Research Article

Consensus Patterns of a Set of Time Series via a Wavelet-Based Temporal Localization: Emphasizing the Utility over Point-Wise Averaging and Averaging under Dynamic Time Warping

Chekhaprabha Priyadarshanee Waduge,[1] Naleen Chaminda Ganegoda [ID],[2] Darshana Chitraka Wickramarachchi,[3] and Ravindra Shanthakumar Lokupitiya[3]

[1]*Faculty of Graduate Studies, University of Sri Jayewardenepura, Nugegoda, Sri Lanka*
[2]*Department of Mathematics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka*
[3]*Department of Statistics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka*

Correspondence should be addressed to Naleen Chaminda Ganegoda; naleen@sjp.ac.lk

Summarizing or averaging a sequential data set (i.e., a set of time series) can be comprehensively approached as a result of sophisticated computational tools. Averaging under Dynamic Time Warping (DTW) is one such tool that captures consensus patterns. DTW acts as a similarity measure between time series, and subsequently, an averaging method must be executed upon the behaviour of DTW. However, averaging under DTW somewhat neglects temporal aspect since it is on the search of similar appearances rather than stagnating on corresponding time-points. On the contrary, the mean series carrying point-wise averages provides only a weak consensus pattern as it may over-smooth important temporal variations. As a compromise, a pool of consensus series termed Ultimate Tamed Series (UTS) is studied here that adheres to temporal decomposition supported by the discrete Haar wavelet. We claim that UTS summarizes localized patterns, which would not be reachable via the series under DTW or the mean series. Neighbourhood of localization can be altered as a user can customize different levels of decomposition. In validation, comparisons are carried out with the series under DTW and the mean series via Euclidean distance and the distance resulted by DTW itself. Two sequential data sets are selected for this purpose from a standard repository.

1. Introduction

Classical topic of summarizing and classifying a sequential data set (i.e., a set of time series) has recently gained a new outlook because of data mining and machine learning procedures [1–3]. Investigating common pattern (or shape) attributes (e.g., monotonic, oscillatory, and periodic) of a set of time series is ever so attainable, but still challenging even with advanced computer technology. The basic reason is that data are ordered unlike in a usual data set. In practice, time series arise in many scenarios such as economic and financial indices, weather forecasting, disease transmission, census data, market surveys, quality control, and inventory studies [4]. Data availability is also unprecedented as digital platforms facilitate gathering, storing, and processing data in profound

ways [5]. Therefore, we always need effective and efficient techniques to extract valuable information from sequential data [6–8].

Finding a consensus or representative time series for a set of time series (i.e., a pattern similar to many series) is often found in a handful of phenomena. Daily rainfall of a region reported by several substations, fluctuation of a financial index in different stock markets and banks, daily sales of goods in different outlets of a supermarket chain, speech and gesture recognition, and genetic sequencing of an organism are among such phenomena [9–11]. The domain of a time series does not necessarily stand for time; for instance, it refers to the order of nucleotides in deoxyribonucleic acid (DNA). Pattern recognition not only in series-wise but also as a whole set of series is a key concern. In this work, we

are interested in the latter. A measure called Dynamic Time Warping (DTW) is frequently used to extract similarity between time series. Then, one can design an averaging method based on the behaviour of DTW that gives a consensus pattern (i.e., a pattern aligned with many series in the set) [9, 12]. DTW has been developed as a nonlinear mapping between time series to map similar features, neglecting temporal distortions [13]. Thus, a resulting series under DTW captures attributes of pattern rather than attributes of point-wise central tendency. We name the "series under DTW" as "Dynamic Time Warping Series (DTWS)." This is similar to the naming of "mean series" giving the "series under point-wise averaging." The main drawback of DTWS is its limitations on representing local patterns (i.e., corresponding patterns upon temporal proximity). DTW approaches two time series at a time structuring a distance matrix and backtracks in the direction of minimum cumulative distances to achieve so called warping path [14, 15]. Formulation of this output (DTWS) for a set of time series via DTW is explained in detail in Section 2. Improved versions of DTW are available in the literature such as NLAAF (Nonlinear Alignment and Averaging Filters), PSA (Prioritized Shape averaging), and RSA (Re-sampling Shape Averaging) which address the limitations of the algorithm in different aspects. Further, there are techniques such as Contrast Enhanced Dynamic Time Warping (CEDTW) and DTW Barycenter Averaging (DBA) which have greater accuracy and higher performance [15]. However, for this study, we stick to the basic DTW algorithm leaving improved versions into future work. This study is the first of this kind that compares UTS with consensus series under DTW.

In contrast to DTWS, the series produced by point-wise averages (the mean series) represents central tendency, but hardly absorbs pattern attributes. The mean $\bar{x}$ for any set of data points x_i; $i = 1, 2, \cdots n$ yields the minimum squared deviation $\sum_{i=1}^{n} (x_i - \bar{x})^2$ with the data set. Therefore, for any set of time series, the mean series assures the minimum point-wise squared deviation. Consequently, the mean series may be an over-smoothed representative, restricting to impart patterns of data series.

To address the limitations of the mean series, an alternative technique called Ultimate Tamed Series (UTS) has been proposed [16–18]. On top of DTWS, it acquires local patterns with customized temporal windows. UTS is based on discrete Haar wavelet transform. In literature, Haar wavelet is recorded as the earliest wavelet approach and also the simplest in structure [19–21]. In formulating UTS, a variant of Haar matrix is implemented transforming a time series into two series, termed as scaling coefficients (father wavelets) and wavelet coefficients (mother wavelets). These two wavelet types are worked out in a specific way called *taming* [22]. Preliminary details of wavelets and taming are recalled in Section 3. This paper extends UTS technique into higher levels of transform aligning with multiterminal binary decision diagrams available to Haar wavelet [19]. It is designed upon father wavelets that behaves as level-wise representatives of the original time series.

Similar to DTWS, UTS is also operated with two time series at a time. After operating for two time series (chosen from a given set of data series), the resultant series and the next data series are operated together and the same process is recursively carried out. Since DTWS and UTS do not accommodate commutative and associative properties, a lot of consensus series can be determined. Note that DTW is a pseudo-metric which preserves the commutative property giving the same warping distance irrespective of the order of concern for two time series. However, DTWS for two time series may vary with its order of concern due to the backtracking process in generating the consensus series. In the backtracking process, different warping paths can be generated when there is a tie between minimum neighbouring elements in the distance matrix. This scenario is explained in [23], and the DTWS formulation is described in Section 2. Thus, DTW and UTS depend on hierarchies of taking time series into the respective computational procedures [22, 23]. In this comparative study, two specific hierarchies are considered catering the highest and the lowest variations among consecutive data points in each series.

Three consensus approaches, the mean series, DTWS, and UTS, establish pattern attributes in different extents. Identifying the capability of representing patterns is worth investigating which is the main task of this paper. We test these measures for OSULeaf data set and Yoga data set available in UEA & UCR Time Series Classification Repository [24]. Section 5 contains details of data processing used to implement consensus approaches mentioned above. In the sequel, we present basic routines of DTW (Section 2) and UTS (Section 3) along with comparison tools (Section 4). Furthermore, Section 3.3 presents theoretical support for the claims tested in Section 6.

2. Dynamic Time Warping Series (DTWS)

DTW algorithm is based on a distance measure called Levenshtein distance [13, 25]. For two time series of equal length N denoted in sequential terminology $S^{(1)} = \{s_n^{(1)}\}_{n=1}^{N}$ and $S^{(2)} = \{s_n^{(2)}\}_{n=1}^{N}$, the Levenshtein string metric D is recursively obtained by

$$
D\left(S_p^{(1)}, S_q^{(2)}\right) = d\left(s_p^{(1)}, s_q^{(2)}\right) + \min \begin{cases} D\left(S_{p-1}^{(1)}, S_{q-1}^{(2)}\right), \\ D\left(S_p^{(1)}, S_{q-1}^{(2)}\right), \\ D\left(S_{p-1}^{(1)}, S_q^{(2)}\right), \end{cases} \tag{1}
$$

where $d(s_p^{(1)}, s_q^{(2)})$ is the distance between $s_p^{(1)}$ and $s_q^{(2)}$ given by $d(s_p^{(1)}, s_q^{(2)}) = |s_p^{(1)} - s_q^{(2)}|$. Here, $S_p^{(1)} = \{s_n^{(1)}\}_{n=1}^{p}$ and $S_q^{(2)} = \{s_m^{(2)}\}_{m=1}^{q}$ are subsequences of $S^{(1)}$ and $S^{(2)}$, respectively (i.e., $1 \le p \mid q \le N$). Using the metric D, a matrix $D_{N \times N}$ is formed where the entry (p, q) is determined by the value $D(S_p^{(1)}, S_q^{(2)})$. Next, a string called the optimal warping path $W = \{w_r\}_{r=1}^{R}$, where $w_r = [a, b]$ for some $a, b \in \mathbb{N}$, is constructed by backtracking the entries of the matrix $D_{N \times N}$ in the direction of minimum cumulative distance starting from the entry (N, N) to the entry $(1,1)$ [26]. Note that, for each a and b, there is a corresponding

entry from series $S^{(1)}$ and $S^{(2)}$ (say($s_a^{(1)}, s_b^{(2)}$)). Then, a consensus series of $S^{(1)}$ and $S^{(2)}$ named as Dynamic Time Warping Series (DTWS) is obtained by averaging the pairs ($s_a^{(1)}, s_b^{(2)}$) (i.e., ($s_a^{(1)} + s_b^{(2)}$)/2) corresponding to each selected entry in the path W. This is the process we refer by "averaging under DTW." This consensus measure need not to have similar length to the original series, and also, the original series need not to be similar in length. D brings a quantification of similarity between two series. Thus, we further use overall similarity given by $D(S_N^{(1)}, S_N^{(2)})$ to design a comparison tool later in Section 4. Here, lesser D shows more similarity.

For more than two time series, DTWS is computed recursively considering the DTWS of the first two series with the third series and that resultant DTWS with the fourth series and so on [15]. In this process, pairs ($s_a^{(1)}, s_b^{(2)}$) should be averaged in a weighted manner by considering number of series infused in each appearance of DTWS. For a pair ($\bar{s}_a^{(j)}$, $s_b^{(j)}$) where $s_b^{(j)}$ is the bth entry of the jth series according to the hierarchy in the set and $\bar{s}_a^{j}$ is the ath entry of the jth averaged (according to DTW) series, the entries are paired as ($j\bar{s}_a^{(j-1)} + s_b^{(j+1)}$)/($j + 1$) [15]. As mentioned, hierarchy of arranging series into the warping process comes into play here giving different DTWS for different hierarchies. Since the final output in the warping process is substantially longer than the original lengths of the series, the uniform scaling is applied to achieve a consensus series of a length similar to that of data series [14, 27]. A series $W = \{w_r\}_{r=1}^R$ can be scaled to a sequence $Z = \{z_n\}_{n=1}^N$ of required length N, with $z_n = w_{\lceil n.R/N \rceil}$. Then, the new series Z yields a scaled DTWS. This constraint on length forms a common platform for upcoming comparisons. We execute scaling after all the series are taken into warping process allowing temporal distortions to remain flexible.

3. Ultimate Tamed Series (UTS)

In this section, we present the structure of UTS, and next two sections direct its implementation as a comparable tool to the mean series and DTWS. We start with two time series where the procedure is to be recursively used upon any number of time series. First, obtaining father and mother wavelets is described in Section 3.1 as designed via Haar wavelet. Next, the taming procedure is structured in Section 3.2 emphasizing extended work on higher levels. In the final implementation, we should decide a hierarchy of taking time series into the taming procedure. Section 3.5 sets that hierarchy, and Section 5 contains some illustrations of it in association with data. In this study, we execute computer programs designed using the MATLAB software for all the implementations of UTS and DTWS, whereas any other mathematical software can also be used.

3.1. The Discrete Haar Wavelet as a Weak Form: Father and Mother Wavelets. In mathematical analysis, weak forms are applicable when it is difficult to deal with strong forms. If we treat pattern recognition of time series as a strong form,

recognizing patterns via wavelet transform resembles a weak form. Then, different transforms lead to different outlooks of statistics of time series. In general, implementation of weak forms may have their own limitations as we see in Finite Element Methods that require careful choice of weight functions and trial solutions [28, 29]. UTS can also be considered as one that arises from a weak form.

The initial step of UTS is to partition time series into two-tuples (two-component vectors). For instance, for a time series $\{s_n\}$, the discrete Haar transform works on two-tuples: $(s_1, s_2), (s_3, s_4), (s_5, s_6), \cdots$, when we start by s_1. This coupling leads to partial associations since two-tuples generated by shifting the start by one data point (i.e., $(s_2, s_3), (s_4, s_5)$, $(s_6, s_7), \cdots$) are not directly approached. In a comprehensive analysis, both these couplings can be trialed to curtail missing associations in one coupling. However, the first data point s_1 is not considered in the shifted case, and inclusion of the last data point is based on whether the series has even or odd number of points. This would not be a major drawback when we deal with lengthy time series in practice. Then, the extreme data points do not hinder much the overall context of the series.

Hereafter, we name $(s_1, s_2), (s_3, s_4), (s_5, s_6), \cdots$ as 1st coupling and $(s_2, s_3), (s_4, s_5), (s_6, s_7), \cdots$ as 2nd coupling. Furthermore, we proceed with time series containing odd number of data points: $\{s_n\}_{n=1}^{2N+1}$ ($N \in \mathbb{N}$), where we ignore the first data point in 2nd coupling and the last data-point in 1st coupling to read as

$$1^{\text{st}} \text{ coupling} : \{(S_{2n-1}, S_{2n})\}_{n=1}^N \text{ and } 2^{\text{nd}} \text{ coupling} : \{(S_{2n}, S_{2n+1})\}_{n=1}^N. \tag{2}$$

Next, father and mother wavelets are determined by right-multiplying the two-tuples by variant Haar matrix $H_v = 1/2 \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ as in Definition 1. Here, the classical Haar matrix $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is scaled by 1/2 to cater average of data points in father wavelets. It facilitates interpretations in a way that father wavelets act as representatives of the original time series in further transform levels.

Definition 1 (H_v-transform). Right-multiplying the two-tuples of $\{s_n\}_{n=1}^{2N+1}$ by $H_v = 1/2 \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ gives

1st *coupling—father wavelets*: $\{f_n = (s_{2n-1} + s_{2n})/2\}_{n=1}^N$.
1st *coupling—mother wavelets*: $\{m_n = (s_{2n-1} - s_{2n})/2\}_{n=1}^N$.
Alterations to the 2nd coupling comes as
2nd *coupling—father wavelets*: $\{f_n = (s_{2n} + s_{2n+1})/2\}_{n=1}^N$.
2nd *coupling—mother wavelets*: $\{m_n = s_{2n} - s_{2n+1}/2\}_{n=1}^N$.

The inverse transform of the above H_v-transform is structured into Definition 2.

Definition 2 (inverse H_ν-transform). Data points of $\{s_n\}_{n=1}^{2N+1}$ can be obtained back by the following inverse operations.

   1st *coupling*: $s_{2n-1} = f_n + m_n$ and $s_{2n} = f_n - m_n$; $n = 1, 2,$ $\cdots, N$.

   2nd *coupling*: $s_{2n} = f_n + m_n$ and $s_{2n+1} = f_n - m_n$; $n = 1, 2,$ $\cdots, N$.

   Note that the inverse of H_ν is given by $H_\nu^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, where right-multiplying the two-tuples (f_n, m_n) provides back the time series $\{s_n\}_{n=1}^{2N+1}$.

Remark 3. The H_ν-transform in Definition 1 is continued into further levels upon father wavelet series. We address this in Section 3.2.1 with required work of UTS.

3.2. Taming Procedure.

So called taming is the key procedure that articulates UTS. It is nurtured in the transformed series obtained via H_ν-transform. Here, we first present the way of taming two series and next extend it into more series and more transform levels. Only 1st coupling is demonstrated here, where the same approach is applicable to 2nd coupling.

Step 1 (father and mother wavelets). Two time series $S^{(1)} = \{s_n^{(1)}\}_{n=1}^{2N+1}$ and $S^{(2)} = \{s_n^{(2)}\}_{n=1}^{2N+1}$ are partitioned into two-tuples of 1st coupling as $\{(s_{2n-1}^{(1)}, s_{2n}^{(1)})\}_{n=1}^{N}$ and $\{(s_{2n-1}^{(2)}, s_{2n}^{(2)})\}_{n=1}^{N}$, respectively. By H_ν-transform, we have

   father wavelets of series $S^{(i)}$; $i = 1, 2$ as $f_n^{(i)} = (s_{(2n-1)}^{(i)} + s_{2n}^{(i)})/2$ and mother wavelets of series $S^{(i)}$; $i = 1, 2$ as $m_n^{(i)} = (s_{2n-1}^{(i)} - s_{2n}^{(i)})/2$.
   Here, $n = 1, 2, \cdots N$.

Step 2 (taming criteria). We define the differences Df_n and Dm_n for the father and mother wavelets of two time series as $Df_n = |f_n^{(1)} - f_n^{(2)}|$ and $Dm_n = |m_n^{(1)} - m_n^{(2)}|$.

The taming procedure is performed via averaging either father wavelet or mother wavelet subject to the highest of Df_n and Dm_n as follows (taming criteria). Superscript (new) stands for new wavelets produced by the taming criteria.

   (i) If $Df_n > Dm_n$, then $f_n^{(new)} = (f_n^{(1)} + f_n^{(2)})/2$ and $m_n^{(new)} = m_n^{(1.)}$

   (ii) If $Df_n < Dm_n$, then $f_n^{(new)} = f_n^{(1)}$ and $m_n^{(new)} = (m_n^{(1)} + m_n^{(2)})/2$

   (iii) If $Df_n = Dm_n$, then $f_n^{(new)} = (f_n^{(1)} + f_n^{(2)})/2$ and $m_n^{(new)} = (m_n^{(1)} + m_n^{(2)})/2$

Remark 4. Geometric motivation for the above procedure has been addressed in [16–18, 22, 30]. To recall along with Figure 1, let L_1 be the line segment joining the data points

$s_{2n-1}^{(1)}$ and $s_{2n}^{(1)}$, and L_2 be the line segment joining the data points $s_{2n-1}^{(2)}$ and $s_{2n}^{(2)}$ in a time series plot. Then, L_1 and L_2

   (i) do not intersect if and only if $Df_n > Dm_n$

   (ii) intersect at a data point or coincide if and only if $Df_n = Dm_n$

   (iii) intersect at a point other than to data points if and only if $Df_n < Dm_n$

In summary, we tend to tame by father wavelets if the difference of them are higher (i.e., line segments are apart) and otherwise, we tame by mother wavelets. Case (ii) compromises qualitatively moderate situation.

Step 3 (tamed series). A tamed series $T^{(1)} = \{t_n^{(1)}\}_{n=1}^{2N}$ is obtained by the inverse H_ν-transform $t_{2n-1}^{(1)} = f_n^{(new)} + m_n^{(new)}$ and $t_{2n}^{(1)} = f_n^{(new)} - m_n^{(new)}$. We denote $T^{(1)} = S^{(1)} \oplus S^{(2)}$ showing $T^{(1)}$ as the end result of taming $S^{(1)}$ and $S^{(2)}$. Thus, $T^{(1)}$ is considered as a consensus series.

Remark 5. Note that $T^{(1)}$ is the same as the mean series of $S^{(1)}$ and $S^{(2)}$, if both father and mother wavelets are averaged in Step 2 instead of one. This coins the word taming as it sounds a lesser smoothing effect compared to usual averaging. Geometric illustration of the taming criteria is available in [17, 30].

Remark 6. Operation $\oplus$ is noncommutative and nonassociative since we bias on first series in deciding new wavelet that is not averaged in (i) and (ii) of Step 2.

We see these as optimistic properties that allow a hierarchical procedure into a consensus measure, which is not available for the mean series. A user can customize the hierarchy according to a required descriptive analysis.

Step 4 (UTS). To find a consensus series for more than two time series, the taming procedure must be extended by taming new series with already tamed series. Figure 2 indicates the flow of this recursive procedure for a set of k time series $\{S^{(1)}, S^{(2)}, \cdots S^{(k)}\}$. Here, $T^{(k-1)}$ is considered as an end result giving a consensus series for all k series and named it as Ultimate Tamed Series (UTS), i.e., $T^{(k-1)} = ((S_1 \oplus S_2) \oplus S_3 \cdots) \oplus S_k$. At the rth taming step, averaging must be done in a weighted manner as $f_n^{(new)} = (rf_n^{(1)} + f_n^{(2)})/(r + 1)$ instead of $f_n^{(new)} = (f_n^{(1)} + f_n^{(2)})/2$ and $m_n^{(new)} = rm_n^{(1)} + m_n^{(2)}/(r + 1)$ instead of $m_n^{(new)} = (m_n^{(1)} + m_n^{(2)})/2$, where now the superscript (1) stands for already tamed series and superscript (2) stands for the next series newly entering to the taming process.

The entries of the above UTS acquire local trend only with an adjacent point partitioned in 1st coupling. Indeed, sign of mother wavelet illustrates increasing/decreasing trends as it caters a difference. In order to capture pattern attributes by more data points, taming procedure should be applied to further transform levels as described in Section
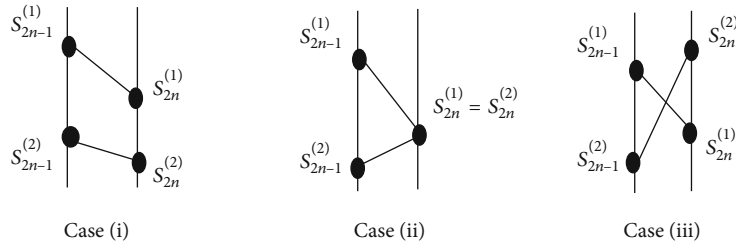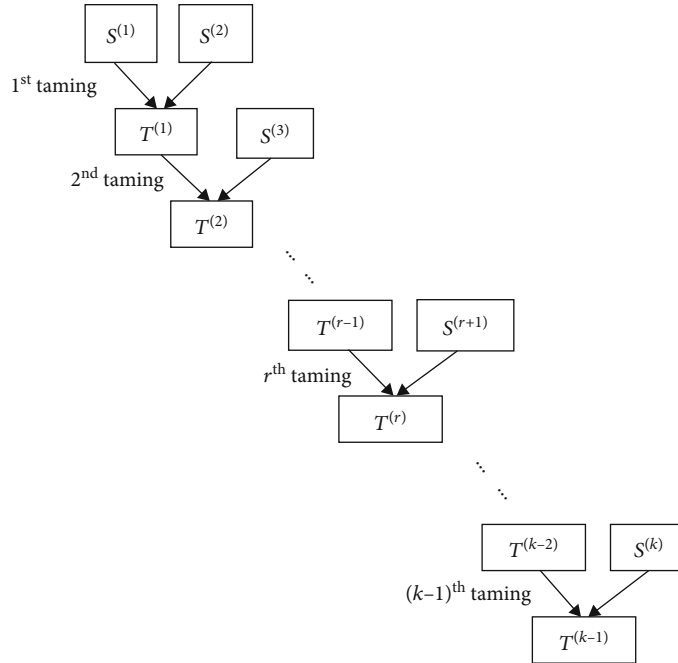
Figure 1: Illustrations of Remark 4.

Figure 2: Taming k number of time series.

3.2.1. As we suggested earlier, repeating the work for 2nd coupling yields a more comprehensive analysis.

3.2.1. Higher Level Taming. After carrying out Step 1 in Section 3.2, we continue the same for resultant father wavelets. Immediate level resulting from Step 1 is termed as level 1. Next, father and mother wavelets of level 2 are taken by coupling father wavelets of level 1 and carrying out H_ν-transform for those two-tuples. In general, father and mother wavelets of level $h = 2, 3, \cdots$ are obtained by carrying out Step 1 for father wavelets of level $h - 1$. Note that still we proceed with the two time series $S^{(1)} = \{s_n^{(1)}\}_{n=1}^{2N+1}$ and $S^{(2)} = \{s_n^{(2)}\}_{n=1}^{2N+1}$ separately. Figure 3 depicts how a time series $\{s_n\}$ with eight data points is transformed into four levels as a binary diagram.

Father wavelets in level h rescale original time series by averaging 2^h adjacent points (data points with temporal proximity), while mother wavelets associate with the difference of such rescaled father wavelets in the previous level. The averaging occurs due to the scaling by 1/2 in H_ν-transform, giving a representative role (i.e., average) to father wavelets. We later illustrate our results with these 2^h-tuple windows.

Now, we can move into higher level taming after transforming two time series $S^{(1)} = \{s_n^{(1)}\}_{n=1}^{2N+1}$ and $S^{(2)} = \{s_n^{(2)}\}_{n=1}^{2N+1}$ up to required levels. Taming for higher levels $h = 2, 3, \cdots$ is carried out via Step 2 in Section 3.2 for father and mother wavelets of respective level. Suppose we select a particular h as the taming level that we implement Step 2. Next, the inverse procedure similar to Step 3 is carried out in level h to obtain new father wavelets of level $h - 1$. Note that here, we have only the father wavelets by inverse transform in Step 3. Moreover, to make a complete taming, we must continue the inverse up to the level of original series. Then, we need corresponding mother wavelets, and for that, we propose two ways as follows.

(i) *Inverse with Average Mother (AM-Inverse).* For all the levels from level $h - 1$ to level 1, we use respective averages of mother wavelets corresponding to two series $S^{(1)}$ and $S^{(2)}$.

(ii) *Inverse with Previous Mother (PM-Inverse).* For all the levels from level $h - 1$ to level 1, we use mother wavelets corresponding to the first series $S^{(1)}$.
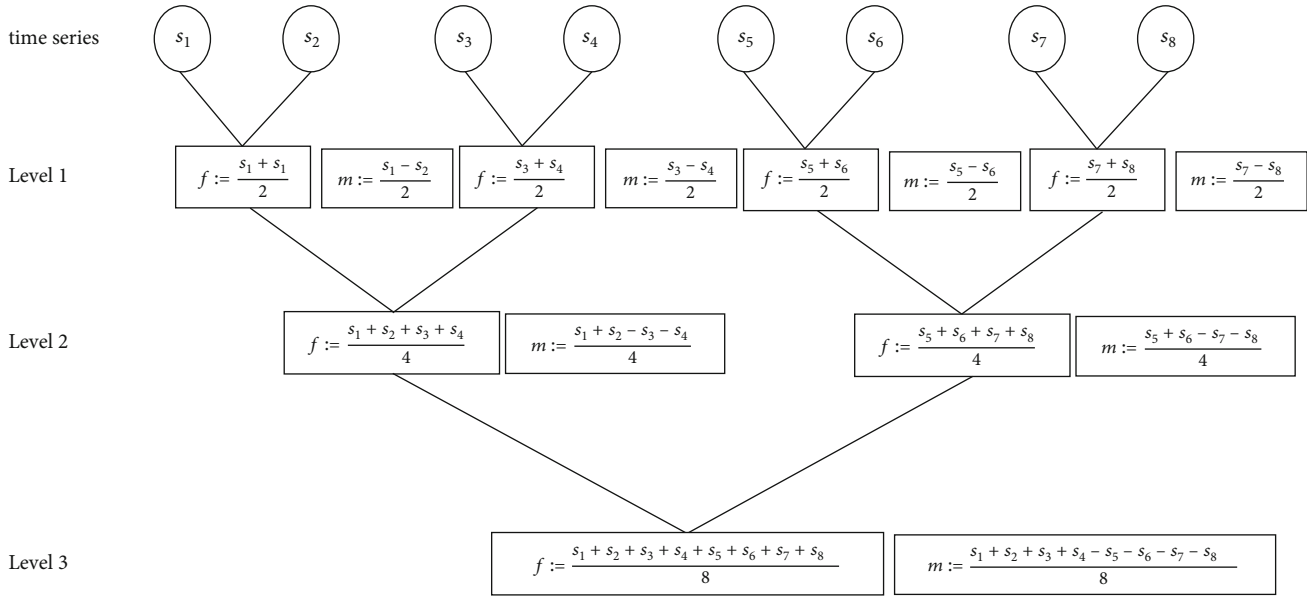
Figure 3: Illustration of H_ν-transform for different levels.

For a set of k time series, we execute the same procedure given in Step 4 (i.e., taming already tamed series with the next series adhering to weighted averages). We denote UTS obtained by taming at level h as UTS-L_h, and to distinguish under AM-inverse and PM-inverse, we further denote it as UTS-L_h(AM) and UTS-L_h(PM), respectively. UTS-L_1 stands for the same UTS formulated in Section 3.2. In this manner, we can generate a pool of consensus series giving options for customizing.

Remark 7. UTS-L_h(PM) is biased more to the pattern of earlier series than the latter series in the hierarchy (order) of taking time series into the taming procedure. It is because of taking new mother wavelets as the mother wavelets of the first series (or already tamed series). UTS-L_h(AM) is unbiased in that sense.

We further discuss about hierarchy of taming in the Section 3.5.

3.3. Well-Posedness of H_ν-Transform and Its Inverse. This subsection presents well-posedness of H_ν-transform and its inverse enclosing existence, uniqueness, and continuous change of the concerned series. We assure these properties for a given set of data series subject to a given coupling method (1st or 2nd) and for a given mother wavelet for inverse (AM-inverse or PM-inverse). Both H_ν-transform and its inverse are encountered here. Without loss of generality, we articulate propositions between level 1 and level 2 that can be generalized between level h and level $h + 1$.

Proposition 8 (existence). *There exist father and mother wavelets of level 2 corresponding to two-tuples of father wavelets of level 1 (via H_ν-transform) and vice-versa (via inverse H_ν-transform).*

Proof. This result is straightforward by Definition 1 applied to the series of father wavelets of level 1 and Definition 2 applied to the inverse operations. □

Proposition 9 (uniqueness).

(i) *Father and mother wavelets of level 2 are uniquely determined by father wavelets of level 1 via H_ν-transform*

(ii) *Conversely, father wavelets of level 1 are uniquely determined by father and mother wavelets of level 2 via inverse H_ν-transform*

Proof. It is enough to show the result for one two-tuple as the same can be repeated throughout. Suppose (f_1, f_2) is a two-tuple of father wavelets of level 1 corresponds to father wavelet F_1 and mother wavelet M_1 of level 2.

Then, we have the following two linear systems by adjusting right-multiplying into left-multiplying appropriately.

$$H_\nu \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} F_1 \\ M_1 \end{pmatrix}, \text{where } H_\nu = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$H_\nu^{-1} \begin{pmatrix} F_1 \\ M_1 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \text{where } H_\nu^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{3}$$

Since the determinants of H_ν and H_ν^{-1} are nonzero, both the systems have unique solutions guaranteeing the result. □

Proposition 10 (continuous dependency).

(i) *Father and mother wavelets of level 2 continuously depend on the changes of father wavelets of level 1*
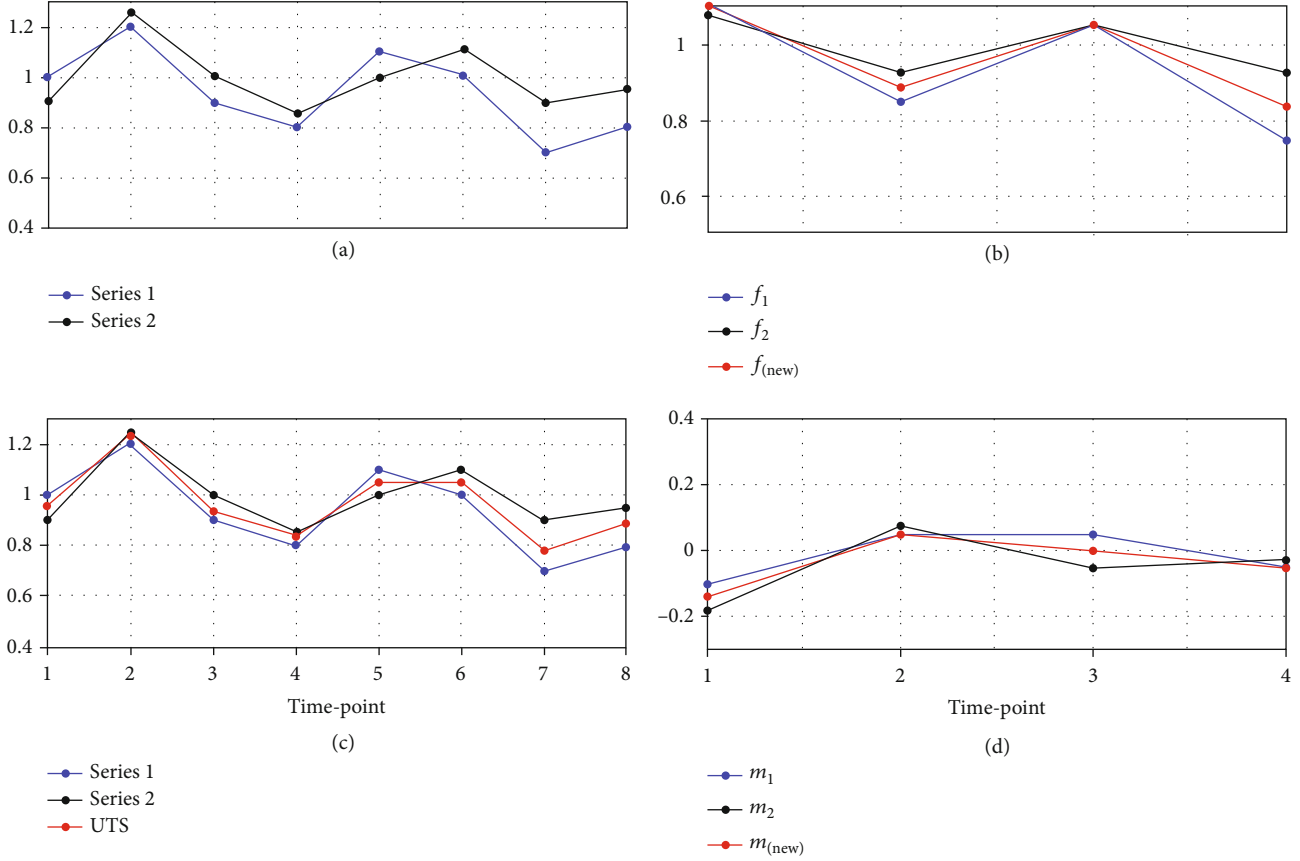
FIGURE 4: Qualitative illustration of taming two close time series (four two-tuples are shown): f_1 and f_2—father wavelets of series 1 and series 2, $f_{(\text{new})}$—new father wavelet, m_1 and m_2—mother wavelets of series 1 and series 2, and $m_{(\text{new})}$—new mother wavelet.

(ii) *Conversely, father wavelets of level 1 continuously depend on the changes of father and mother wavelets of level 2*

Proof. Here also, it is enough to show the result for one two-tuple. Suppose (f_1, f_2) is a two-tuple of father wavelets of level 1 corresponds to father wavelet F_1 and mother wavelet M_1 of level 2. If we change the two-tuple into $(f_1^{(\text{new})}, f_2^{(\text{new})})$, then resulting wavelets of level 2 are changed (say F_1 into $F_1^{(\text{new})}$ and M_1 into $M_1^{(\text{new})}$). Observe that

$$
\begin{aligned}
\left| F_1^{(\text{new})} - F_1 \right| &= \left| \frac{f_1^{(\text{new})} + f_2^{(\text{new})}}{2} - \frac{f_1 + f_2}{2} \right| \\
&= \left| \frac{f_1^{(\text{new})} - f_1}{2} - \frac{f_2^{(\text{new})} - f_2}{2} \right| \\
&\leq \frac{1}{2} \left| f_1^{(\text{new})} - f_1 \right| + \frac{1}{2} \left| f_2^{(\text{new})} - f_2 \right|.
\end{aligned} \tag{4}
$$

For a given $\varepsilon > 0$, there exists $\delta = 2\varepsilon > 0$ such that $|f_1^{(\text{new})} - f_1| < \delta$ and $|f_2^{(\text{new})} - f_2| < \delta \Longrightarrow |F_1^{(\text{new})} - F_1| < \varepsilon$. Similarly, we can show that, $|f_1^{(\text{new})} - f_1| < \delta$ and $|f_2^{(\text{new})} - f_2| < \delta \Longrightarrow |$

$M_1^{(\text{new})} - M_1| < \varepsilon$, establishing Proposition 10 (i).

For (ii), observe that

$$
\begin{aligned}
\left| f_1^{(\text{new})} - f_1 \right| &= \left| \left(F_1^{(\text{new})} + M_1^{(\text{new})} \right) - (F_1 + M_1) \right| \\
&= \left| \left(F_1^{(\text{new})} - F_1 \right) + \left(M_1^{(new)} - M_1 \right) \right| \\
&\leq \left| F_1^{(\text{new})} - F_1 \right| + \left| M_1^{(\text{new})} - M_1 \right|.
\end{aligned} \tag{5}
$$

For a given $\varepsilon > 0$, there exists $\delta = \varepsilon/2 > 0$ such that $|F_1^{(\text{new})} - F_1| < \delta$ and $|M_1^{(\text{new})} - M_1| < \delta \Longrightarrow |f_1^{(\text{new})} - f_1| < \varepsilon$.

Similarly, we can show that, $|F_1^{(\text{new})} - F_1| < \delta$ and $|M_1^{(\text{new})} - M_1| < \delta \Longrightarrow |f_2^{(\text{new})} - f_2| < \varepsilon$, establishing Proposition 10 (ii). □

Remark 11. Above propositions can be advanced by taking level 1 as original time series and level 2 as level 1. Then, data points of time series (or entries of already tamed series) replacing father wavelets of level 1 and wavelets of level 2 are replaced by those of level 1.

3.4. Qualitative Illustration. As guaranteed by the Proposition 10 (i) in the view of Remark 11, wavelets of two close time series do not vary much. Here, by the term "close," we
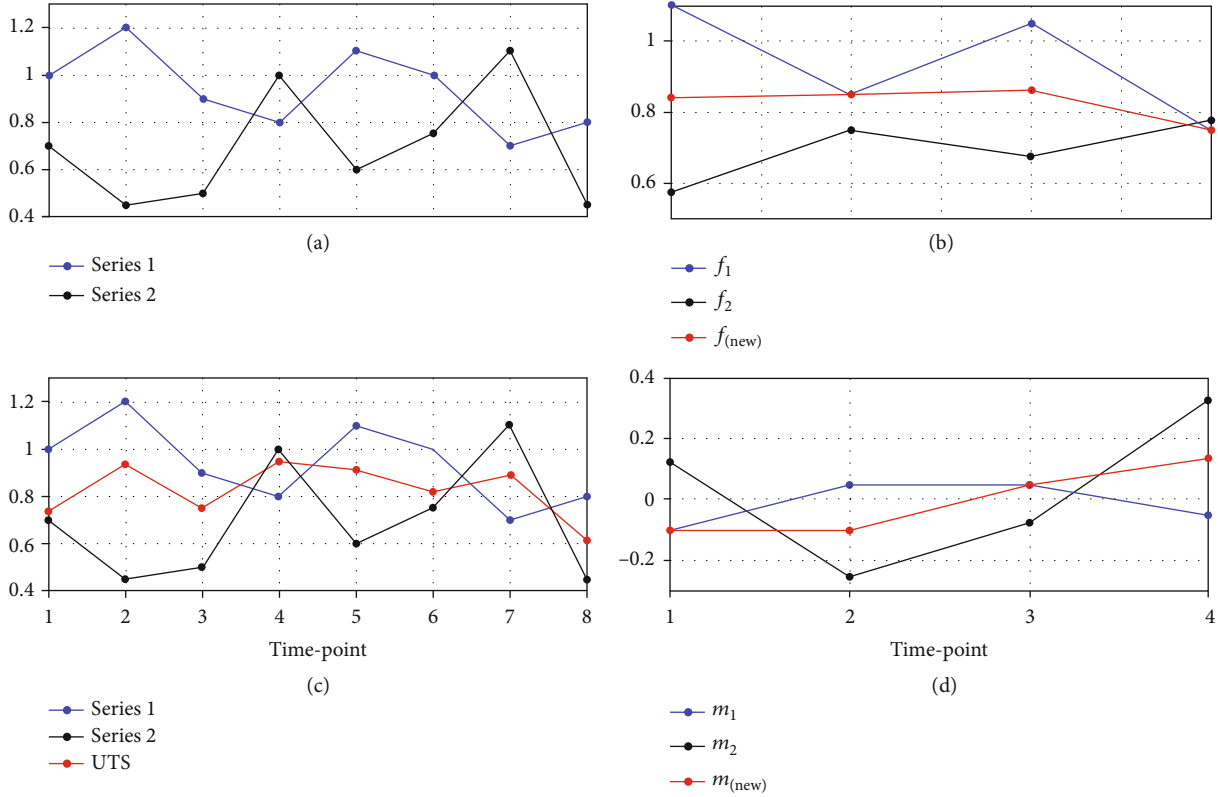
FIGURE 5: Qualitative illustration of taming two time series not close to each other (four two-tuples are shown): f_1 and f_2—father wavelets of series 1 and series 2, $f_{(\text{new})}$—new father wavelet, m_1 and m_2—mother wavelets of series 1 and series 2, and $m_{(\text{new})}$—new mother wavelet.

mean positioning of corresponding data points is close (i.e., one can say that one series is obtained by making small changes to the other series). Figure 4(a) illustrates two time series that are close to each other compared to the two time series in Figure 5(a). Corresponding wavelets are shown in Figures 4(b), 4(d), 5(b), and 5(d) that illustrate small changes in data points lead to small changes in wavelets and vice versa.

When wavelets are close, resulting new wavelets by the taming criteria are also close to corresponding wavelets. Then, the resultant UTS is also close to two time series verifying Proposition 10 (ii). This is illustrated in Figures 4(c) and 5(c).

3.5. Hierarchy for DTWS and UTS. As seen in formulation, both DTWS and UTS vary according to the hierarchy (order) of taking time series into the taming procedure. In our trials, we focus on two specific hierarchies based on variations of consecutive data points. As an aggregated measure, we consider sum of squared differences of consecutive data points of each series. For a set of k time series $\{S^{(1)}, S^{(2)}, \cdots, S^{(k)}\}$ of length $2N + 1$, these sums are given by $A^{(i)} = \sum_{n=1}^{2N} \left(s_{n+1}^{(i)} - s_n^{(i)}\right)^2$; $i = 1, 2, \cdots, k$. Now the set of time series is hierarchically arranged according to the ascending and descending order of $A^{(i)}$ values. Thus, we look forward to immediate closeness/variation of data points rather than their magnitude. $A^{(i)}$ of a particular series $S^{(i)}$ does not depend on other series. In addition, it caters every consecutive variation without restricting to two-tuples in coupling technique. There-

fore, hierarchy does not depend on coupling technique. These characteristics motivate deciding hierarchy via $A^{(i)}$.

4. Tools for Comparison

The main task of this paper is to compare UTS-L_h, DTWS, and the mean series in the sense of grasping consensus patterns of a set of time series. Meanwhile, one can see how far central tendency can be reflected by UTS-L_h measures. By central tendency, we mean how each entry in a consensus series represents the center of corresponding data points. We introduce two comparison tools: Euclidean distance and DTW distance between series are utilized. They are the well-known distance measures used in similar work [9, 12, 31].

4.1. Euclidean Distance. Point-wise squared deviation is taken in the Euclidean sense to estimate central tendency acquired by consensus series. The Euclidean distance between a consensus series $C = \{c_n\}_{n=1}^{2N}$ and each time series $S^{(i)} = \{s_n^{(i)}\}_{n=1}^{2N}$; $i = 1, 2, \cdots, k$ (say $P_i(C)$) is given by

$$P_i(C) = \sqrt{\sum_{n=1}^{N} \left(s_n^{(i)} - c_n\right)^2}. \tag{6}$$

Next a representative of $P_i(C)$; $i = 1, 2, \cdots, k$ should be determined (say $P(C)$) for comparisons between consensus series. Thus, $P(C)$ represents overall deviation of time series
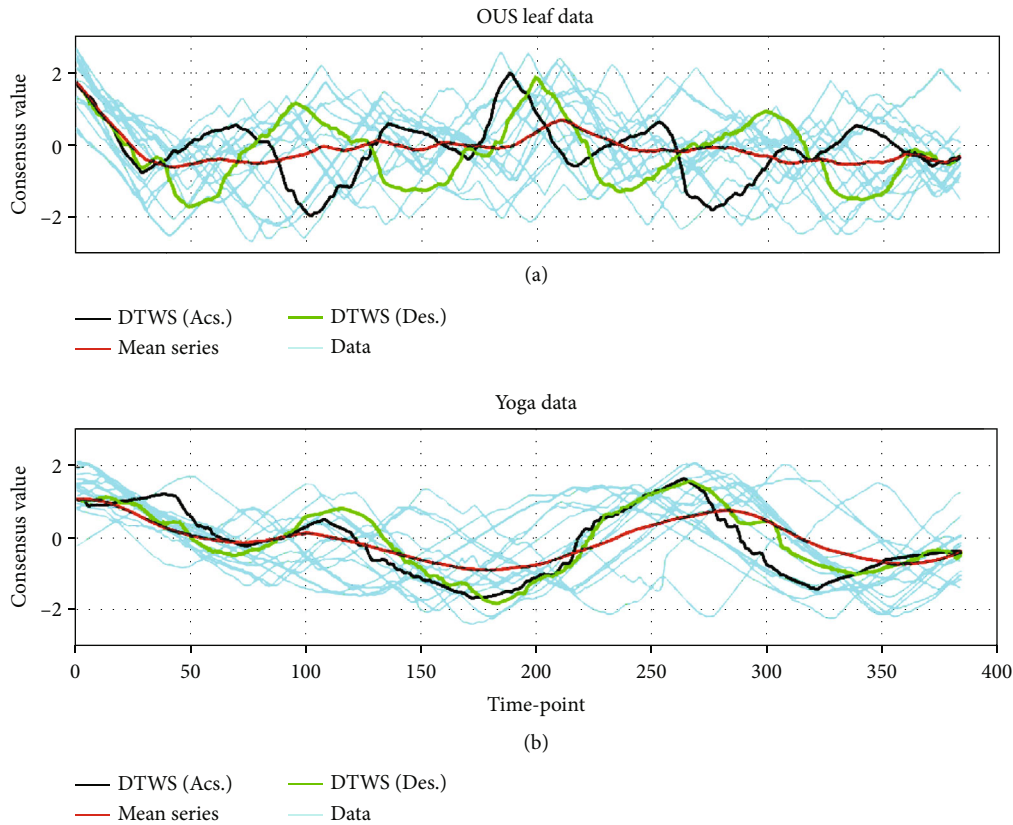
Figure 6: Mean series, DTWS with ascending hierarchy (Asc.) and DTWS with descending hierarchy (Des.) of two data sets: OSULeaf and Yoga.

from a consensus series C in the Euclidean sense. Mean and median of $P_i(C)$'s are two possible candidates for $P(C)$.

4.2. Dynamic Time Warping Distance. The string metric D in Section 5 yields the optimal warping path generated for two time series. We can make use of this measure to quantify the similarity between a consensus series $C = \{c_n\}_{n=1}^{2N}$ and each data series $S^{(i)} = \{s_n^{(i)}\}_{n=1}^{2N}$; $i = 1, 2, \cdots, k$ in the sense of pattern. Suppose $D_i(C)$ stands for that similarity between C and $S^{(i)}$ (i.e., $D_i(C) = D(C_{2N}, S_{2N}^{(i)})$). This comes recursively via Equation (7), which is in accordance with Equation (1).

$$D\left(C_p, S_q^{(i)}\right) = d\left(c_p^{(1)}, s_q^{(i)}\right) + \min \begin{cases} D\left(C_{p-1}, S_{q-1}^{(i)}\right), \\ D\left(C_p, S_{q-1}^{(i)}\right), \\ D\left(C_{p-1}, S_q^{(i)}\right). \end{cases} \quad (7)$$

We recall that lesser D indicates more similarity in patterns. A representative DTW distance can be obtained by taking an appropriate central measure (e.g., mean or median) of $D_i(C)$; $i = 1, 2, \cdots, k$ (say $D(C)$). Thus, we expect $D(C)$ to represent overall deviation of time series from a consensus series C in the sense of pattern. Determining $D(C)$ and $P(C)$ is further discussed in Section 6.2.

Remark 12. Suppose $C^{(1)}$ and $C^{(2)}$ be two consensus series of the same set of time series. In summary, if $D(C^{(1)}) > D(C^{(2)})$ then we would expect time series are more deviated from $C^{(1)}$ than $C^{(2)}$ when pattern attributes are concerned. If $P(C^{(1)}) > P(C^{(2)})$ then $C^{(2)}$ shows more central tendency than $C^{(1)}$.

5. Data Processing

We test consensus series using two data sets from UEA & UCR Time Series Classification Repository, namely, OSU-Leaf and Yoga [24] (http://www.timeseriesclassification .com). All the data series are stored as normalized series by z-scores. It is a usual measure for many experimental settings in pattern recognition [12, 32]. Such normalization is compatible with the work of UTS too. OSULeaf data set contains one-dimensional outlines of leaves obtained by colour image segmentation and boundary extraction from digitized leaf images of six classes. Angles measured at each pixel point along the boundary of leaves form a set of series. We are motivated to choose this particular data set due to shifted fluctuating patterns in different phases. In order to preserve the meaning of summarizing a set of time series, we stick into a one class of data, class six, which includes 15 time series each having 427 data points. Note that to see a complete decomposition up to a certain level, say level h, we need number of data points to be divisible by 2^h. We choose first 384 consecutive data points to attain maximal comparison up
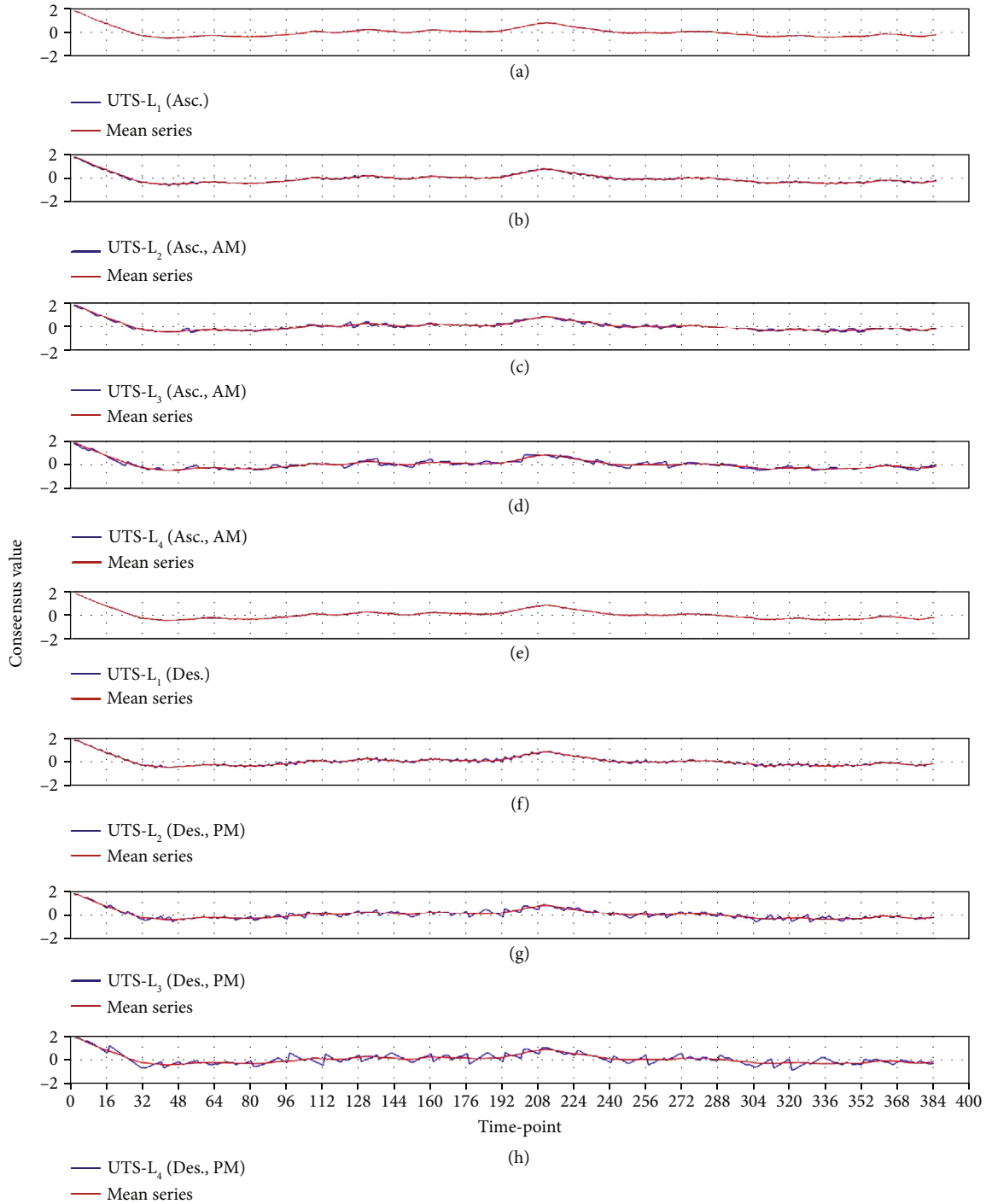
FIGURE 7: UTS-L_1–UTS-L_4: (a–d) ascending hierarchy with AM-inverse (Asc., AM) and (e–h) descending hierarchy with PM-inverse (Dec., PM). Data track: OSULeaf data.

to seven levels of UTS. We start by the second data point for 2nd coupling.

The second data set, Yoga, has been generated by capturing two actors transiting between yoga poses in front of a green screen giving to two classes of data. Each image has converted to a one-dimensional series by finding the outline and measuring the distance of the outline to the center. That is, the motion of an actor captured by sensors is transformed

to a time series. We proceed with class 1 containing 137 series. Out of 426 data points, we proceed with the first 384 consecutive data points, again to compatible with seven levels of UTS.

All 15 series of OSULeaf data are shown in Figure 6(a) along with the mean series and DTWS. Figure 6(b) depicts those of Yoga data. Here, only a few of data series (15) are shown for clear appearance, picking every 10th series in the
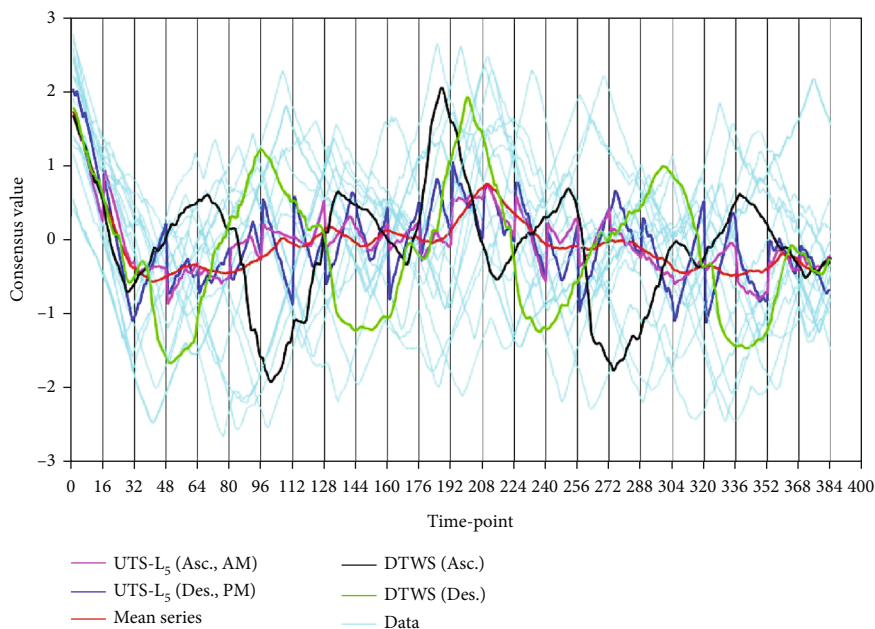
FIGURE 8: UTS-L_5: descending hierarchy with PM-inverse (Des., PM) and ascending hierarchy with AM-inverse (Asc., AM). Data track: OSULeaf data.
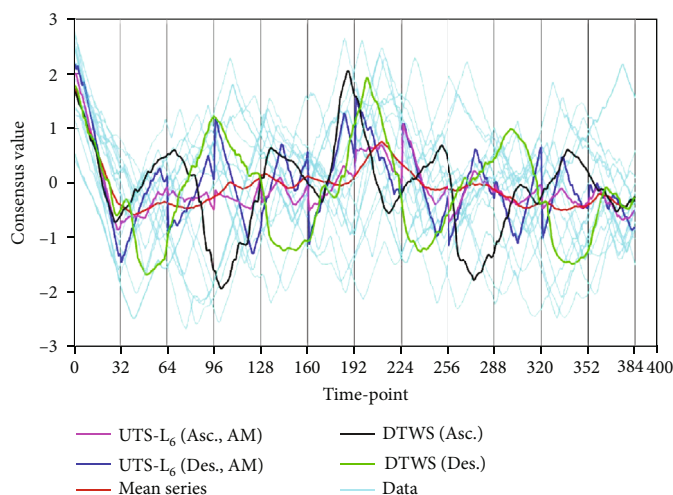

FIGURE 9: UTS-L_6: ascending hierarchy with AM-inverse (Asc., AM) and descending hierarchy with PM-inverse (Des., PM). Data track: OSULeaf data.

ascending order of $A^{(i)}$ (see Section 3.5) and the series having the least and the highest $A^{(i)}$.

Notably, two cases of DTWS of OSULeaf data follow opposite patterns (increasing and decreasing) in many segments of time-points. However, Yoga data do not show such a significant difference in pattern. Sensitivity of ascending and descending hierarchies to the pattern of UTS-L_h can be scrutinized later in Section 6.

6. Results

We present consensus series (UTS-L_h, DTWS, and the mean series) in a systematic way for comparisons. Applicability of UTS-L_h is overseen on top of DTWS and the mean series.

6.1. Illustrations of Consensus Series. We illustrate results on four pillars: (i) coupling approach—1st coupling and 2nd coupling (from Section 3.1), (ii) taming levels and mother wavelet for inverse—average mother (AM-inverse) and previous mother (PM-inverse) (from Section 3.2.1), (iii) hierarchy of data series—ascending and descending (from Section 3.5), and (iv) comparisons—via the Euclidean distance and DTW distance (from Section 4). Pattern recognition becomes more challenging when adjacent data points are highly varied in dissimilar way. If it is required to prioritize sizable adjacent variations, the preferable combination is to execute descending hierarchy with PM-inverse. As projected by Remark 6, new wavelets in taming level are biased on that of first series. Therefore, descending hierarchy brings
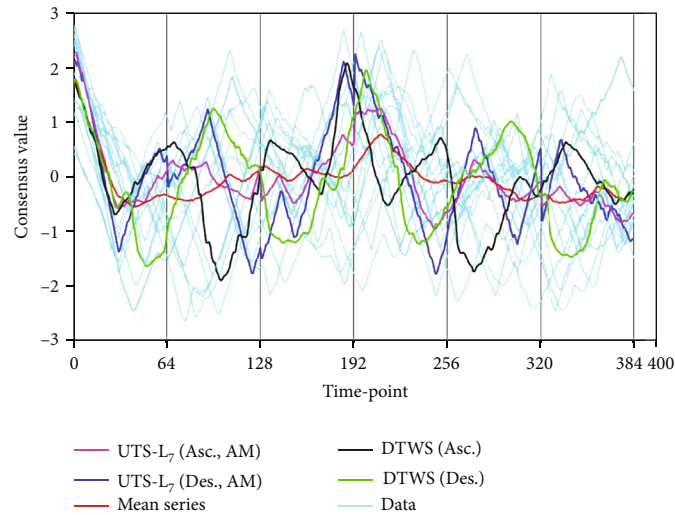
FIGURE 10: UTS-L_7: ascending hierarchy with AM-inverse (Asc., AM) and descending hierarchy with PM-inverse (Des., PM). Data track: OSULeaf data.
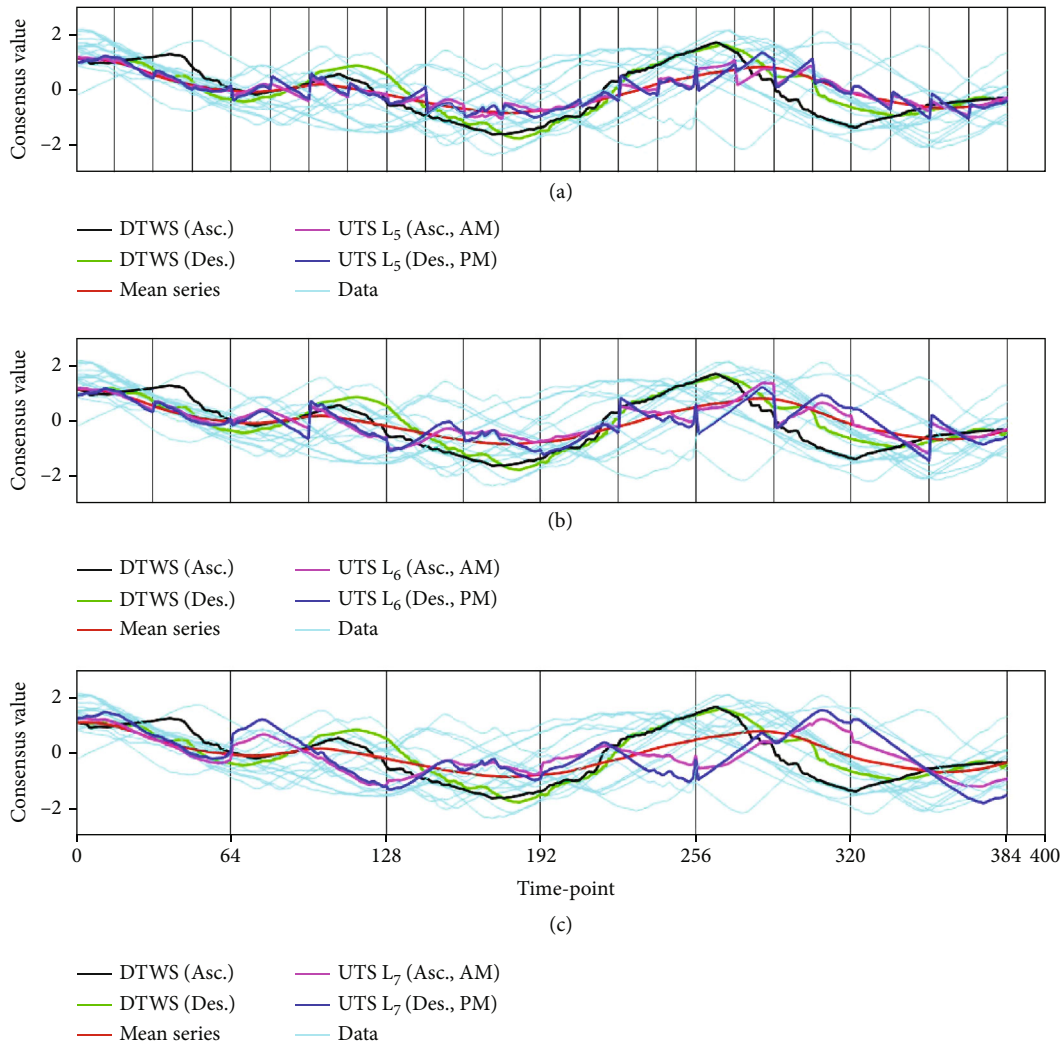

FIGURE 11: (a) UTS-L_5, (b) UTS-L_6, and (c) UTS-L_7: descending hierarchy with PM-inverse (Des., PM) and ascending hierarchy with AM-inverse (Asc., AM). Data track: Yoga data.
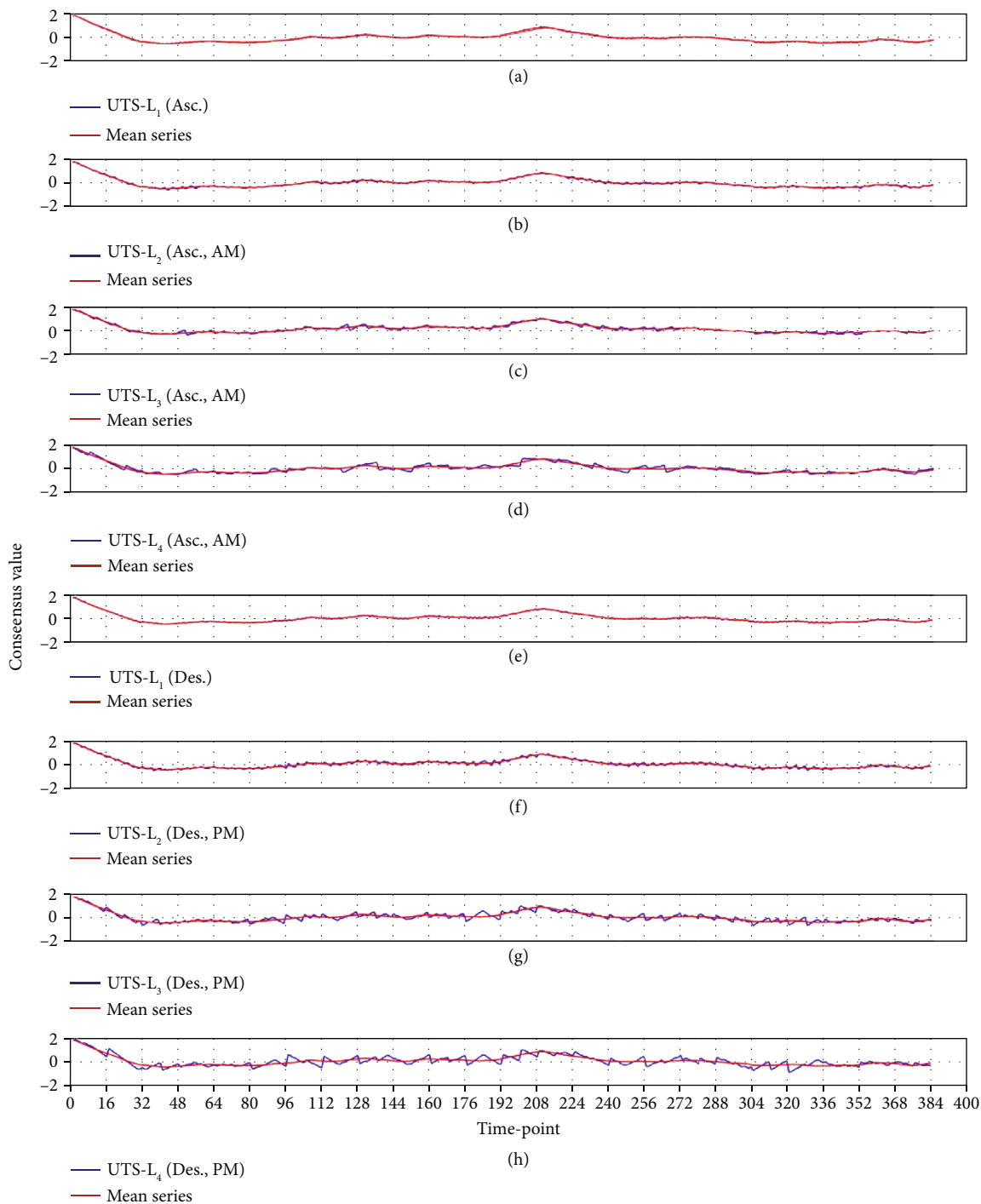
FIGURE 12: UTS-L_1–UTS-L_4 (2nd coupling): (a–d) ascending hierarchy with AM-inverse (Asc., AM) and (e–h) descending hierarchy with PM-inverse (Des., PM). Data track: OSULeaf data.

more sensitivity to adjacent variations. This is further supported by PM-inverse as per Remark 7. According to that, not only at the taming level but also all the earlier levels in inverse procedure are biased on the first series. In addition, a theoretical support is established by Proposition 10 (generalized version) that assures higher variations in tamed series for higher variations in data series. Finally, such variations are infused into consensus series UTS-L_h.

In the other extreme, ascending hierarchy with AM-inverse yields much smoothed output close to the mean series. It seems OSULeaf data do not show significant adjacent variations. It is observed in Figure 7 that even descending hierarchy with PM-inverse provides UTS-L_h, $h = 1, 2, 3, 4$, more or less the same as the mean series. In general, if we see this similarity until we proceed into higher levels, then we have less variation in early windows of temporal proximity.
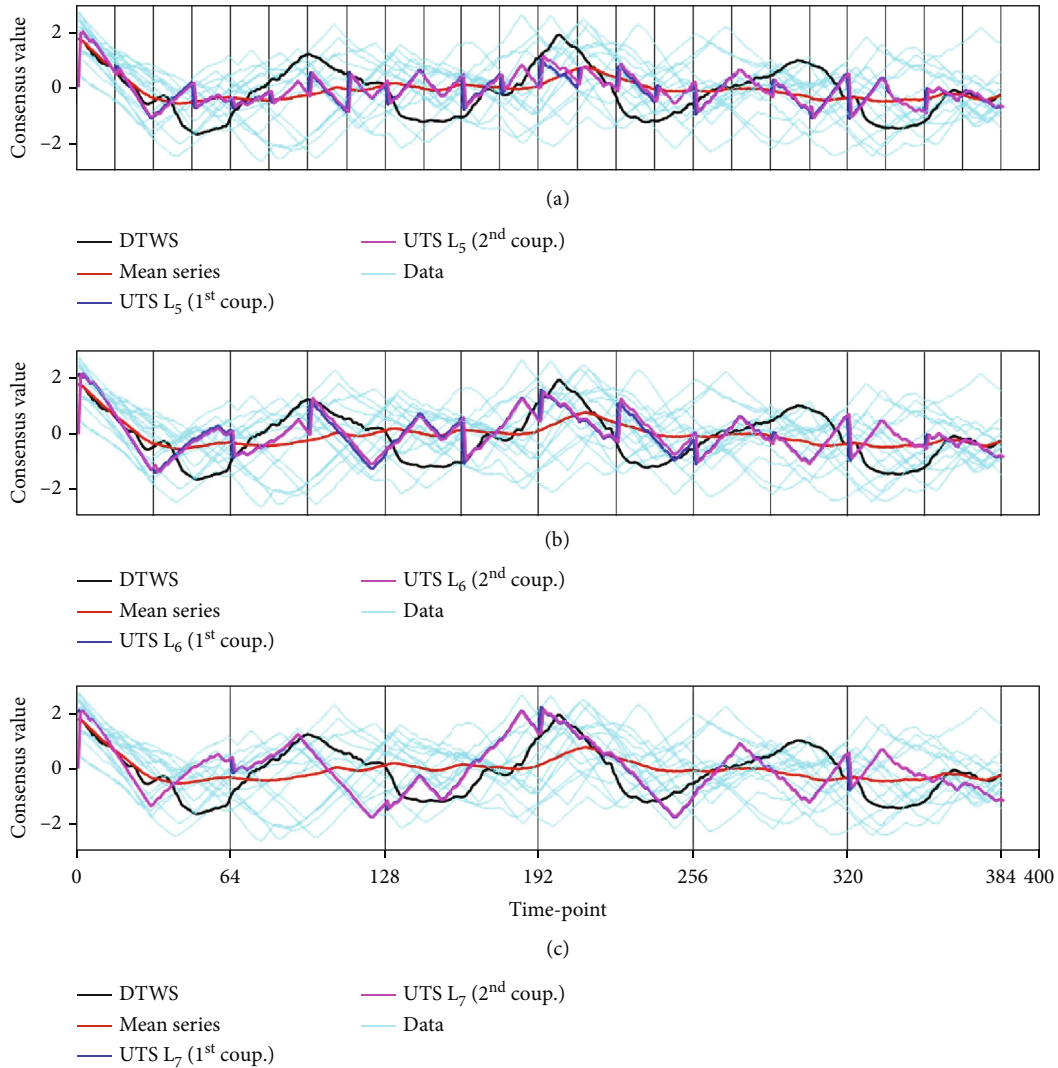
FIGURE 13: 1st coupling and 2nd coupling of (a) UTS-L_5, (b) UTS-L_6, and (c) UTS-L_7: descending hierarchy with PM-inverse (Des., PM). Data track: OSULeaf data.

Once we consider the wavelet-based temporal aspect, four windows relevant to UTS-L_h, $h = 1, 2, 3, 4$, consist of 2-tuples, 4-tuples, 8-tuples, and 16-tuples, respectively. This is due to each wavelet in level h is responsible for 2^h data points (Section 3.2.1). Next, UTS-L_5 has a window of 32 data points, where a wavelet in level 5 is obtained using averages of 16-tuples. Thus, taming at level 5 deals with pattern of the series of 16-tuple averages. According to Figure 8, UTS deviates from the mean series, in particular with PM-inverse, suggesting considerable variations in 16-tuple averages. In 16-tuple windows (between vertical lines), we have the relevant consensus series. It can be observed that two DTWS do not adhere to such localized patterns. Moreover, DTWS is highly sensitive to hierarchy (ascending/descending) in many windows.

A similar visualization is carried out for UTS-L_6 and UTS-L_7 in Figures 9 and 10, respectively. Note that UTS-L_6 and UTS-L_7 bring consensus patterns in 32-tuples and 64-tuples, respectively. Now, UTS-L_h follow the pattern with much relaxed localization. It is visually verified by closeness

to DTWS than the mean series. In a holistic approach, a user can adjust the windows either by size or position to see respective consensus patterns, subject to the requirements of H_v-transform. This customizing option facilitates a setting of moving-window that can be conceptualized analogous to moving-average.

Observe that skips appear in UTS-L_h at the vertical lines separating windows. This is a result of independent backward calculation (inverse procedure) in each window that yields loss of connection across those vertical lines. In these trials of OSULeaf data, we used first 384 data points out of total length of 427 points. This is to satisfy the requirement of level 7, where the tail or rest of the data points cannot be compiled into one complete window. If someone needs to generate UTS-L_h, say up to level 7 including the tail, we may select last 384 data points. This is only a matter of one input command in an algorithm for UTS-L_h.

As a further verification, we produce the same illustrations for Yoga data (Figure 11). Here also, two DTWS fails

TABLE 1: Statistics of the Euclidean distance between consensus series and data series. Data track: OSULeaf data.

Consensus series (C)	Mean	Median	Min	Max	Standard deviation
Mean series	316.79	346.01	184.72	479.84	86.36
UTS-L_1 (Asc.)	317.08	346.44	185.04	480.02	86.30
UTS-L_2 (Asc., AM)	317.84	347.28	186.24	480.57	86.03
UTS-L_3 (Asc., AM)	319.45	349.58	190.20	481.17	84.85
UTS-L_4 (Asc., AM)	323.64	350.87	193.46	485.17	85.48
UTS-L_5 (Asc., AM)	331.10	345.90	218.54	467.95	73.69
UTS-L_6 (Asc., AM)	337.21	321.61	227.04	527.77	85.61
UTS-L_7 (Asc., AM)	370.84	336.31	225.00	651.82	114.22
DTWS (Asc.)	554.60	497.50	325.54	919.09	181.40
UTS-L_1 (Des.)	317.43	346.79	185.75	480.65	86.19
UTS-L_2 (Des., PM)	319.56	349.39	190.34	482.43	85.21
UTS-L_3 (Des., PM)	325.42	356.63	201.91	490.22	83.54
UTS-L_4 (Des., PM)	337.16	346.41	225.98	490.45	77.75
UTS-L_5 (Des., PM)	367.26	349.37	247.46	535.26	75.14
UTS-L_6 (Des., PM)	427.95	426.61	216.20	619.07	107.64
UTS-L_7 (Des., PM)	573.80	624.46	110.61	899.26	206.04
DTWS (Des.)	543.82	528.70	243.52	793.35	132.99

TABLE 2: Statistics of DTW distance between consensus series and data series. Data track: OSULeaf data.

Series	Mean	Median	Min	Max	Standard deviation
Mean series	188.30	174.53	151.43	240.59	30.11
UTS-L_1 (Asc.)	183.59	170.75	147.25	234.94	30.17
UTS-L_2 (Asc., AM)	178.60	166.70	142.48	229.29	30.36
UTS-L_3 (Asc., AM)	174.33	159.81	135.78	225.57	30.24
UTS-L_4 (Asc., AM)	166.83	154.30	134.56	215.11	26.46
UTS-L_5 (Asc., AM)	157.25	142.02	125.26	209.32	28.41
UTS-L_6 (Asc., AM)	150.56	143.95	114.81	204.70	29.03
UTS-L_7 (Asc., AM)	138.90	124.01	104.93	220.32	32.00
DTWS (Asc.)	87.43	80.93	60.37	144.11	24.82
UTS-L_1 (Des.)	184.30	172.92	149.16	236.35	30.08
UTS-L_2 (Dec., PM)	180.37	172.12	146.78	231.15	29.22
UTS-L_3 (Dec., PM)	168.48	157.73	139.92	220.45	28.16
UTS-L_4 (Dec., PM)	154.50	145.63	124.92	211.09	25.95
UTS-L_5 (Dec., PM)	140.42	131.75	113.14	196.65	23.80
UTS-L_6 (Dec., PM)	116.24	103.95	91.84	174.46	22.98
UTS-L_7 (Dec., PM)	95.21	92.70	63.79	174.48	25.19
DTWS (Des.)	80.40	84.85	36.72	144.53	30.77

to show localized pattern attributes of the data series. However, both cases of (Asc., AM) and (Des., PM) show almost similar patterns compared to that of OSULeaf data. This interprets a lesser sensitivity of hierarchy in producing localized consensus patterns via UTS-L_h. The same is hinted by the similarity of two DTWS in more nonlocalized aspect. In such a way, observation of UTS-L_h along with DTWS yields more useful descriptive analysis of a set of time series.

Next, we demonstrate consensus series associated with 2nd coupling. As assured by general versions of Proposition 8 and Proposition 9, existence and uniqueness preserve for both couplings, subject to a choice for hierarchy (descending/ascending) and mother wavelet for inverse (AM-inverse/PM-inverse). Here, the aim is to see whether UTS-L_h differ in general agreement shown by 1st coupling. We proceed with 384 data points starting from the second point of OSULeaf data. Figure 12 illustrates UTS-L_1–UTS-L_4 analogous to that of Figure 7. It demonstrates no significant difference showing slight wobbles around the mean series similar to 1st coupling.
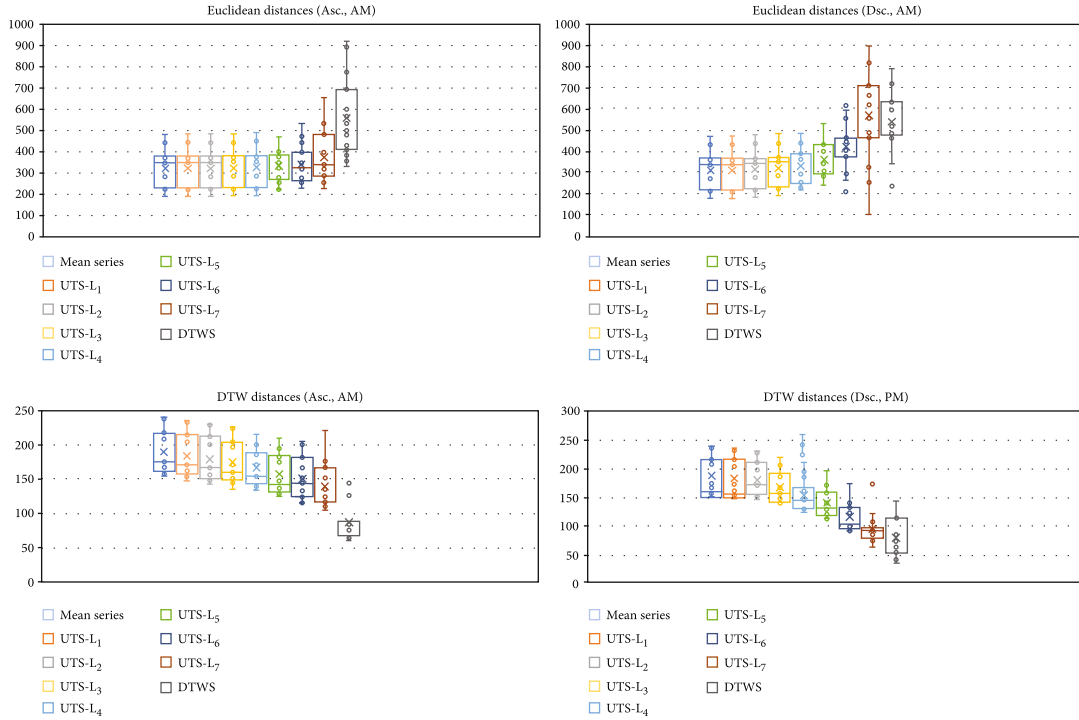
FIGURE 14: Box-plot diagrams of the Euclidean distances and DTW distances (lower box limit: 1st quartile (Q_1), upper box limit: 3rd quartile (Q_3), middle line: median, cross ($\times$): mean, lower whisker: $Q_1 - 1.5$ IQR, and upper whisker: $Q_3 + 1.5$ IQR). Data track: OSULeaf data.

Figure 13 illustrates the rest of UTS-L$_h$ preserving similar windowing effect followed earlier, subject to a shift by one point. If UTS-L$_h$ (Des., PM) measures do not deviate much from 1st coupling counterparts, then it is hard to expect deviations in other options of descending/ascending, AM-inverse/PM-inverse. Therefore, we present only that category showing comparison between 1st and 2nd couplings for the sake of brevity too. According to Figure 13, UTS-L$_h$ behave almost the same way in both the couplings.

6.2. Comparisons of Consensus Series. Earlier section is devoted to visualize the behaviour of UTS-L$_h$. In this section, we provide numerical quantification to how far they deviate from data series in the sense of the Euclidean distance and DTW distance. These two distance measures are presented in Section 4 as $P_i(C)$ and $D_i(C)$, respectively, where i stands for data series and C is the concerned consensus series.

We have 15 data series for the trial of OSULeaf data. Table 1 contains basic statistics of corresponding $P_i(C)$; $i = 1, 2, \cdots, 15$. Here also, we consider UTS-L$_h$ of descending hierarchy with PM-inverse and ascending hierarchy with AM-inverse. For brevity, only 1st coupling is implemented. The Euclidean distances when C is the mean series and DTWS are also included as baselines for comparisons.

According to Table 1 The mean series shows the least mean, while DTWS leans towards the greatest mean. Therefore, UTS-L$_h$ are characterized with much compromised overall deviations. Mean increases when level of taming h increases in both cases of descending hierarchy with PM-inverse and ascending hierarchy with AM-inverse. Notably,

this guides that higher level taming directs to lesser concern on point-wise central tendency. Median does not show the same monotonic behaviour as mean. However, median of UTS-L$_h$ with higher h is closer to that of two DTWS than the mean series. It also shows the loss of point-wise central tendency. In contrast, we expect that UTS-L$_h$ with higher h reflect pattern attributes rather than point-wise location. To claim this, we present statistics of DTW distances in Table 2. This pattern-oriented characteristic can be declared as a key applicability of UTS-L$_h$ as consensus series.

According to Table 2 also, UTS-L$_h$ are in a compromised context as far as the mean series and DTWS are concerned. This is implied by acquiring the least mean by DTWS and the greatest mean by the mean series. The same is observed in median as well. Both mean and median decrease when level of taming h increases (except in median of UTS-L$_6$ (Asc., AM)), giving opposite to what we observe in Table 1. This verifies that UTS-L$_h$ with higher h align more with pattern attributes. Interestingly, UTS-L$_h$ still follow temporal windows, where DTWS is not abide by such localization.

The standard deviation of certain UTS is comparatively high. This may be due to outliers of distance measures resulting from outlier data series rather than unsuitability of UTS as a consensus series. Therefore, as a holistic view, we depict box-plot diagrams carrying statistics of above two tables in Figure 14. The boxes are generated for UTS measures at each level together with DTWS and mean series representative. The spread of Euclidean distances and DTW distances are produced for extreme hierarchies ascending with average mother inversion and descending with previous mother
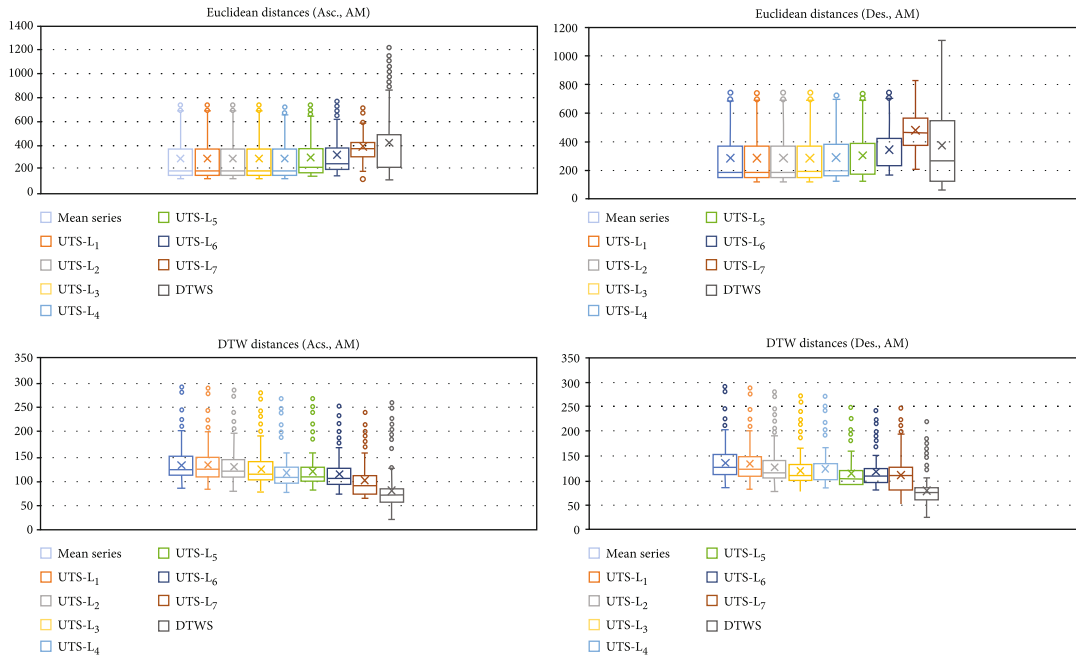
FIGURE 15: Box-plot diagrams of the Euclidean distances and DTW distances (lower box limit: 1 st quartile (Q_1), upper box limit: 3rd quartile (Q_3), middle line: median, cross ($\times$): mean, lower whisker: $Q_1 - 1.5$ IQR, and upper whisker: $Q_3 + 1.5$ IQR). Data track: Yoga data.

inversion to generate the box plots. Now, increasing trend in the Euclidean distance and decreasing trend in DTW distance as h increases are clearly visualized.

We demonstrate the same for Yoga data in Figure 15, and the same trends can be observed as in Figure 14. In a data mining process, one may investigate all those UTS-L_h and select a fair enough representative for further work, knowing which temporal windowing prevails.

7. Discussion

Wavelet techniques are utilized in time series analysis in order to investigate trends, recurring fluctuations, and autocorrelation effects [33, 34]. These applications refer to longitudinal aspect of time series. The approach via the discrete Haar wavelet proposed in this work adds on a utility in terms of cross-sectional characteristics of a set of time series. It is associated with the central topic of consensus patterns. Its novelty lies with the ability of customizing a consensus pattern by maintaining localized temporal concern in different extents.

The study is centralized around a pool of measures designated as Ultimate Tamed Series (UTS). A theoretical platform is also established enclosing existence, uniqueness, and continuous dependency (Proposition 8, 9, and 10). Implementation of UTS in different hierarchies (ascending and descending), taming levels and inverse procedure (AM-inverse and PM-inverse), and coupling formation (1st coupling and 2nd coupling) bring the essence of the overall performance, again highlighting customizing options. Hierarchical arrangement facilitates bringing adjacent variations of data points into proposed consensus series. A window of 2^h data points is catered at a time in taming by

level h. UTS-L_h measures owe this characteristic by the routine set by the discrete Haar wavelet decomposition. Wavelet approaches gain their reputation due to such localized treatment as evident in many applications [35–37]. In fact, father wavelets in different levels allow rescaling data series, while mother wavelets affix localized trends. The work relevant to this begins in Definition 1 (H_v-transform), and it is further illustrated in Figure 3.

If we shift the implementation of UTS by one point of each data series (2nd coupling), then all outputs are different. We accept this as an optimistic feature than a limitation, since more pattern attributes can be extracted. Both classes of UTS (1st and 2nd couplings) show similar pattern if there are no contrasting variations because of the shift, which is the case for trialed data in this work (Figures 12 and 13). Otherwise, it is preferred to determine UTS-L_h for both couplings in all illustrations since we identify discrete Haar wavelet as a weak form (Section 3.1).

As baselines, we consider two consensus series, the mean series (series of point-wise averages) and Dynamic Time Warping Series (DTWS), to show how far UTS-L_h can illustrate patterns emphasizing local attributes. Note that we have illustrated normalized data series, and hence, the mean series acts as a guide for over-smoothed representation for pattern. Option of PM-inverse avoids over-smoothing, which is often the main advantage of UTS-L_h over the mean series. Higher level UTS-L_h visualize pattern attributes fairly, while lower level UTS-L_h locate closer to the mean series. By moving back and forth with different levels of taming, one can extract important characteristics of a set of time series. Thus, UTS-L_h measures are pooled as a collection of statistics, where a user can see more complete profile rather than seeing one consensus series. Its computational burden is redundant in modern computer technology. A DTWS (without scaling)

is substantially longer compared to the original series. This expansion happens in a nonlinear manner so that uniform scaling techniques may not align the series properly to the temporal axis. Therefore, improved versions such as NLAAF and PSA [14, 26] can be tested with UTS-L$_h$ in a future work.

8. Conclusion

UTS-L$_h$ bridge the gap between the mean series and DTWS in the sense of displaying consensus patterns. The classical DTW approaches only recognize patterns somewhat neglecting the correspondence to the temporal axis while the mean series restricts into point-wise concerns. UTS-L$_h$ with different levels of taming allow seeing consensus patterns by expanding and shrinking the localization of time-points (Figures 14 and 15). We verify these effects via two comparison tools: Euclidean distance and DTW distance.

UTS-L$_h$ can also be tested for data sets that are highly varied. For that, clustering tools must be proposed to see informative consensus patterns. Several motivations can be obtained by existing clustering algorithms including well-known k-means [32, 38, 39]. In the broader scope of data mining, one needs not only extracting important data but also extracting useful patterns [40]. Thus, we further propose UTS as a technique for mining consensus patterns of sequential big data.

Data Availability

UEA & UCR Time Series Classification Repository can be found at http://www.timeseriesclassification.com/index.php. OSU Leaf data is available at http://www.timeseriesclassification.com/description.php?Dataset=OSULeaf. Yoga data can be found at http://www.timeseriesclassification.com/description.php?Dataset=Yoga.

Conflicts of Interest

Authors declare that there is no conflict of interest.

References

[1] B. Pavlyshenko, "Machine-learning models for sales time series forecasting," *Data*, vol. 4, no. 1, p. 15, 2019.

[2] L. Kirichenko, T. Radivilova, and V. Bulakh, "Machine learning in classication time series with fractal properties," *Data*, vol. 4, no. 1, 2018.

[3] D. Afolabi, S. U. Guan, K. L. Man, P. W. H. Wong, and X. Zhao, "Hierarchical meta-learning in time series forecasting for improved interference-less machine learning," *Symmetry*, vol. 9, no. 11, p. 283, 2017.

[4] NIST/SEMATECH, "e-handbook of statistical methods," http://www.itl.nist.gov/div898/handbook/.

[5] G. Anadiotis, "Processing time series data: what are the options?," 2018, https://www.zdnet.com/article/processing-time-series-data-what-are-the-options.

[6] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, 1996.

[7] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, "Querying and mining of time series data: experimental comparison of representations and distance measures," *Proceedings of the VLDB Endowment*, vol. 1, no. 2, pp. 1542–1552, 2008.

[8] S. W. Kim, S. Park, and W. W. Chu, "An index-based approach for similarity search supporting time warping in large sequence databases," in *Proceedings 17th International Conference on Data Engineering*, pp. 607–614, Heidelberg, Germany, 2001.

[9] F. Petitjean and P. Gancarski, "Summarizing a set of time series by averaging: from Steiner sequence to compact multiple alignment," *Theoretical Computer Science*, vol. 414, no. 1, pp. 76–91, 2012.

[10] J. Aach and G. M. Church, "Aligning gene expression time series with time warping algorithms," *Bioinformatics*, vol. 17, no. 6, pp. 495–508, 2001.

[11] Z. Bar-Joseph, G. Gerber, D. K. Giord, T. S. Jaakkola, and I. Simon, "A new approach to analyzing gene expression time series data," in *Proceedings of the sixth annual international conference on Computational biology - RECOMB '02*, pp. 39–48, Washington, DC, USA, 2002.

[12] F. Petitjean, A. Ketterlin, and P. Gancarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

[13] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[14] V. Niennattrakul and C. A. Ratanamahatana, "Shape averaging under time warping," in *2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 626–629, Chonburi, Thailand, 2009.

[15] S. Ongwattanakul and D. Srisai, "Contrast enhanced dynamic time warping distance for time series shape averaging classication," in *2nd Int. Conf. Interaction Sciences: Information Technology, Culture and Human (ICIS '09)*, pp. 976–981, Seoul, Korea, 2009.

[16] N. C. Ganegoda, *Mathematical Analysis of Parasite Dynamics of Lymphatic Filariasis Using Range Modelling and Other Approaches in Relation to Observational Data from Sri Lanka, [Ph.D. thesis]*, University of Sri Jayewardenepura, Nuge-goda, Sri Lanka, 2011, http://dr.lib.sjp.ac.lk/handle/123456789/4927.

[17] N. C. Ganegoda, K. K. W. A. S. Kumara, D. A. Tantrigoda, S. K. Boralugoda, and S. S. N. Perera, "An approach for visualizing error and obtaining a measure of central tendency regarding a set of time series using discrete Haar wavelet," *Journal of Wavelet Theory and Applications*, vol. 10, no. 1, pp. 1–18, 2016.

[18] C. P. Waduge and N. C. Ganegoda, "Case study on measuring central tendency of a set of time series based on discrete Haar wavelet," in *Proceedings of the International Conference on Computational Modelling and Simulation 2017*, pp. 312–315, Colombo, Sri Lanka, 2017.

[19] R. S. Stankovic and B. J. Falkowski, "The Haar wavelet transform: its status and achievements," *Computers and Electrical Engineering*, vol. 29, no. 1, pp. 25–44, 2003.

[20] P. Porwik and A. Lisowska, "The haar–wavelet transform in digital image processing: its status and achievements," *Machine Graphics and Vision*, vol. 13, pp. 79–98, 2004.

[21] R. C. Guido, "A note on a practical relationship between filter coefficients and scaling and wavelet functions of discrete wavelet transforms," *Applied Mathematics Letters*, vol. 24, no. 7, pp. 1257–1259, 2011.

[22] C. P. Waduge, N. C. Ganegoda, D. C. Wickramarachchi, R. S. Lokupitiya, and G. H. J. Lanel, "A group of representative measures of a set of time series and its decision support: a trial for dengue incidence data of Sri Lanka," *International Journal of Information and Decision Sciences*, In press.

[23] V. Niennattrakul and C. Ratanamahatana, "On clustering multimedia time series data using k-means and dynamic time warping," in *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*, pp. 733–738, Seoul, Korea, 2007.

[24] W. V. Anthony Bagnall, J. Lines, and E. Keogh, "Ucr time series classication archive," 2020, http://www.timeseriesclassification.com.

[25] H. Sakoe and S. Chiba, "A dynamic programming approach to continuous speech recognition," *Proceedings of the Seventh International Congress on Acoustics*, vol. 3, pp. 65–69, 1971.

[26] L. Gupta, D. Molfese, R. Tammana, and P. Simos, "Nonlinear alignment and averaging for estimating the evoked potential," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 4, pp. 348–356, 1996.

[27] A. W. C. Fu, E. J. Keogh, L. Y. H. Lau, C. A. Ratanamahatana, and R. C. W. Wong, "Scaling and time warping in time series querying," *The VLDB Journal*, vol. 17, no. 4, pp. 899–921, 2008.

[28] J. Fish and T. Belytschko, *A First Course in Finite Elements*, John Wiley & Sons, 2007.

[29] L. C. Evans, *Partial Differential Equations. Graduate Studies in Mathematics*, American Mathematical Society, 1998.

[30] N. C. Ganegoda, K. K. W. A. S. Kumara, D. A. Tantrigoda, S. K. Boralugoda, and S. S. A. Perera, "A discrete Haar wavelet based approach for visualizing error regarding a simulated time series," *GSTF Journal on Computing*, vol. 3, no. 1, pp. 56–58, 2013.

[31] V. Niennattrakul and C. A. Ratanamahatana, "Inaccuracies of Shape Averaging Method Using Dynamic Time Warping for Time Series Data," in *Computational Science–ICCS 2007*, Y. Shi, G. Albada, J. Dongarra, and P. Sloot, Eds., pp. 513–520, Springer, Berlin Heidelberg, 2007.

[32] L. Gupta and R. Tammana, "A discrepancy measure for improved clustering," *Pattern Recognition*, vol. 28, no. 10, pp. 1627–1634, 1995.

[33] S. Ahmad, T. Taskaya-Temizel, and K. Ahmad, "Summarizing Time Series: Learning Patterns in `Volatile' Series," in *Intelligent Data Engineering and Automated Learning (IDEAL)-2004*, Z. R. Yang, H. Yin, and R. M. Everson, Eds., vol. 3177 of Lecture Notes in Computer Science, , Springer, 2004.

[34] S. O. Renaud, J. L. Starck, and F. Murtagh, "Prediction based on a multiscale decomposition," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 1, no. 2, pp. 217–232, 2003.

[35] D. K. Ruch and P. J. V. Fleet, *Wavelet Theory: An Elementary Approach with Applications*, John Wiley and Sons Inc, 2009.

[36] J. Baccou and J. Liandrat, "Definition and analysis of a wavelet/fictitious domain solver for the 2d-heat equation on a general domain," *Mathematical Models and Methods in Applied Sciences*, vol. 16, no. 6, pp. 819–845, 2006.

[37] C. Canuto and I. Cravero, "A wavelet-based adaptive finite element method for advection-diusion equations," *Mathematical Models and Methods in Applied Sciences*, vol. 7, no. 2, pp. 265–289, 1997.

[38] C. A. Ratanamahatana and E. Keogh, "Making time-series classification more accurate using learned constraints," in *Proceedings of the 2004 SIAM International Conference on Data Mining*, pp. 11–22, Florida, USA, 2004.

[39] S. Zolhavarieh, S. Aghabozorgi, and Y. W. Teh, "A review of subsequence time series clustering," *The Scientific World Journal*, vol. 2014, Article ID 312521, 19 pages, 2014.

[40] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann publishers, 2001.