Presentation of Coupling Analysis Techniques of Maximum and Minimum Values Between N Sets of Data Using Matrix $[\mu]^{[MKN]}$

K. N. Makris
Department of Mathematics,
School of Applied Mathematical and Physical Sciences,
National Technical University of Athens, Zografou, Greece.
Corresponding author: constantinosmakris@yahoo.gr

I. Vonta
Department of Mathematics,
School of Applied Mathematical and Physical Sciences,
National Technical University of Athens, Zografou, Greece.
E-mail: vonta@math.ntua.gr

Abstract
This paper deals with the presentation and study of alternative coupling techniques for maximum and minimum values between data sets, namely the problem which is examined in this work is the possible appearance of maximum or minimum values between data sets in the same or neighboring time points. The data can be time-dependent (time series) or non-time-dependent. In this work, the analysis is focused on time series and novel indices are defined in order to measure whether the values of N sets of data display in terms of time, the maximum or minimum values at the same instances or at very close instances. For this purpose, two methods will be compared, one direct method and one indirect method. The indirect method is based on Matrices of dimensionless indicators which are denoted by $[\mu]^{[MKN]}$, and the direct method is based on a variance-type measure which is denoted by $[V]^{[MKN]}$.

Keywords- Coupling techniques, Time series, Indices, Index matrix $[\mu]^{[MKN]}$.

1. Introduction

The purpose of data analysis is from one point of view to understand the variation of the data, in order to be able to predict future values for the case that the data depend on time or their frequency of occurrence for the case that the data are independent of time. The tools to achieve this goal are the use of statistical methods such as descriptive statistics, hypothesis testing, regression analysis, analysis of variance, quality control, regression models (Álvarez et al., 2021), time series analysis, etc. Many papers have been published in the field of time series analysis, publications with applications to data derived from demography, economics, financial stability (Nguyen and Bui, 2020) and financial indicators, results from biological laboratories, health science (clinical trials), industrial production lines, etc. The analysis of time series coming from a great variety of situations, is based on the study of various characteristics of the data, such as the Trend, the Periodicity, the Stationarity or Similarity between time series.

The similarity between two (or even many) time series attempts with various techniques to study the common changes of two-time series. The methods of calculating the similarity between two-time series can be achieved with simple mathematical measures (Iglesias and Kastner, 2013), by using some transformations in the data (Lin et al., 2003) or by using algorithms (Keogh et al., 2001; Morse and Patel, 2007; Nakamura et al., 2013; Serra and Arcos, 2012), or are based on local

autopatterns (Baydogan and Runger, 2016), or by using correlation-aware measures (Mirylenka et al., 2017). Another approach are the measures of divergence defined between two probability distributions, the most well-known being the measure of Kullback-Leibler (Kullback and Leibler, 1951), the Csiszar's φ-divergence family of measures (Ali and Silvey, 1966; Csiszár, 1964), the Cressie and Read measures including well-known measures such as the Pearson's χ2 measure and the classical loglikelihood ratio statistics (Cressie and Read, 1984), the BHHJ deviation measure of Basu (Basu et al., 1998), the generalized BHHJ family of measures (Mattheou et al., 2009), and entropy-type measures and divergences with applications in engineering (Koukoumis and Karagrigoriou, 2021). Finally, there are many dissimilarity measures (Huber-Carol et al., 2002; Meselidis and Karagrigoriou, 2020; Toma, 2008, 2009; Zhang, 2002).

The indices and the corresponding matrices that will be presented in the present work are based on and are a continuation of the idea of *MKN* indices defined in Makris (2017) and Makris (2018). Through these indices the simultaneous time pairing (i.e., time display) of the maximum (or minimum) values between time series was compared, based on the values of the time series. *MKN* indices were defined for the first time in experimental data measuring forces, moments, displacements, and rotations for two types of floating wind turbines. More specifically, the combined effect of the anchor lines and the turbine (wind turbine) was considered in relation to the response of the floating body in order to study the hydrodynamic aspects of the floating wind turbine and the undulations of the combined wind and wave action for two data cases, firstly for the case that the data come from regular waves and secondly for the case that the data come from irregular waves (Makris, 2017). Epidemiological data from Greece collected during the period of 2004-2017 and related to influenza were considered in Makris (2018).

The analysis begins in Section 2 with the presentation of a direct way of solving the above problem (of the simultaneous time pairing), illustrated through examples with the help of a modified measure of variance. In Section 3 we introduce an indirect method of comparing time series and the *MKN* indices. In Section 4 we generalize the *MKN* indices by introducing an extra parameter and rename the indices as $\mu^{[MKN]}$ and in the last section we define the Matrix $[\mu]^{[MKN]}$ based on the indices $\mu^{[MKN]}$.

2. Introduction of the Method *V*
For a better understanding and comparison of the concepts that will be presented in the present work, we will first present an alternative method of calculating the coupling of maximum and minimum values between time series. The method is denoted by the letter *V* (since it is related to the definition of Variance) and is a direct method (as opposed to the indirect method that will be presented later) and is calculated based on a modification of the Variance. The analysis in this work will focus on time-dependent data, i.e., time series.

2.1 Definition of the Matrix $[V]^{[MKN]}$
The first part of the analysis begins with the presentation of the measure $V^{[MKN]}$ which is defined for *N*=2 time series, by:

$$V_{(i,j)}^{[M=\text{Max},K,N=2]} = \frac{1}{K}\sum_{r=1}^{K}\left(t^{i(n-r+1)} - t^{j(n-r+1)}\right)^2 \tag{1}$$

In relation (1), *i* and *j* are two time series of equal length (i.e. each time series has n observations), where the notation $t^{i(n-r+1)}$ expresses the time when the time series *i* displays its $(n-r+1)^{th}$

maximum value, and $t^{j(n-r+1)}$ denotes the time when the time series j displays its $(n-r+1)^{th}$ maximum value, while $i_{(n-r+1)}$ represents the $(n-r+1)^{th}$ ordered observation of the time series i and $j_{(n-r+1)}$ represents the $(n-r+1)^{th}$ ordered observation of the time series j (see also Makris et al., 2021).

It is noted that the function defined in (1) as well as others that will follow, depend on three parameters denoted by M, K and N, where the parameter M denotes whether the values of the time series under investigation are Maximum or Minimum, i.e., M is a binary parameter which takes as values the functions Min and Max, or in code 0 for the Minimum values and 1 for the Maximum values. The parameter K stands for the number of the maximum or minimum values based on which the calculations are made, (for instance if K takes the value 3 then the 3 highest values of a time series $\{i_r\}_{r=1}^{n}$ are $i_{(n)}$, $i_{(n-1)}$, $i_{(n-2)}$), and parameter N is the number of time series that are simultaneously compared where in (1) as previously said, the N parameter is equal to 2 (time series i and j).

If M=min the measure defined in (2) takes the form
$$V_{(i,j)}^{[M=\text{Min},K,N=2]} = \frac{1}{K}\sum_{r=1}^{K}\left(t^{i(r)} - t^{j(r)}\right)^2. \tag{2}$$

For the case where the number of time series that are analyzed simultaneously is $N>2$, the matrix of functions $[V]^{[MKN]}$ of dimensions $N \times N$ (relation (3) below) is created, combining the time series in pairs, based on (1) or (2), and the letter V is enclosed in brackets to distinguish it from definition (1) or (2) which refer to a value:
$$[V]^{[MKN]} = \begin{bmatrix} V_{(1,1)}^{[MKN]} & \cdots & V_{(1,N)}^{[MKN]} \\ \vdots & \vdots & \vdots \\ V_{(N,1)}^{[MKN]} & \cdots & V_{(N,N)}^{[MKN]} \end{bmatrix} \tag{3}$$

It is easy to notice the following (for M=min but analogously it holds for the M=max):
$$V_{(i,i)}^{[MKN]} = \frac{1}{K}\sum_{r=1}^{K}\left(t^{i(r)} - t^{i(r)}\right)^2 = 0 \tag{4}$$
$$V_{(i,j)}^{[MKN]} = \frac{1}{K}\sum_{r=1}^{K}\left(t^{i(r)} - t^{j(r)}\right)^2 = \frac{1}{K}\sum_{r=1}^{K}\left(t^{j(r)} - t^{i(r)}\right)^2 = V_{(j,i)}^{[MKN]} \tag{5}$$

and therefore, the matrix $[V]^{[MKN]}$ is symmetric with diagonal elements equal to 0.

The problems created by the application of the above direct method are that in some cases incorrect results are extracted, as it will be evident from the presentation of Examples 1 and 2 in Section 6.

2.2 Definition of the Matrix $[V]^{[M^{Cross}KN]}$

In the previous section the function V was defined for non-complementary variables. In this section the function V will be defined for complementary variables. Complementary variables can be the demand of two substitute goods, where substitute goods are two products that the consumer can use for the same purpose, (some substitute goods are the tea and coffee, the water in one company and the water in another company). The parameters K and N remain unchanged as in the previous section in terms of definitions, while for the parameter M in the setting of this section, maximum and minimum values are used in combination in order to compare the maximum values of one time series with the minimum values of the other time series simultaneously (the parameter M is denoted

here by M^{Cross}). For example, for the case that we study the demand of tea and coffee, as it is well known, when the price of coffee is increased, then the demand of coffee will reduce and the demand of tea (the complementary good of coffee) will increase, in order to replace the decrease in coffee demand.

The measure $V_{(i,j)}^{[M^{Cross},K,N=2]}$ is defined for $N=2$ time series, by:

$$V_{(i,j)}^{[M^{Cross},K,N=2]} = \frac{1}{K}\sum_{r=1}^{K}\left(t^{i_{(n-r+1)}} - t^{j_{(r)}}\right)^2 \tag{6}$$

while the measure $V_{(j,i)}^{[M^{Cross},K,N=2]}$ is defined for $N=2$ time series, by:

$$V_{(j,i)}^{[M^{Cross},K,N=2]} = \frac{1}{K}\sum_{r=1}^{K}\left(t^{j_{(n-r+1)}} - t^{i_{(r)}}\right)^2 \tag{7}$$

where the notations $i_{(n-r+1)}$, $j_{(n-r+1)}$, $t^{i_{(n-r+1)}}$ and $t^{j_{(n-r+1)}}$ have been defined in the previous section (2.1).

For the case where $N>2$, the matrix of measures $[V]^{[M^{Cross},KN]}$ of dimension $N \times N$ is based on (6) and (7) and is defined by:

$$[V]^{[M^{Cross},KN]} = \begin{bmatrix} V_{(1,1)}^{[M^{Cross},KN]} & \cdots & V_{(1,N)}^{[M^{Cross},KN]} \\ \vdots & \vdots & \vdots \\ V_{(N,1)}^{[M^{Cross},KN]} & \cdots & V_{(N,N)}^{[M^{Cross},KN]} \end{bmatrix} \tag{8}$$

In contrast to the previous section (where the matrix $[V]^{[MKN]}$ has the diagonal elements equal to zero), in this section the diagonal elements of matrix $[V]^{[M^{Cross},KN]}$ are not equal to zero, that is,

$$V_{(i,i)}^{[M^{Cross},K,N=2]} = \frac{1}{K}\sum_{r=1}^{K}\left(t^{i_{(n-r+1)}} - t^{i_{(r)}}\right)^2 \neq 0 . \tag{9}$$

Finally, in the present section, the symmetric property (5) of the previous section does not apply, namely,

$$V_{(i,j)}^{[M^{Cross},K,N=2]} \neq V_{(j,i)}^{[M^{Cross},K,N=2]}. \tag{10}$$

3. Definition of the Indices *MKN*
In this section we present the *MKN* indices (Maximum or Minimum *K* values of *N* time series), which were proposed in order to study the existence of time coupling of maximum or minimum values of the time series (depending on what we are interested in each time i.e., maximum or minimum values) and constitute an indirect method of pairing maximum or minimum values, as the time points are correlated based on the actual values of the time series.

These indices are defined under a condition and are calculated from the definition (11) below. For two time series *i* and *j,* the calculation of the *MKN* index of the time series *j* given the series *i* (which is denoted by $MKN_{(j\backslash i)}$) is performed based on the *K* time points that the *K* larger (or smaller values) of the time series *i* are materialized (which constitutes the aforementioned condition). Let $t^{i_{\{n-K+1\}}}$ denote the *K* time points where the time series *i* displays its *K* larger values. The numerator of the index fraction in (11) calculates the average of the *K* values of the time series *j* conditionally on the time points that the time series *i* displays its *K* larger values, when the maximum values are

under consideration (or the minimum, when the minimum values are under consideration), while in the denominator is the average value of all the values of the time series j, denoted by j^{Total}.

The comparison between the two time series (N=2) is achieved by comparing the index $MKN_{(j \backslash i)}$ with the corresponding MKN index of the time series i itself, which is defined as the ratio of the average of the K maximum values of the time series i divided by the average of all the values of the time series i (and which is denoted by $MKN_{(i \backslash i)}$). If the value of the index $MKN_{(j \backslash i)}$ is close to the value of the index $MKN_{(i \backslash i)}$, then this means that the two time series display their K maximum values in the same or closely situated time points. For the maximum values, the index is defined as:

$$MKN_{(j \backslash i)} = (\text{Max}, KN)_{(j \backslash i)} = \frac{Average\left(\left[j \backslash t^{i\{n-K+1\}}\right]\right)}{Average\left(\left[j^{Total}\right]\right)} \tag{11}$$

4. Generalization of the Indices to $\boldsymbol{\mu}^{[MKN]}$

In this section we generalize the definition of the index MKN defined in (11) by introducing the definition (12) below, of which definition (11) is a special case. The generalized indices are denoted by $\mu^{[MKN]}$, where the parameter μ introduced plays the role of a statistical function or statistic. More specifically, the $\mu^{[MKN]}$ index is defined as the ratio of the statistical functions μ^K and μ^T, where μ^K is a statistical function of the K maximum or minimum values (based on the condition defined in the previous section), while respectively μ^T is the statistical function of all (total) values of the time series under consideration. The statistical functions μ^K and μ^T can be one of the following known statistical functions such as the Average, the Median, the Variance, the Standard Deviation, the Covariance or Correlation or some other statistical characteristic or even a constant value. In fact, μ^T does not have to be the same function as μ^K (i.e., the numerator can be the Standard deviation and the denominator the Average and as a result a generalized definition of the coefficient of variation arises).

For the case that the statistical functions in the numerator and denominator are the same, for example the Average, i.e., $\mu = \frac{Average^K}{Average^T}$, then it will be written explicitly as $Average$ in place of µ. In general, the definition of the index $\mu^{[MKN]}$ for M=Max is as follows

$$\mu_{(j \backslash i)}^{[\text{Max}, KN]} = \frac{\mu_{(j \backslash i)}^K}{\mu_{(j \backslash j)}^T} = \frac{\mu\left(\left[j \backslash t^{i\{n-K+1\}}\right]\right)}{\mu\left(\left[j^{Total}\right]\right)} \tag{12}$$

Similarly, for M=Min we have

$$\mu_{(j \backslash i)}^{[\text{Min}, KN]} = \frac{\mu_{(j \backslash i)}^K}{\mu_{(j \backslash j)}^T} = \frac{\mu\left(\left[j \backslash t^{i\{K\}}\right]\right)}{\mu\left(\left[j^{Total}\right]\right)} \tag{13}$$

where by $t^{i\{K\}}$ we denote the K time points where the time series i displays its K smaller values.

5. Definition of the Matrix $[\boldsymbol{\mu}]^{[MKN]}$

In this section we will present the Matrix $[\mu]^{[MKN]}$ based on the generalized index $\mu^{[MKN]}$ defined in (12) and (13) in the previous section, for the case that N time series are analyzed simultaneously.

The N time series on which the $[\mu]^{[MKN]}$ matrix is calculated should have the same sampling step (not necessarily constant) and the same duration. The calculations in the matrix $[\mu]^{[MKN]}$ are being

done by row, i.e., in the ith row of the matrix the indices $\mu_{(j\backslash i)}^{[MKN]}$ are calculated conditionally on the time points of the K maximum (or minimum) values of the time series i. More specifically, one first needs to locate the K time points where the corresponding K maximum (or minimum) values of the i time series are presented   and then based on these time points the corresponding indices $\mu_{(j\backslash i)}^{[MKN]}$ are calculated for all $j=1,..,N$, including also the diagonal index $\mu_{(i\backslash i)}^{[MKN]}$ which as we previously discussed is the value with which all other time series are compared with the time series i in terms of similarity.

For example, in the general case of the matrix $[\mu]^{[MKN]}$ defined in (14) below which is of dimension $N\times N$, the value of the $\{1, N\}$- element of the matrix denoted by $\mu_{(N\backslash 1)}^{[MNK]}$ calculates the index of the time series N based on the K time points in which the corresponding K maximum (or minimum) values of the first time series are presented. The matrix is defined as

$$[\mu]^{[MKN]} = \begin{Vmatrix} \mu_{(1\backslash 1)}^{[MKN]} & \cdots & \mu_{(N\backslash 1)}^{[MKN]} \\ \vdots & \vdots & \vdots \\ \mu_{(1\backslash N)}^{[MKN]} & \cdots & \mu_{(N\backslash N)}^{[MKN]} \end{Vmatrix} \tag{14}$$

As an example, if $\mu=$ average, the maximum values are under consideration and the value of the parameter N is equal to 2, (i.e., two time series are analyzed) the format of the above matrix reduces to the matrix defined in (15).

$$[Average]^{[Max,K,N=2]} = \begin{Vmatrix} Average_{(1\backslash 1)}^{[Max,K,N]} & Average_{(2\backslash 1)}^{[Max,K,N]} \\ Average_{(1\backslash 2)}^{[Max,K,N]} & Average_{(2\backslash 2)}^{[Max,K,N]} \end{Vmatrix} \tag{15}$$

6. Results and Discussion

It this section we study the aforementioned two methods (the direct method via function V and the indirect method using the proposed indices μ) through two numerical examples for two different cases of pairs of time series.

6.1 Example 1 (Function V)

Consider two time series A and B consisting of 20 values, where the data are given in rows 2 and 3 respectively of Table 1.

In Figure 1 it is shown that there is in fact same coupling between the two time series (A and B) in terms of the time instances of the four maximum values ($K=1,..,4$), as it is observed that both time series display their four largest values simultaneously in the following four time points, $t^{A(20)} = t^{B(20)} = 20$, $t^{A(19)} = t^{B(19)} = 19$, $t^{A(18)} = t^{B(18)} = 18$, $t^{A(17)} = t^{B(17)} = 5$.   More specifically the time series A and B have their maximum value which corresponds to the case $K=1$, at the same time point as it can be seen from Figure 1, which is in fact the twentieth observation for both data sets, (which leads to $t^{A(20)} - t^{B(20)} = 20 - 20 = 0$). Also, they have their second maximum value at the same time point (which is the nineteenth observation for both data sets), and at the same time they both have their third and fourth maximum values at the eighteenth and fifth observation respectively.

Resulting in the values of the function V for $K=1,..,4$ to be:
$$V_{(A,B)}^{[Max,K=1,N=2]} = V_{(A,B)}^{[Max,K=2,N=2]} = V_{(A,B)}^{[Max,K=3,N=2]} = V_{(A,B)}^{[Max,K=4,N=2]} = 0 \tag{16}$$

6.2 Example 2 (Function *V*)

Consider two time series *C* and *D* with 20 values each, where the data are given in rows 4 and 5 respectively of Table 1.

Figure 2 graphically shows the values of the two time series where it appears that the two time series *C* and *D* display their four largest values within the set of times {1,2,19,20} but in an alternating order and more specifically: $t^{C(20)} = 1, t^{C(19)} = 20, t^{C(18)} = 2, t^{C(17)} = 19$ and $t^{D(20)} = 20, t^{D(19)} = 1, t^{D(18)} = 19, t^{D(17)} = 2$. This combination has a significant effect to the values of function *V* as in this case the function *V* takes very large values for all the values of *K*=1,..,4, although this result is misleading because the two time series have their four maximum values as a combination at similar time points. The values of the function *V* for *K*=1,…4 are equal to:

$$V_{(C,D)}^{[Max,K=1,N]} = 361, \quad V_{(C,D)}^{[Max,K=2,N]} = 361, \quad V_{(C,D)}^{[Max,K=3,N]} = 337, \quad V_{(C,D)}^{[Max,K=4,N]} = 325 \qquad (17)$$

Based on these results, it appears that the direct method (i.e., definition (1)) gives satisfactory results in the first example, but large values in the second example for *K*=1,..,4. The above two examples show the inability of the direct method to give satisfactory results for some cases. For this reason, the index μ will be preferable which produces, for both cases, more reliable results.

Table 1. Data of time series *A*, *B*, *C* & *D*.

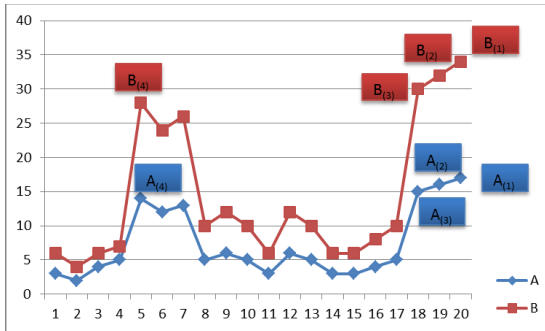| Time order of data	A	B	C	D
1	3	6	34	18
2	2	4	27	14
3	4	6	13	8
4	5	7	15	9
5	14	28	16	10
6	12	24	17	11
7	13	26	14	9
8	5	10	12	8
9	6	12	18	13
10	5	10	17	12
11	3	6	14	9
12	6	12	13	8
13	5	10	9	5
14	3	6	12	7
15	3	6	13	8
16	4	8	10	6
17	5	10	13	9
18	15	30	18	10
19	16	32	24	17
20	17	34	32	25

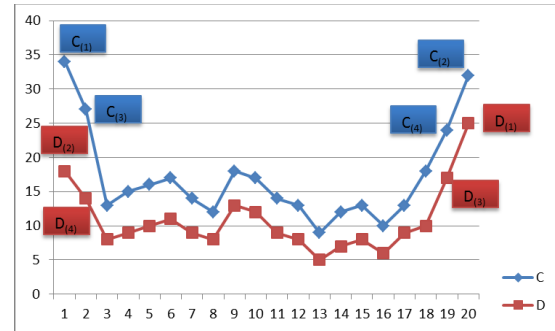**Figure 1.** Time series *A&B*.

Figure 2. Time series *C&D*.

6.3 Examples 1 and 2 (Indices μ)

For the case of Example 1 introduced before notice in Table 2 that the values of the indices $MKN_{(A\backslash B)}$ are equal to the indices $MKN_{(A\backslash A)}$ as well as the indices $MKN_{(B\backslash A)}$ are equal to the indices $MKN_{(B\backslash B)}$ for each value of the parameter *K* considered, a result which shows that the time series *A* and *B* display their maximum values at the same time points for all *K* = 1, 2, 3, 4. We arrive at the same conclusion as with the direct method and the measure $V^{[MKN]}$. Also observe that as the value of the parameter *K* increases the value of the *MKN* indices decreases since the numerator of the index tends to the value of the denominator.

Table 2. *MKN* indices for time series *A & B*.

	$MKN_{(A\backslash B)} = MKN_{(A\backslash A)}$	$MKN_{(B\backslash A)} = MKN_{(B\backslash B)}$
K=1	2.33	2.37
K=2	2.26	2.30
K=3	2.19	2.23
K=4	2.12	2.16

For the case of Example 2 we notice in Table 3 that the values of the indices $MKN_{(C\backslash D)}$ are equal to the indices $MKN_{(C\backslash C)}$ and the values of the indices $MKN_{(D\backslash C)}$ are equal to the indices $MKN_{(D\backslash D)}$ for the values of the parameter *K*=2 and *K*=4, a result which shows that the time series *C* and *D* display their *K* larger values in alternating order in terms of time. This index gives better results than the direct method while the measure $V^{[MKN]}$ between *C* and *D* gave extremely high values.

Table 3. *MKN* indices for time series *C & D*.

	$MKN_{(C\backslash D)}$	$MKN_{(C\backslash C)}$	$MKN_{(D\backslash C)}$	$MKN_{(D\backslash D)}$
K=1	1.88	1.99	1.67	2.32
K=2	1.94	1.94	1.99	1.99
K=3	1.76	1.82	1.76	1.85
K=4	1.72	1.72	1.71	1.71

One of the advantages of the *MKN* indices is that they provide more information, that is, in addition to recognizing the possible common placement of the maximum values between the time series, they also provide through their value a new measure of the similarity between the time series.

7. Conclusions

In this work an indirect method of data analysis based on the indices $\mu^{[MKN]}$ as well as a direct

method based on the measure $V^{[MKN]}$ were presented. The first method is defined in terms of coupling of maximum or minimum values of the time series considered while the second method is defined in terms of the time instances of occurrence of the maximum or minimum values. The index $\mu^{[MKN]}$ method has an advantage over the direct method, since besides the information about the similarity of the time series, as far as the times of occurrence of the maximum and minimum values are concerned, it also provides information about statistical characteristics of the data such as the Average. This work is part of an ongoing research and further exploration of the capabilities of the $\mu^{[MKN]}$ indices will be performed.

References

Ali, S.M., & Silvey, S.D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, *28*(1), 131-142.

Álvarez, M.A.N., Ibáñez, J.C., & de Mingo, C.S. (2021). Reliability Assessment of repairable systems using simple regression models. *International Journal of Mathematical, Engineering and Management Science*, *6*(1), 180-192.

Basu, A., Harris, I.R., Hjort, N.L., & Jones, M.C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, *85*(3), 549-559.

Baydogan, M.G., & Runger, G. (2016). Time series representation and similarity based on local autopatterns. *Data Mining and Knowledge Discovery*, *30*(2), 476-509.

Cressie, N., & Read, T.R. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, *46*(3), 440-464.

Csiszár, I. (1964). Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten. Magyer Tud. Akad. Mat. Kutato Int. Koezl., *8*, 85-108.

Huber-Carol, C., Balakrishnan, N., Nikulin, M., & Mesbah, M. (2002). *Goodness-of-fit tests and model validity*. Springer Science & Business Media. Birkhauser, Boston.

Iglesias, F., & Kastner, W. (2013). Analysis of similarity measures in times series clustering for the discovery of building energy patterns. *Energies*, *6*(2), 579-597.

Keogh, E., Chakrabarti, K., Pazzani, M., & Mehrotra, S. (2001). Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, *3*(3), 263-286.

Koukoumis, C., & Karagrigoriou, A. (2021). On entropy-type measures and divergences with applications in engineering, management and applied sciences. *International Journal of Mathematical, Engineering and Management Science*, *6*(3), 688-707.

Kullback, S., & Leibler, R.A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79-86.

Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003, June). A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery* (pp. 2-11). USA. https://doi.org/10.1145/882082.882086.

Makris, K. (2017). *Statistical analysis of random waves in SPAR-type and TLP-type wind turbines*. Doctoral dissertation, MSc. Thesis, National Technical University of Athens, Greece.

Makris, K. (2018). *Statistical analysis of epidemiological time series data*. Doctoral dissertation, MSc. Thesis, National Technical University of Athens, Greece.

Makris, K., Karagrigoriou, A., & Vonta, I. (2021). On divergence and dissimilarity measures for multiple time series.In: Dimotikalis, Y., Karagrigoriou, A., Parpoula, C., Skiadas, C.H. (eds) *Applied Modeling Techniques and Data Analysis 1: Computational Data Analysis Methods and Tools*. pp. 249-260. https://doi.org/10.1002/9781119821588.ch14.

Mattheou, K., Lee, S., & Karagrigoriou, A. (2009). A model selection criterion based on the BHHJ measure of divergence. *Journal of Statistical Planning and Inference*, *139*(2), 228-235.

Meselidis, C., & Karagrigoriou, A. (2020). Statistical inference for multinomial populations based on a double index family of test statistics. *Journal of Statistical Computation and Simulation*, *90*(10), 1773-1792.

Mirylenka, K., Dallachiesa, M., & Palpanas, T. (2017, June). Data series similarity using correlation-aware measures. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management* (pp. 1-12). Chicago, USA. https://doi.org/10.1145/3085504.3085515.

Morse, M.D., & Patel, J.M. (2007, June). An efficient and accurate method for evaluating time series similarity. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data* (pp. 569-580). Beijing, China.

Nakamura, T., Taki, K., Nomiya, H., Seki, K., & Uehara, K. (2013). A shape-based similarity measure for time series data with ensemble learning. *Pattern Analysis and Applications*, *16*(4), 535-548. https://doi.org/10.1007/s10044-011-0262-6.

Nguyen, M.-L.T., & Bui, T.N. (2020). The real estate market and financial stability. *International Journal of Mathematical, Engineering and Management Science, 5*(6), 1270-1283.

Serra, J., & Arcos, J.L. (2012, September). A competitive measure to assess the similarity between two time series. *In International Conference on Case-Based Reasoning* (pp. 414-427). Springer, Berlin, Heidelberg.

Toma, A. (2008). Minimum Hellinger distance estimators for multivariate distributions from the Johnson system. *Journal of Statistical Planning and Inference*, *138*(3), 803-816.

Toma, A. (2009). Optimal robust M-estimators using divergences. *Statistics & Probability Letters*, *79*(1), 1-5.

Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(2), 281-294.