-PCA) of order $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ of a multivariate time series $((x_t^i)_{i=1}^m)_{t=1}^T$, $T \in \mathbb{N}$ is defined as

$$\text{MPE-PCA}_{d,\tau,i}(Z^i) = -\sum_{j=1}^{d!} p_j^{\tau,d} \ln p_j^{\tau,d}, \tag{17}$$

where $p_j^{\tau,d}$ is the frequency of ordinal pattern j in the i-th principal component Z^i.

For completeness, we define a pooled version of all principal components as follows.

Algorithm 2: Computation of MPE-PCA/PPE-PCA

Input: Multivariate Time Series $X^{m\times T}$, Order d, Delay τ, Principal Component(s) (i, r)

$Z \leftarrow$ perform `PCA(`X`)` `// see Alg. (1)`
if $r == i$ **then**
- **return** PE(Z^i) `// see Definition 3`

else
- **Function** `pooling(`Z`,`d`,`τ`)`**:**
  - $P^{m\times d!} \leftarrow$ pooling matrix initialised with zeros
  - **for** *every principal component* $Z^i \in Z$ **do**
    - **for** *every ordinal pattern* $j = 1, \ldots, d!$ **do**
      - $c \leftarrow$ # time steps with pattern j
      - $P_{ij} \leftarrow c$ divided by $\delta \cdot m$
  - **return** P
- **Function** `marginalisation(`P`)`**:**
  - $p^{1\times d!} \leftarrow$ vector initialised with zeros
  - **for** *every column* $j = 1, \ldots, d!$ *in* $P^{m\times d!}$ **do**
    - $p_j \leftarrow$ sum up p_{ij}
  - **return** p
- **return** PE(p) `// see Definition 3`

Definition 6. The *pooled permutation entropy based on principal component analysis* (PPE-PCA) of a multivariate time series $X = ((x_t^i)_{i=1}^m)_{t=1}^T$ is defined as PE of the marginal frequencies $p_{.j}^{\tau,d} = \sum_{i=1}^m p_{ij}^{\tau,d}$ for $j = 1, ..., d!$ of all principal components $Z \in \mathbb{R}^{m\times T}$ and can be calculated as

$$\text{PPE-PCA}_{d,\tau}(Z) = -\sum_{j}^{d!} p_{.j}^{\tau,d} \ln p_{.j}^{\tau,d}. \tag{18}$$

Depending on the application, e.g., for classification, it may be helpful to use only a reduced representation $Z \in \mathbb{R}^{r\times T}$ instead of the whole representation $Z \in \mathbb{R}^{m\times T}$. This is based on the assumption that if the first $r < m$ basis vectors cover a sufficiently large percentage of the total variance, then the new r basis vectors are sufficient for the information content of the data. Keeping only the first r PCs of the data X gives the truncated transformation $Z_r = V_r^T X$, where $V \in \mathbb{R}^{m\times r}$ is a matrix of weights whose columns are the eigenvectors of Σ sorted in descending order of the r highest corresponding eigenvalues and is used for dimensionality reduction. Algorithm 2 for computing MPE-PCA and PPE-PCA can also be found on Github and Python Package Index (PyPI)[1]. In previous work, MPE-PCA corresponds to a one-dimensional projection, i.e., $r = i = 1$. Of course, it depends on the data whether the first principal component also explains "enough variance". In [19] we show that on 15 out of 25 different real multivariate time series data sets from the UEA MTSC archive [34] we achieve an improvement in accuracy in terms of classification with the first principal component alone compared to the standard measures discussed in Section III. For the sake

[1] https://github.com/marisamohr/mpePy, https://pypi.org/project/mpePy

of completeness, we analyse all principal components Z^i for $i = 1, ..., m$ and not only $i = 1$ in the following.

B. Principal Component Analysis Applied to Multivariate Fractional Brownian Motion

Given a sample of mfBm of variable dimension m, i.e., $X = ((x_t^i)_{i=1}^m)_{t=1}^T \in \mathbb{R}^{m\times T}$. Remember that as in Lemma 1 mfBm marginally is an fBm. PCA converts a set of observations of possibly correlated variables into a linear combination of uncorrelated variables. That is, a decomposition is found such that the matrix-vector multiplication $\mathbf{X} \to \mathbf{Z} = \mathbf{V}^T\mathbf{X}$ gives the equations

$$\begin{array}{rcl} Z^1 &=& v_{11}X^1 + \cdots + v_{m1}X^m \\ Z^2 &=& v_{12}X^1 + \cdots + v_{m2}X^m \\ \vdots &=& \vdots \\ Z^m &=& v_{1m}X^1 + \cdots + v_{mm}X^m \end{array}, \tag{19}$$

where X^i is a single fBm of length T. The variance of Z^1 is $\mathrm{Var}(Z^1) = \lambda_1$.

Van Zanten [35] shows that the local almost sure behaviour of a linear combination of independent fBms is equivalent to a multiple of a single fBm.

Corollary 1 ([35])**.** Let $X = \sum_i^m a_k X^k$, where $X^1, \ldots, X^m$ are independent fBms with Hurst parameters $H_1 < \cdots < H_m$ and $a_1, \ldots, a_m \in \mathbb{R}\backslash\{0\}$.

(i) If X is equivalent to a multiple of an fBm on $[0, T]$ for some $T > 0$, then X is equivalent to a_1X^1 and $H_2 - H_1 > 1/4$.

(ii) If $H_2 - H_1 > 1/4$ then X and a_1X^1 are locally equivalent.

Thus, if all marginal fBms X^i of mfBm are independent, and for the Hurst parameters $H_2 - H_1 > 1/4$ holds, then Z^i and $v_{1i}X^1$ from Eq. (19) are locally equivalent. If there are cross-correlations between the variables of mfBm, PCA identifies eigenvectors that decorrelate the data and decorate the behaviour of fBm with a multiple. This results in different representations or principal components for mfBm with the same Hurst parameter H and different cross-correlations ρ_{ij}.

C. Ordinal Pattern Distributions of Principal Components

In general, (permutation) entropy increases with the degree of disorder and is maximum for absolutely random states. A time series with Hurst parameter $H = 1/2$ is called a random walk, while for Hurst parameter $H < 1/2$ or $H > 1/2$ the increments of (m)fBm are negatively or positively correlated, respectively. Thus, if the Hurst parameter varies, the entropy of (m)fBm also varies, i.e., if the Hurst parameter $H < 1$ increases, the information content of (m)fBm increases due to strong positive correlations. In case of high cross-correlation between the variables, the variables adjust their behaviour to each other as can be seen in Figure 1 (low cross-correlation at the top, high cross-correlation at the bottom). For example, as the cross-correlation increases, the steep downward trend of the 5-th variable in Figure 1(c), dominated by the high self-similarity or Hurst parameter $H = 0.75$, changes to a slightly upward trend in Figure 1(f), adjusted by the remaining variables with Hurst parameters $H_i \in [0.35, 0.5]$ for $i = 1, \ldots, 4$.

Hereafter we restrict ourselves to the case $H_2 - H_1 > 1/4$. Since in this paper we study the detection of cross-correlations using MPE-PCA, we consider two cases:

- $\rho_{ij} = 0$: Since performing PCA or decorrelation has no effect, the distributions of ordinal patterns and the behaviour of MPE-PCA and PPE-PCA on the principal components of mfBm are the same as in Section II-C.
- $\rho_{ij} \neq 0$: Since the i-th principal component Z^i and $v_{1i}X^1$ are locally equivalent and each element $v_{ij} \in V$ represents a loading, namely the correlation between the original variable and the principal component, Z^i behaves like an fBm decorated by its loading. As the behaviour of fBm is directly related to the Hurst parameter H, but the distribution of ordinal patterns of order $d = 2$ does not depend on the Hurst parameter H of an fBm, neither do MPE-PCA$_{2,\tau,i}$ for all τ, i and PPE-PCA$_{2,\tau}$. It holds MPE-PCA$_{2,\tau,i} = -\ln(1/2) =$ PPE-PCA$_{2,\tau}$ (see Section II-C 1)). In contrast, PE and PPE of orders $d = 3$ and $d = 4$ are monotonically dependent on the Hurst parameter H (see Section II-C 2) and 3)) and Theorem 2), i.e., the decorrelation and the loadings, respectively, influence MPE-PCA and PPE-PCA. Although there are no closed formulas for $d > 4$ (see Section II-C, 4)), similar behaviour is to be expected as in the previous case [18].

In contrast to MPE-PCA$_i$ and PPE-PCA, the computations of PE$_i$ on the i-th variable and PPE are independent of the Hurst parameter H and on cross-correlations, such that these measures are not able to detect cross-correlations.

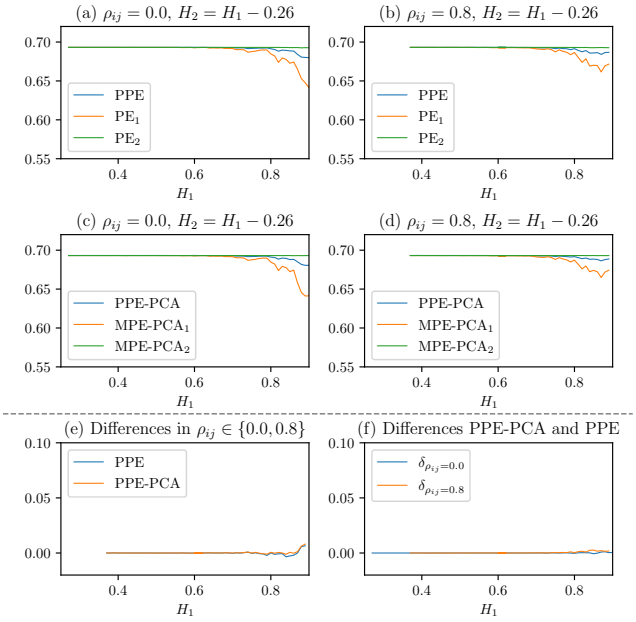

Fig. 3. Comparison of PPE and MPE-PCA of order $d = 2$ on mfBm.

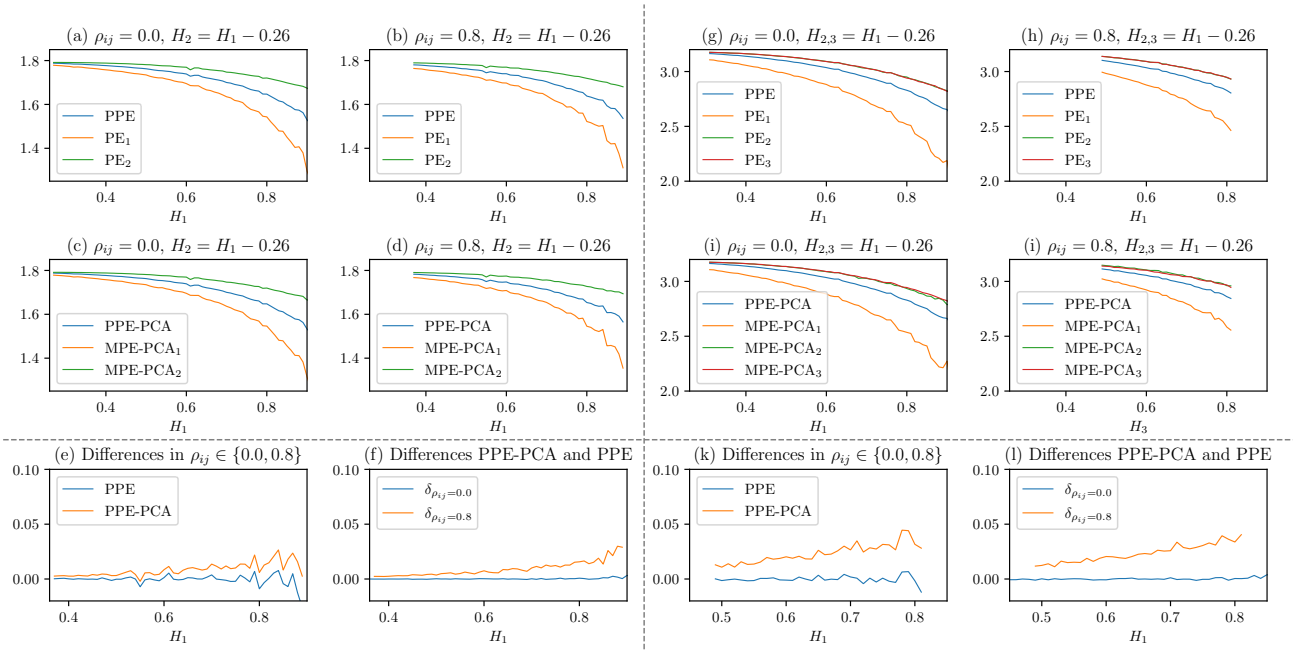

Fig. 4. Comparison of PPE and MPE-PCA of order $d = 3$ (left) and $d = 4$ (right) on mfBm.

V. Empirical Evaluation

In this section, we evaluate the behaviour of MPE-PCA and PPE-PCA on mfBm in an experimental setting underpinning our theoretical findings. The following experimental calculations are based on a simulation of mfBm using Lemma 1 and a corresponding algorithm implemented by Amblard et al. [20]. The length $T = 7500$ of mfBms is assumed to be large. For the simulation of general mfBms, we compare a variation in Hurst parameter H as well as different cross-correlations $\rho_{i,j} = \{0.0, 0.8\}$. In order for Corollary 1 to be satisfied, the relationship $H_2 = H_1 - 0.26$ or $H_{2,3} = H_1 - 0.26$ is chosen for $m = 2$ or $m = 3$ variables, respectively. The time-reversibility parameter $\eta_{i,j}$ is set to $0.1/(1 - H_i - H_j)$. PE$_i$ and MPE-PCA$_i$ are calculated on the single-dimensional i-th variable and i-th principal component, respectively. The visualised values correspond to the mean of 20 simulated experiments.

Order $d = 2$. In Figure 3 we compare PPE (top) and MPE-PCA (middle) of order $d = 2$ and delay $\tau = 1$ on mfBms with $m = 2$ variables and different cross-correlations $\rho_{ij} = 0.0$ (left) and $\rho_{ij} = 0.8$ (right). Figure 3(a) shows a constant entropy with value $-\ln(1/2)$ for all variables i, the pooled version PPE, and Hurst parameters H confirming Theorem 1. The fluctuations in Figure 3(a)-(d) for increasing Hurst parameter H result from length restriction $T < \infty$ in the simulation, i.e., a completely equal distribution of the ascending and descending patterns can only be expected when $T \to \infty$ converges [36]. At $T < \infty$ the trend is dominated by high Hurst parameter H and (permutation) entropy decreases. As PE and PPE of order $d = 2$ are independent of the Hurst parameter H as well as the cross-correlation, Figure 3(b), (c) and (d) are equal to Figure 3(a). Figure 3(e) and (f) confirm that neither PPE nor MPE-PCA of order $d = 2$ are able to detect cross-correlations of the variables.

Orders $d = 3$ and $d = 4$. In Figure 4 we compare of PPE (top) and MPE-PCA (middle) of order $d = 3$ (left) and $d = 4$ (right), respectively, and delay $\tau = 1$ on mfBms with $m = 2$ (left) and $m = 3$ (right) variables, respectively, and different cross-correlations $\rho_{ij} = \{0.0, 0.8\}$. The fluctuations with increasing Hurst parameter H can be explained by length restriction in the simulation, as in the case $d = 2$. Figure 4(a)-(d) and (g)-(i) confirm that PE, PPE, MPE-PCA and PPE-PCA depend monotonically on the Hurst parameter H, i.e., entropy decreases for increasing H. With Corollary 1, the principal components are locally equivalent to an fBm decorated by its loading. Figure 3(e)-(l) confirm that unlike PPE and PPE-PCA of order $d = 2$, PPE-PCA of orders $d = 3$ and $d = 4$ are able to detect cross-correlation of variables, since decorrelation of variables using PCA decorates the behaviour of mfBm.

VI. Conclusion and Future Work

In this work, we investigate the behaviour of multivariate permutation entropy based on principal component analysis (MPE-PCA) and pooled permutation entropy based on principal component analysis (PPE-PCA) on multivariate fractional Brownian motion (mfBm) from a theoretical and experimental point of view. We show that the entropies of the principal components are monotonically dependent on the Hurst parameter H, i.e., entropy decreases as H increases. Thus, MPE-PCA and PPE-PCA are appropriate for solving inverse problems, i.e., given an observed realisation of mfBm, the calculation of MPE-PCA or PPE-PCA provide information about the level

of H and thus parameters of the generating mfBm. Moreover, we show that, unlike PPE, MPE-PCA or PPE-PCA of order $d > 2$ can uncover large cross-correlations at large Hurst parameter H. Since information about the Hurst parameter H and correlations of the variables can be derived single-handed from MPE-PCA or PPE-PCA, this approach offers interesting advantages. In fact, this paper does not fully solve the inverse problem but focuses on the theoretical relationships that motivate the solution of inverse problems. Experiments need to be evaluated in a detailed study to assess performance, especially compared to other point estimators mentioned in the introduction.

The main limitation of PCA is that it projects the data linearly, with many real-world challenges containing complex, non-linear relationships between variables. Since this work examines counts or sums of ordinal patterns, a linear relationship may be reasonable. Nevertheless, it remains to be investigated whether, for example, kernel PCA, independent component analysis or functional PCA analysis can improve the results.

REFERENCES

[1] T. Konstantopoulos and S.-J. Lin, "Fractional brownian approximations of queueing networks," in *Stochastic Networks*, P. Glasserman, K. Sigman, and D. D. Yao, Eds. Springer New York, 1996, pp. 257–273.

[2] R. J. Elliott and J. van der Hoek, "Fractional brownian motion and financial modelling," in *Mathematical Finance*, M. Kohlmann and S. Tang, Eds. Basel: Birkhäuser Basel, 2001, pp. 140–151.

[3] B. S. Nath and B. Mousumi, "Long memory in stock returns: A study of emerging markets," *Iranian Journal of Management Studies*, vol. 5, no. 2, pp. 67–88, Jan. 2012.

[4] S. Arianos and A. Carbone, "Cross-correlation of long-range correlated series," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 03, p. P03037, Mar. 2009.

[5] G. Didier and V. Pipiras, "Integral representations and properties of operator fractional Brownian motions," *Bernoulli*, vol. 17, no. 1, pp. 1–33, Feb. 2011.

[6] C. R. Dietrich and G. N. Newsam, "Fast and Exact Simulation of Stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix," *SIAM Journal on Scientific Computing*, vol. 18, no. 4, pp. 1088–1107, Jul. 1997.

[7] B. D'Ambrogi-Ola, "Inverse Problem for Fractional Brownian Motion with Discrete Data," Apr. 2009. [Online]. Available: https://helda.helsinki.fi/handle/10138/21230

[8] C.-F. Li, "Rescaled-range and power spectrum analyses on well-logging data," *Geophysical Journal International*, vol. 153, no. 1, pp. 201–212, Apr. 2003.

[9] V. A. Setty and A. S. Sharma, "Characterizing Detrended Fluctuation Analysis of multifractional Brownian motion," *Physica A: Statistical Mechanics and its Applications*, vol. 419, pp. 698–706, Feb. 2015.

[10] T. Özkurt, T. Akgül, and S. Baykut, "Principal component analysis of the fractional brownian motion for $0 < h < 0.5$," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 3, 06 2006.

[11] J. Amigó, *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That*, ser. Springer Series in Synergetics. Berlin Heidelberg: Springer-Verlag, 2010.

[12] C. Bandt and B. Pompe, "Permutation Entropy: A Natural Complexity Measure for Time Series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, Apr. 2002, publisher: American Physical Society.

[13] C. Bandt and F. Shiha, "Order Patterns in Time Series," *Journal of Time Series Analysis*, vol. 28, no. 5, pp. 646–665, Sep. 2007.

[14] M. Sinn and K. Keller, "Estimation of ordinal pattern probabilities in gaussian processes with stationary increments," *Computational Statistics and Data Analysis (CSDA)*, vol. 55, no. 4, p. 1781–1790, Apr. 2011.

[15] M. Mohr, F. Wilhelm, and R. Möller, "On the Behaviour of Weighted Permutation Entropy on Fractional Brownian Motion in the Univariate and Multivariate Setting," *The International FLAIRS Conference Proceedings*, vol. 34, Apr. 2021.

[16] L. A. Gil-Alana, "A fractional multivariate long memory model for the US and the Canadian real output," *Economics Letters*, vol. 81, no. 3, pp. 355–359, Dec. 2003.

[17] S. Achard, D. S. Bassett, A. Meyer-Lindenberg, and E. Bullmore, "Fractal connectivity of long-memory networks," *Physical Review E*, vol. 77, no. 3, Mar. 2008.

[18] M. Mohr, N. Finke, and R. Möller, "On the Behaviour of Permutation Entropy on Fractional Brownian Motion in a Multivariate Setting," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2020 (APSIPA-ASC)*, 2020.

[19] M. Mohr, F. Wilhelm, M. Hartwig, R. Möller, and K. Keller, "New Approaches in Ordinal Pattern Representations for Multivariate Time Series," in *Proceedings of the 33rd International Florida Artificial Intelligence Research Society Conference (FLAIRS-33)*, 2020.

[20] P.-O. Amblard, J.-F. Coeurjolly, F. Lavancier, and A. Philippe, "Basic properties of the Multivariate Fractional Brownian Motion," *Séminaires et congrès*, vol. 28, pp. 65–87, 2013.

[21] P.-O. Amblard and J.-F. Coeurjolly, "Identification of the Multivariate Fractional Brownian Motion," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5152–5168, Nov. 2011, conference Name: IEEE Transactions on Signal Processing.

[22] B. Chiu, E. Keogh, and S. Lonardi, "Probabilistic Discovery of Time Series Motifs," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '03. ACM, 2003, pp. 493–498.

[23] I. Stolz and K. Keller, "A General Symbolic Approach to Kolmogorov-Sinai Entropy," *Entropy*, vol. 19, no. 12, p. 675, Dec. 2017.

[24] K. Keller, T. Mangold, I. Stolz, and J. Werner, "Permutation Entropy: New Ideas and Challenges," *Entropy*, vol. 19, no. 3, p. 134, Mar. 2017.

[25] A. B. Piek, I. Stolz, and K. Keller, "Algorithmics, Possibilities and Limits of Ordinal Pattern Based Entropies," *Entropy*, vol. 21, no. 6, p. 547, Jun. 2019.

[26] F. C. Morabito, D. Labate, F. La Foresta, A. Bramanti, G. Morabito, and I. Palamara, "Multivariate Multi-Scale Permutation Entropy for Complexity Analysis of Alzheimer's Disease EEG," *Entropy*, vol. 14, no. 7, pp. 1186–1202, Jul. 2012.

[27] B. Fadlallah, B. Chen, A. Keil, and J. Principe, "Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 87, 02 2013.

[28] K. Keller and H. Lauffer, "Symbolic Analysis of High-Dimensional Time Series," *International Journal of Bifurcation and Chaos*, vol. 13, no. 09, pp. 2657–2668, Sep. 2003.

[29] S. He, K. Sun, and H. Wang, "Multivariate permutation entropy and its application for complexity analysis of chaotic systems," *Physica A: Statistical Mechanics and its Applications*, vol. 461, pp. 812–823, 2016.

[30] K. Keller, "Permutations and the Kolmogorov-Sinai entropy," *Discrete & Continuous Dynamical Systems - A*, vol. 32, no. 3, p. 891, 2012.

[31] A. Antoniouk, K. Keller, and S. Maksymenko, "Kolmogorov-Sinai entropy via separation properties of order-generated sigma-algebras," *Discrete and Continuous Dynamical Systems*, vol. 34, no. 5, pp. 1793–1809, Oct. 2013.

[32] Y. Rayan, Y. Mohammad, and S. A. Ali, "Multidimensional Permutation Entropy for Constrained Motif Discovery," in *Intelligent Information and Database Systems*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing.

[33] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, 2nd ed., ser. Springer Series in Statistics. New York: Springer-Verlag, 2009.

[34] A. Bagnall, H. A. Dau, J. Lines, M. Flynn, J. Large, A. Bostrom, P. Southam, and E. Keogh, "The UEA multivariate time series classification archive, 2018," *arXiv:1811.00075 [cs, stat]*, Oct. 2018.

[35] H. van Zanten, "When is a linear combination of independent fBm's equivalent to a single fBm?" *Stochastic Processes and their Applications*, vol. 117, no. 1, pp. 57–70, Jan. 2007.

[36] A. Dávalos, M. Jabloun, P. Ravier, and O. Buttelli, "Theoretical study of multiscale permutation entropy on finite-length fractional gaussian noise," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1087–1091.