ISSN: 2834-7706

Research Article

Journal of Mathematical Techniques and Computational Mathematics

A Wavelet-Based Approach for Similar Pattern Detection in Time Series

Pierpaolo Massoli*

Directorate for Methodology and Statistical Process Design (DCME), Italian National Institute of Statistics (ISTAT), Italy

***Corresponding Author**
Pierpaolo Massoli, Directorate for Methodology and Statistical Process Design (DCME), Italian National Institute of Statistics (ISTAT), Italy.

Submitted: 2025, Feb 18; **Accepted:** 2025, Mar 18; **Published:** 2025, Mar 25

Citation: Massoli, P. (2025). A Wavelet-Based Approach for Similar Pattern Detection in Time Series. *J Math Techniques Comput Math, 4*(2), 01-09.

Abstract

The analysis of randomness in time series is crucial for extracting relevant pat- terns from noisy data. Noise can obscure underlying dynamics, posing challenges in various research fields such as financial analysis, biomedical signal processing, and environmental monitoring. This study proposes a novel method for detecting similar patterns in large-scale time series datasets. The approach employs a denoising technique based on the Morlet wavelet transform to enhance pattern recognition. The similarity-search method leverages Locality Sensitive Hashing in order to detect denoised similar patterns embedded within time series. A significant reduction in entropy in the reconstructed data reveals hidden patterns that were previously masked by noise. This study avails of entropy as a measure of detection accuracy, incorporating a well-known technique from the Conformal Prediction framework. The pre-defined confidence level is closely related to the minimum cosine similarity of the detected patterns which exchange high values of mutual information as a consequence of the noise removal as demonstrated in the case study.

Keywords: Wavelet Denoising, Pattern Recognition, Mutual Information, Locality Sensitive Hashing, Conformal Prediction

1. Introduction

Exploring the randomness in time series is a fundamental task when discovering structured patterns in data pertaining to various research fields [1]. Noise disturbance in the data may mask interesting dynamics, requiring advanced techniques for effective signal processing. Wavelet denoising is a prominent approach that leverages wavelet transforms to isolate noise from relevant patterns [2]. This type of transforms pro- vides a time-frequency decomposition suitable for non-stationary signals, as opposed to the well-known Fourier transform, which is appropriate for frequency analysis of periodic signals but may lack in time resolution [3]. The Morlet wavelet, which com- bines a Gaussian envelope with a sinusoidal wave, is particularly effective in isolating local oscillatory behaviors within noisy time series [4]. Wavelets are extensively used for data compression, noise reduction, and feature extraction [5]. Soft-thresholding techniques when applied to wavelet coefficients are effective in reducing random noise in order to enhance the structure of the signal [6]. The proposed wavelet-based denoising method leverages the Morlet wavelet transform for decomposing a time series into its components in the frequency domain in accordance with a user-defined number of resolution scales, isolating high-frequency noise from the low-frequency signal. Mutual information shared between these components is evaluated in order to set the components with minimal information contribution which are typically related to noise equal to zero [7]. The remaining components are used to reconstruct a denoised version of the input data. The proposed method acts as an adaptive filter by dynamically selecting wavelet components based on mutual information minimization, filtering noisy components without requiring the application of a unique threshold to all time series, thus avoiding potential information loss. The Locality Sensitive Hashing is successfully adopted for exploring time series randomness by detecting segments also referred to as blocks of time series pertaining to similar fluctuations around the mean of the values in the block [8]. This method efficiently detects local recurring patterns in large datasets of time series, providing an effective scalable approach to pattern recognition. This study demonstrates that filtering the inherent noise in time series effectively increases the number of similar patterns identified revealing hidden dynamics. The effect of the proposed wavelet pre-processing is evaluated by measuring the

entropy in each time series subsequent to their denoising. Entropy quantifies the unpredictability or information content within a time series by measuring the distribution of probabilities associated to its observed states. It is widely used in analyzing the complexity and randomness of patterns in diverse fields [9]. In this study the evaluation of the entropy is an effective approach for analyzing pairs of similar blocks for a subsequent application of a Conformal Prediction technique in order to evaluate the reliability of the proposed detection algorithm. The entropy of the first block in the pair is considered as being the true value while the entropy of the other is considered as being the predicted value of an ideal model so that the residuals are computed for each pair detected as being the difference of these values and subsequently transformed into non- conformity scores, which are standardized by means of the standard deviation of the residuals estimated in a calibration set. In this study prediction intervals reflect the uncertainty of the detection algorithm in maintaining a pre-defined confidence level for the coverage of the solution found by using the proposed detection algorithm as it is related to the minimum value of the similarity in the detected pairs. Conformal Prediction is a reliable distribution-free framework for uncertainty quantification in

machine learning [10-12]. The entropy-based validation further reinforces its adaptive nature, ensuring optimal pattern detection in time series data. A real-world case study is reported in order to investigate the potential of the proposed approach.

2. Theoretical Background

In order to introduce the fundamental aspects of the detection process to the reader, some notions from the Wavelet Theory as well as the basic concepts of the Locality Sensitive Hashing, Entropy, Mutual Information and Conformal Prediction framework are reported in this section.

2.1. Basics of Wavelet Transform

The Wavelet Transform (WT) is a mathematical technique used to analyze signals by decomposing them into different frequency components while preserving time localization. Unlike the Fourier Transform, which represents a signal as a sum of infinite sinusoidal waves, the Wavelet Transform uses localized basis functions called wavelets, allowing for multi-resolution analysis. The Continuous Wavelet Transform (CWT) is defined as:

$$W(s,\tau) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-\tau}{s} \right) dt \qquad (1)$$

where ψ(t) is the mother wavelet, s is the scale parameter controlling frequency resolution, and τ is the translation parameter determining time localization. The Discrete Wavelet Transform (DWT) provides a computationally efficient alternative by applying filtering and down sampling, leading to a multi-level signal decomposition. It is given by:

$$W(j,k) = \sum_n x(n) \psi^*_{j,k}(n) \qquad (2)$$

where j represents the decomposition level and k denotes the position index. The
inverse wavelet transform allows the reconstruction of the original signal as:

$$x(t) = \sum_j \sum_k W(j,k) \psi_{j,k}(t) \qquad (3)$$

A key application of wavelets is signal decomposition, where the signal is iteratively broken down into approximation and detail coefficients. At each level j, the signal is split into an approximation Aj (low-frequency content) and a detail Dj (high- frequency content), using low-pass h[n] and high-pass g[n] filters:

$$A_j[n] = \sum_k h[k-2n] A_{j-1}[k], \quad D_j[n] = \sum_k g[k-2n] A_{j-1}[k] \qquad (4)$$

This decomposition continues recursively, forming a multi-resolution analysis (MRA). Among the commonly used wavelets, the Morlet wavelet is widely adopted due to its optimal balance between time and frequency localization, defined as:

The wavelet power spectrum, expressed as W (f, t) 2, provides insight into the distribution of signal energy across different frequencies and time intervals. Normalization techniques ensure comparability with the Fourier spectrum, preserving the total energy:

$$\psi(t) = e^{i\omega_0 t} e^{-t^2/2} \qquad (5)$$

where S˜(f) is the wavelet spectrum and u′2 is the variance of the signal. This approach enables accurate analysis of time-varying signals, detecting transient fluctuations and frequency components that may be obscured in traditional stationary methods. Wavelet denoising using the Morlet transform involves thresholding wavelet coefficients in the CWT in order to remove noise while preserving significant signal features. This method is especially useful for enhancing pattern detection in noisy time series.

2.2. Locality Sensitive Hashing Fundnamentals

In data science Locality Sensitive Hashing (LSH) refers to a method designed for an approximate similarity search in high-dimensional spaces where traditional search methods become computationally expensive. There are several metrics that LSH encompasses for finding near-duplicates by means of a suitable family of hash functions $h(\cdot)$ which establish a relation between two input data points $(x_k, x_h) \in X$ and the probability of sharing the same hash code: $sim(x_k, x_h) = Pr[h(x_k) = h(x_h)]$. The choice of the hash function determines the metric to approximate. Every family associates input data to integers which are thought of as being buckets with the purpose of hashing is to group similar data points together into the same bucket so that neighboring data fall into the same bucket with a high probability while data which are likely to be distant in the input space belong to different buckets. In a database context, this facilitates the detection of pairwise similar observations in accordance with varying degrees of similarity. The LSH family known as Random Projections adopted in this study

is tailored for evaluating the cosine similarity between numerical sequences. This family implements the Johnson-Linden Strauss lemma which states that data belonging to high dimensional spaces can be projected onto a lower-dimensional space nearly pre-serving pairwise distances. In order to carry out this task, a set of randomly generated hyperplanes in the input space is used to project every sequence onto a lower dimensional space. Each hyperplane is considered as a decision boundary so that neighboring data points are inserted into the same bucket while they are inserted into different buckets if they are not neighbors. To be more precise, by generating a matrix **P** with elements $\{p_{ij}\} \sim \mathcal{N}(0, 1)$ which has as many rows as the dimensions of the input space and a number of columns equal to the pre-defined number n of hyperplanes, the hash code of the sequence xk is given by setting every ith element of the vector-matrix prod-uct (x_k, **P**) equal to 0 if the product of the sequence and the ith column of the matrix is negative and equal to 1 otherwise. The distribution $\mathcal{N}(0, 1)$ denotes a standard nor- mal distribution with mean 0 and standard deviation 1. The number of hyperplanes affects the maximum number of buckets to which the data points are associated: In practical applications the typical values that ensure hashing with a reduced number of collisions are $n = 32$ or higher values. By multiplying every input sequence of L elements by a sequence of H randomly generated ($L \times n$) matrices $\{P_1, P_2, \ldots, P_H\}$, the input dataset is transformed into a dataset of signatures which are sequences of H i.i.d. hash codes. As a result the input dataset is transformed into a ($L \times H$) signature matrix. Subsequent to the creation of the signatures matrix in order to speed up the near-duplicates search, LSH shrinks the signatures into B bands. Each band consists of R adjacent hash codes combined together so that the relation H = BR holds. Similar sequences are finally detected by sorting the ($N \times B$) banded matrix and sequentially scanning it B times. Every pair of consecutive signatures with at least one corresponding equal band indicates a pair of near-duplicate input sequences. The probability of being a pair of similar objects with a similarity value σ is given by:

$$p(\sigma, R, B) = 1 - (1 - \sigma^R)^B \qquad (7)$$

It is widely reported in the literature that the LSH is an approximate method that can give rise to *false duplicates* in the solution. The rate of the same is usually controlled by an appropriate tuning process of the hyperparameters.

2.3. Entropy: A Measure of Uncertainty

In the context of information theory entropy is a fundamental measure of uncertainty in a system. It is a non-negative measure for quantifying the information contained in a probability distribution in order to assess the unpredictability as well as the dispersion of data. entropy is defined as:

$$\mathcal{H}(X) = -\sum_{i=1}^{N} p_i \log_b p_i \qquad (8)$$

where p is the probability that a random variable X may be in state i, the same being selected among N possible states. Shannon entropy is calculated by setting the base b in the formula equal to 2. A low entropy value indicates that the system is highly predictable, while a high entropy value suggests a greater unpredictability in the data. It is straightforward that a sequence consisting of only one repeated value has entropy zero (no uncertainty) in contrast to a sequence in which all values occur with equal probability, implying a maximum value of the entropy (system uncertainty). Applications of Entropy In this study entropy is proposed as a measure of the accuracy of the proposed detection algorithm.

2.4. Mutual Information: Basic Concepts

Mutual information is a well-known measure of the dependency in time series analysis. This measure accounts for non-linear dependencies and requires no specific theoretical probability distribution assumption in order to be estimated. Mutual information as well as correlation are measures of association between variables, each capturing different aspects of the relationship. Correlation measures the strength and direction of a linear relationship between two continuous variables while mutual information measures the amount of information exchanged between variables which captures any type of relationship. It appears obvious that they are complementary measures describing different aspects of the association between two random variables (X, Y). Mutual information is related to entropy $\mathcal{H}(\cdot)$ as reported below:

$$\mathcal{MI}(X, Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y) \qquad (9)$$

which can be normalized by dividing it by $max(\mathcal{H}(X), (\mathcal{H}(Y)$. Equation 9 indicates that mutual information gains as the degree of regularity increases, implying that observing one variable provides a better prediction in relation to the other. Therefore information sharing within time series can be investigated by estimating the mutual information exchanged by their similar detected blocks. Randomness is higher in less probable events, i.e. high degrees of randomness may emerge, following the detection of a few pairs of similar blocks. As it is well known from the literature, mutual Information is closely related to the *Jensen-Shannon Divergence* (JSD), which quantifies the dissimilarity between two probability distributions. The JSD is defined as the average of the Kullback-Leibler Divergence between each distribution and their mean. Since JSD measures the divergence between the entropy distributions of detected blocks, its reduction implies an increase in mutual information, reinforcing the statistical dependency between the identified segments. As a matter of fact, mutual information is bounded above by a function of JSD, meaning that as JSD decreases, the detected patterns tend to share more information, leading to stronger statistical dependence."

2.5. A Quick Glance at Conformal Prediction

Conformal Prediction (CP) is a versatile framework for quantifying uncertainty in a machine learning model, yielding to a guaranteed coverage relating to its prediction intervals without requiring assumptions about data distribution. This framework is particularly robust insofar as model reliability is a critical concept as far as practical applications are concerned. Prediction intervals

are evaluated by means of a calibration set in order to compute *non-conformity scores*. In this study these latter are calculated as standardized residuals defined as follows:

$$s_i = \frac{|y_i - \hat{y}_i|}{\sigma}, \tag{10}$$

where y_i and $\hat{y}_i$ are respectively the observed (true) and predicted values and σ rep-resents the standard deviation of the residuals computed by using data belonging to the calibration set. The prediction intervals for every observation x belonging to the test set are evaluated as follows:

$$\mathcal{C}(\mathbf{x}) = [\hat{y}_i - \hat{q}\sigma, \hat{y}_i + \hat{q}\sigma], \tag{11}$$

where $\hat{q}$ is the (1 α) (n+1) /n quantile with n equal to the number of observations in the calibration dataset and α is the user-defined level of accuracy. This split-conformal approach requires that the dataset is partitioned into disjoint calibration and test sets. The CP framework requires a dataset of n observations $\{(X = x_i, Y = y_i)\}$ (i = 1, 2, . . . , n) with features $X \in \mathbb{R}^d$ and response $y \in \mathbb{R}$, the prediction interval $C(X) = [L(X), U(X)]$ covers the true value with probability:

$$\Pr(y \in \mathcal{C}(X)) \geq 1 - \alpha, \tag{12}$$

In order to further evaluate the properties of prediction intervals whilst ensuring a pre-fixed coverage the *Mean Interval Width* (MIW) is used. It is defined as:

$$MIW = \frac{1}{n}\sum_{i=1}^{n}(U_i - L_i) \tag{13}$$

where U_i and L_i are the upper and lower bounds of the prediction interval for obser-vation i in a test set containing n observations. In order to normalize this measure relative to the scale of the target variable, the *Relative Interval Width* (RIW):

$$RIW = \frac{\text{MIW}}{y_{\max} - y_{\min}} \tag{14}$$

where y_{max} and y_{min} denote the maximum and minimum observed values of the target variable. A lower MIW indicates narrower intervals, while a lower RIW facilitates comparisons across different datasets or models by adjusting for data scale variations. These measures ensure a precise evaluation of the adaptivity of the prediction intervals.

3. Detection of Similar Patterns in Time Series

The proposed algorithm is devised for detecting similar blocks embedded in a large dataset of time series. The workflow of the algorithm is described in this section.

3.1 Time Series Denoising by Using the Morlet Decomposition

In the proposed approach, a time series of length T is processed by means of a wavelet filtering based on the Morlet transform for decomposing the time series into components in the frequency domain in accordance with a pre-fixed number of resolution scales N_{scale}. The basics of the Morlet wavelet have been briefly described in Section 2. These aforementioned components make up the columns of a / T (N_{scale} + 1)) complex-valued matrix in which each column is a component related to a specific frequency scale. Columns are aligned so that those on the left are connected to lower frequency components while those on the right are related to higher frequency components. Noise filtering solely takes the real part of this matrix into account. Commencing from the first column on the left side of the matrix, the components are iteratively considered in order to reconstruct the time series for an evaluation of the mutual information exchanged between the same and the other time series reconstructed in which the remaining components are used. The process is carried out for j = 1, 2, . . . , N_{scale} in order to seek for the index J which splits the matrix in two disjoint partitions so that the mutual information exchanged between the time series has a minimum value.

As a consequence, the time series which is reconstructed by using those low-frequency components which are scarcely correlated to the noise. The reconstructed time series is defined as follows: where ψ is the Morlet transform defined as follows:

Algorithm 1 Decomposition of the time series `ts.decomp`

input: $ts.input$, # scales N_{scale}, time series length T, central frequency ω^*
Output: matrix of wavelet transform coefficients $WT(j,k)$
for $j = 0$ to N_{scale} **do**
  for $k = 1$ to T **do** $WT(k, j+1) \Leftarrow \psi(t = 1 : T, \omega = \omega^*, j = j, \tau = k) \cdot ts.input$
  end for
end for
return WT

$$\psi(t) = \frac{1}{\sqrt{d}}\pi^{-\frac{1}{4}} e^{i\omega\frac{(t-\tau)}{d}} e^{-\frac{(t-\tau)^2}{2d^2}} \tag{15}$$

where t is the time variable, τ is the translation parameter of the wavelet, ω is the central frequency of the wavelet, j is a scale parameter, and d is the dilation factor, defined as $d = 2\ 2^{\ j \cdot 0.125}$. This form represents a complex wavelet, characterized by an exponential modulation with frequency ω and a Gaussian envelope ensuring localization in both time and frequency. The proposed filtering algorithm is described in Algorithm 3.1 The Inverse Morlet Wavelet Transform in the filtering algorithm is described in Algorithm 3.1, while the function for the reconstruction of the time series in the time domain is reported in in Algorithm 3.1. In order to evaluate the mutual information exchanged between the time series of first-order differences resulting from the denoised time series and the complementary one as 9s reported in Algorithm 3.1 is required a further processing provided by the function discretize which returns 1 when the first-order deferred is equal to 1 and 0 otherwise.

3.2. Similar Blocks Detection

Consider an input dataset X with N time series each one being T time periods long[1]. A time window of length $L < T$ is used to break every input series down into blocks $x^{(k)} = \{x_{(k-1)+1}\ \ \mu_k, x_{(k-1)+2}\ \ \mu_k, \ldots, x_{(k-1)+L}\ \ \mu_k$ by shifting them one period k at a time so that the total number of resulting blocks C from the input dataset is equal to $N\ (T - L + 1)$. The variable μ_k is the average of all values in the block. The thus created input dataset is transformed into the signatures matrix as described in Section 2. The LSH-family of Random Projections approximates the pairwise cosine similarity between the blocks. The solution set is composed by all pairs with a high probability of being similar with a high degree of similarity. Due to the probabilistic nature of the LSH, the presence of false duplicates must be controlled by carefully selecting the parameters $\{H, B, R\}$. Their setting is generally a critical aspect of the

Algorithm 2 Wavelet filtering

input: time series $ts.input$, number of scales N_{scale}
output: filtered time series $ts.lo$
compute first-order difference: $y \leftarrow \texttt{diff}(\texttt{ts.input})$
perform wavelet decomposition: $WT \leftarrow \texttt{ts.decomp}(\texttt{y}, \texttt{N}_{\texttt{scale}})$
extract real part of wavelet coefficients: $WT.real \leftarrow \texttt{Re}(\texttt{WT})$
initialize mutual information matrix: $MI \leftarrow \emptyset$
for $k = 1$ to N_{scale} **do**
    initialize $WT.lo \leftarrow WT.real$
    initialize $WT.hi \leftarrow WT.real$
    set the first k components of $WT.lo[tozero :, 1 : k] \leftarrow 0$
    set the $(N_{scale} + 1) - k$ components of $WT.hi$ to zero: $WT.hi[:, -\{1 : k\}] \leftarrow 0$
    reconstruct time series: $ts.lo \leftarrow \texttt{ts.reconstruct}(\texttt{imwt}(\texttt{WT.lo}), \texttt{ts.input}[1])$
    reconstruct complementary series: $ts.hi \leftarrow \texttt{ts.reconstruct}(\texttt{imwt}(\texttt{WT.hi}), \texttt{ts.input}[1])$
    compute mutual information: $MI[k] \leftarrow \texttt{mutual\_information}(\texttt{discretize}(\texttt{diff}(\texttt{ts.lo})), \texttt{discretize}(\texttt{diff}($
end for
find the index of the minimum mutual information: $id \leftarrow \arg\min(MI)$
select the last occurrence of the minimum: $id \leftarrow id[\texttt{length}(\texttt{id})]$
zero out the first id components in $WT.lo$: $WT.lo[:, 1 : id] \leftarrow 0$
reconstruct the final filtered time series: $ts \leftarrow \texttt{ts.reconstruct}(\texttt{imwt}(\texttt{WT.lo}), \texttt{ts.input}[1])$
return $ts.lo$

Algorithm 3 Inverse Morlet Wavelet Transform: `imwt`

input: wavelet coefficient matrix WT
output: reconstructed time series $ts.denoised$
$T \leftarrow$ number of rows in WT (time series length)
$J \leftarrow$ number of columns in WT minus one
extract real part of wavelet coefficients: $WT.real \leftarrow \texttt{Re}(\texttt{WT})$
compute dilation factors: $dial \leftarrow 2 \cdot 2^{(0:J \cdot 0.125)}$
initialize reconstructed series: $ts.denoised \leftarrow \emptyset$ of length T
for $l = 1$ to t **do**
    $ts.denoised[l] \leftarrow 0.2144548 \cdot \sum \frac{WT.real[l,j]}{\sqrt{dial[j]}}$
end for
return $ts.denoised$

nearest neighbors search insofar as a wrong setting could compromise the goodness of the solution. The parameters in the algorithm proposed here are therefore set to achieve an almost zero false negatives rate in opposition to a probable higher false positives rate. In order to lower the rate of false positives, the number of the pairs detected usually can be reduced by filtering out all the pairs whose cosine similarity is below a pre-defined threshold τ from the set of the detected pairs. This study proposes a Conformal Prediction approach in order to refine the solution set in a pre-defined confidence level without an explicit similarity threshold requirement.

Algorithm 4 Reconstruction of the Denoised Time Series

input: differenced time series $ts.diff$, initial value $ts.x0$
output: reconstructed time series $ts.denoised$
set $T^{'} \leftarrow$ length of $ts.diff$
Initialize reconstructed series: $ts.denoised \leftarrow \emptyset$ of length $T^{'} + 1$
Set initial value: $ts.denoised[1] \leftarrow ts.x0$
for $i = 2$ to $T^{'} + 1$ **do**
    $ts.denoised[i] \leftarrow ts.diff[i-1] + ts.denoised[i-1]$
end for
return $ts.denoised$

3.3. Evaluating the Entropy in the Detected Blocks

Following the detection of the pair of similar blocks the entropy H of each block is evaluated by recurring to the Shannon entropy of Section 2. The function for the evaluation is described in Algorithm 3.3.

Algorithm 5 Evaluating the entropy

input: block of z-scores $\mathbf{x} = z_{score}$, discretization categories b
output: entropy in the block $\mathcal{H}(\mathbf{x})$
discretize block $\mathbf{x}$ into b intervals
evaluate the frequency of each interval in the block
calculate the probabilities p_i by normalizing the frequencies
calculate entropy $\mathcal{H}(\mathbf{x}) = -\sum p_i \log_2 p_i$
return entropy $\mathcal{H}(\mathbf{x})$

3.4. Evaluating the Accuracy of the Detection Approach

The entropy pertaining to the similar blocks belonging to every pair detected is considered as being the true value $y = H_{left}$ and the predicted value $\hat{y} = H_{right}$ of an ideal black-box predictive model which is evaluated by adopting the Conformal Pre- diction approach reported in Section 2. The solution set of the previously mentioned pairs is split into calibration and test datasets, having taken the solution set as being the result of a pre-trained model $y = f(x)$ into account. Vector x is one of the blocks in the pair with the underlying assumption that the blocks in the pair are the same. The prediction intervals pertaining to the dependent variable of this model (entropy of the blocks belonging to the test set partition) are estimated by evaluating the quantile $(1 - \alpha)(n + 1) / n$ of the non-conformity scores by using the calibration data as described in Section 2. In the context of Conformal Prediction, the confidence level $1 - \alpha$ and the accuracy of prediction intervals are closely related. The confidence level

represents the probability that the true value of the target variable falls within the prediction interval and is a user-defined parameter, higher confidence level provides wider intervals, increasing the likelihood of containing the correct value while lower confidence level results in narrow intervals. The accuracy of the prediction intervals refers to the ability of the model to ensure that the empirical coverage matches the expected one. A coverage of 95% implies that the true value of the entropy falls into the intervals 95% of the times. Wide intervals may lack in being informative, whereas narrow intervals may exclude true values. This study proposes a 95% confidence level in order to refine the number of the pairs so that the true value of the entropy is inside the prediction intervals.

4. Application to Financial Markets

As it is known from the literature, financial markets are not perfectly efficient implying that the exploitation of past information is not useless. As market efficiency is closely related to the randomness of the same, it is worth detecting the repeated patterns of price movements over time. On the basis of these concepts, a dataset containing various financial market indices is explored as a case study.

4.1. The Input Dataset

Consider a dataset comprising daily values of $N = 27$ market indices (see Table 1). All data was collected via web scraping. The downloaded time series may have covered different time periods ranging from *2000-01-03* to *2024-02-15*, resulting in varying lengths as indicated in Table 1. The time window used to divide all the series into blocks is set to $L = 10$ days.

The presence of missing data has to be managed prior to starting the proposed method, thus, it was necessary to identify missing data in each time series. Blocks containing at least one missing value were excluded from the analysis. No further data preprocessing was carried out. For this study, the input dataset contains a total of C = 161868 blocks to be investigated for similarities. Standard pairwise comparisons for similar blocks would require approximately 1.31 10^{10} comparisons, rendering the search computationally infeasible. In order to investigate the effect of the proposed wavelet filtering approach for noise removal in time series analysis, the detection of pairs of similar blocks is carried out in the original dataset of eprld market indices (case A) as well as on the denoised version of the same (case B) for comparison.

4.2. Comparison of the Results

The total number of pairs of similar blocks detected in case A is equal to 207713 while in the case B the total number increases to 454052 fir an increment equal to 118.6%. In order to simplify the reading of the results of the comparison between the detections of the similar blocks within the time series with and without noise, only the counts of similar blocks belonging to the same time series are reported in Table 2. In this case the hyperparameters of the detection algorithm are set up to provide that every block is identified by a signature of H = 200 i.i.d. hash codes. Every hash is a n = 32 long

N	Index	Description	T
1	AORD	Ordinaries	6100
2	AXJO	S1&PASX 200	6095
3	BFX	BEL 20	6164
4	BSESN	S&P BSE SENSEX	5948
5	BUK100P	Cboe UK 100	3397
6	DJI	Dow Jones Industrial Average	6069
7	FCHI	CAC 40	6167
8	FTSE	FTSE 100	6093
9	GDAXI	DAX Performance-Index	6127
10	GSPC	S&P 500	6069
11	GSPTSE	S&PTSX Composite index	6060
12	HSI	HANG SENG	5945
13	IMOEX.ME	MOEX Russia Index	2699
14	IXIC	NASDAQ Composite	6069
15	JKSE	IDX Composite	5867
16	KLSE	FTSE Bursa Malaysia KLCI	5916
17	KS11	KOSPI Composite Index	5949
18	N100	Euronext 100 Index	6170
19	N225	Nikkei 225	5911
20	NYA	NYSE Composite (DJ)	6069
21	NZ50	S&PNZX 50 Index Gross & Gross	5204
22	RUT	Russell	6069
23	STI	STI Index	6035
24	STOXX50E	ESTX 50 PR.EUR	4234
25	TWII	TSEC weighted	5919
26	VIX	CBOE Volatility Index	6069
27	XAX	NYSE AMEX Composite Index	6069

Table 1: World Market Indices In in the Dataset

integer which is a sufficient length (in bits) for hashing the blocks with a low number of collisions. Each signature is grouped into B = 50 bands of R = 4 hashes combined in *bitwise XOR*. Due to these hyperparameters the probability of detecting a pair of blocks with a cosine similarity equal to 0.8 equates to 1 (see Equation 7) dropping the probability of false negatives to zero while the probability of false positives increases. As mentioned in the previous Section, the proposed Conformal Prediction approach for the evaluation of the accuracy of the detection of similar blocks implies that a coverage of the intervals of 95% as well as a good adaptivity of the same implies a high similarity between the blocks not requiring therefore the setting of a user-defined threshold in order to decrease the number of false positives.

The evaluation of RIW reveals a significant adaptivity of the prediction intervals subsequent to wavelet denoising. The RIW

decreased from 0.6667 to 0.2579, corresponding to a 61.31% reduction. This substantial decrease suggest a significant reduction of uncertainty in comparing the entropies pertaining to the blocks of each detected pair. A further assessment of the performance of the wavelet denoising is highlighted by the reduction in Jensen-Shannon Divergence between the distribution of entropies pertaining to the first block of the pairs and the distribution of those pertaining to the second pnes,, which dropped from a value equal to 0.002353539 to 8.363279×10^{-5}, a −96.45% decrease. This suggests that wavelet denoising effectively removes irrelevant variations, enhancing the detection of similar patterns embedded in the time series. Due to the inverse relation between JSD and mutual information, a reduction in JSD implies an increase in mutual information within the detected blocks, confirming that the removal of high-frequency components from the time series is effective in unveiling structured, non-random patterns in the data. A further comparison between the case of the detection of the original market indices and the denoised version of the same is reported in Table 3.

Index	case A	case B	Increment (%)
AORD	651	6660	922.58
AXJO	523	6606	1162.14
BFX	604	7570	1153.64
BSESN	935	12356	1211.23
BUK100P	105	15097	14283.81
DJI	429	10099	2254.30
FCHI	360	6209	1624.72
FTSE	348	7081	1935.06
GDAXI	459	8345	1717.65
GSPC	334	7505	2146.71
GSPTSE	575	6998	1117.91
HSI	492	53253	10622.56
IMOEX.ME	108	25435	23537.04
IXIC	475	46745	9744.21
JKSE	518	9771	1786.48
KLSE	895	24119	2595.64
KS11	615	5951	867.64
N100	390	7138	1729.23
N225	423	40742	9529.31
NYA	359	7046	1862.12
NZ50	833	6085	630.85
RUT	378	7700	1937.57
STI	650	35242	5329.54
STOXX50E	169	7119	4113.02
TWII	614	8108	1220.52
VIX	159	11701	7260.38
XAX	660	12999	1869.55

Table 2: Pairs of Similar Blocks in Time Series

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
case A						
CS > 0	0.8502	0.9993	0.9999	0.9970	1.0000	1.0000
CS < 0	-1.0000	-1.0000	-1.0000	-0.9958	-0.9986	-0.8501
case B						
CS > 0	0.9646	0.9999	1.0000	0.9998	1.0000	1.0000
CS < 0	-1.0000	-1.0000	-0.9998	-0.9992	-0.9992	-0.9453

Table 3: Cosine Similarity of the Pairs Detected

In the context of the proposed approach based on Conformal Prediction, the requirement of 95% coverage of the prediction intervals for the entropy in pairs of similar blocks implies a high value of their cosine similarity, both when the pairs have concordant trend ($CS > 0$ and when they have opposite trend ($CS < 0$). The results reported in Table 3 highlight the positive effect of the noise removal achieved by means of the wavelet filtering. In order to assess the statistical informativeness of the results, a further investigation focuses on the mutual information exchanged within the blocks detected. A summary is reported in Table 4.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
case A						
CS > 0	0.1889	0.6407	0.7004	0.6935	0.7547	1.0000
CS < 0	0.1717	0.6407	0.7004	0.6931	0.7621	1.0000
case B						
CS > 0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
CS < 0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4: Mutual Information Exchanged in the Pairs Detected

As is reported from the results, wavelet denoising together with the evaluation of prediction intervals of the entropies in the blocks in order to refine the solution pro- vides a sensitive improvement of the detection of repeated structured patterns in the input time series resulting in a high cosine similarity which indicates that the time series exhibit local similar patterns in direction, which as it is well known from the literature is uncorrelated to mutual information exchanged by the detected patterns. The removal of high-frequency noise while preserving the fundamental structure of the time series implies that cosine similarity renders a more robust measure of patterns similarity as it is not affected by noise. As a consequence, these denoised structured patterns are strongly similar in direction and exhibit almost equal entropy values in each detected pair, implying high values of the mutual information exchanged between them as reported in Table 4.

5. Conclusion

This study proposed an enhanced methodology for detecting similar patterns in time series by integrating wavelet-based denoising with Locality Sensitive Hashing and Con- formal Prediction. The approach leverages the Morlet wavelet transform to filter out high-frequency components, thus highlighting patterns with a structure which might remain obscured in the original time series. Although LSH is devised for detecting similar patterns based on geometric properties, such as cosine similarity as is the case for Random Projections hashing functions, the integration of wavelet denoising modifies the nature of the detected similarities. By applying entropy-based validation in a Conformal Prediction framework, the method identifies patterns that, beyond their geometric resemblance, exchange a high mutual information. This aspect is particularly relevant since mutual information is a statistical measure of dependency that is not necessarily related to cosine similarity. The results demonstrate that noise filtering

not only increases the number of detected patterns but also refines their statistical coherence, ensuring that detected patterns are not just structurally similar but also statistically significant. The application to financial market indices validated the effectiveness of this approach as it resulted is a substantial increase in the number of detected patterns subsequent to denoising and a significant reduction in uncertainty in the evaluation of their entropy. The observed increase in mutual information between detected pairs reinforces the hypothesis that reducing high-frequency noise enhances the identification of relevant patterns that go beyond simple geometric similarity. These findings highlight the potential of integrating wavelet filtering and entropy-based validation within similarity search frameworks, providing a more robust approach to uncovering complex dependencies in time series data.

References

1. Percival, D. B., & Walden, A. T. (2000). Wavelet methods for time series analysis (Vol. 4). Cambridge university press.
2. Mallat, S. (1999). A wavelet tour of signal processing. *Elsevier*.
3. Daubechies, I. (1992). Ten lectures on wavelets. Society for industrial and applied mathematics.
4. Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. Bulletin of the *American Meteorological society, 79*(1), 61-78.
5. Strang, G., & Nguyen, T. (1996). Wavelets and filter banks. SIAM.
6. Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *biometrika, 81*(3), 425-455.
7. Cover, T. M. (1999). Elements of information theory. *John Wiley & Sons*.
8. Massoli, P. (2024). Exploring Time Series Randomness. *Curr Res Stat Math, 3*(1), 01-07.
9. Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal, 27*(3), 379-423.
10. Shafer, G., & Vovk, V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research, 9*(3).
11. Angelopoulos, A. N., & Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*.
12. Tibshirani, R. (2023). Conformal Prediction. Stat 154/254: Modern Statistical Prediction and Machine Learning, Spring 2023. University of Berkeley