A Hybrid Time Series Prediction Model Based on Fuzzy Time Series and Maximal Overlap Discrete Wavelet Transform

Nevin GULER DINCER [iD] , Muhammet Oguzhan YALCIN* [iD] , Oznur ISCI GUNERI [iD]

Department of Statistics, Faculty of Science, University of Mugla Sıtkı Kocman

Highlights
• This study proposes a new time series prediction method.
• This method combines FTS based on fuzzy clustering and MODWT.
• Proposed method is based on decomposing of time series into sub-time series through MODWT.
• The main objective is to improve the prediction and forecasting performance of existing FTS methods.

Abstract

This study proposes a new time series prediction method that combines Fuzzy Time Series (FTS) based on fuzzy clustering and Maximal Overlap Discrete Wavelet Transform (MODWT). Time series generally consist of subseries, each of which reflects the different behavior of the time series and using of a single prediction method for all subseries can be negatively impacted the prediction and forecasting accuracy. Proposed method is based on decomposing of time series into sub-time series through MODWT and predicting an FTS model for each sub-time series separately. Besides, time series can contain noise, outlier or unwanted data points and these points can hide the actual behavior of the time series. MODWT has the ability of eliminating negative effects of these kind of data points on the predictions. Besides, proposed method has also all advantages of FTS methods. The main objective of this study based on these advantages is to improve the prediction and forecasting performance of existing FTS methods based on fuzzy clustering. In order to show the performance of proposed method, three FTS methods based on fuzzy clustering and wavelet-based versions of them are applied to eight real time series and experimental results clearly showed that proposed method achieves the best prediction and forecasting results.

1. INTRODUCTION

Time series is one of the data types widely used in many researches relating to fields such as statistics, economics, finance, medicine, astronomy, engineering etc. It can be simply described as a set consisted of observations measured at successive time intervals (denoted as $y=[y\_1,y\_2,\ldots,y\_n]$). The main objective of time series analysis is to reveal stochastic behavior of time series such as trend, seasonality, irregularity and to predict its future values. For this objective, many modeling methods are used that differ according to the characteristics of time series. Autoregressive Moving Average for stationary time series, Integrated Autoregressive Moving Average [1-3] for the time series which are non-stationary but made stationary by taking difference and Artificial Neural Networks [4] and modeling methods based on fuzzy logic [5-7] for time series having nonlinear structure are generally used. However, these methods only consider the time domain of the series and not the frequency domain. Also, many of these methods may fail to provide high prediction performance when time series are not stationary and they include the noise, outliers, unwanted data points and sudden changes. To overcome these drawbacks of mentioned methods, the hybrid modeling techniques, based on especially discrete wavelet transform (DWT), maximal overlap discrete wavelet transform (MODWT), Fourier transform (FT) and empirical mode decomposition (EMD) have been developed in many studies [8-27].

This study also proposes a hybrid time series prediction method, which combines FTS based fuzzy clustering and MODWT. There are several reasons to choose MODWT as decomposition method instead of DWT, EMD and FT widely used in the literature [8-10,12-18,20-28]. MODWT allows to effectively analyze the time series both in time and frequency domain and thus, it preserves the time information of the time series unlike FT [28]. MODWT has also several advantages in comparison with DWT. The most important ones of these advantages are that MODWT can be applied to the time series with any length unlike DWT, is not sensitive against to the changing of the starting point of time series; and that wavelet and scaling coefficients as much as length of time series are obtained in each decomposition level of MODWT [19]. Although EMD has some superior properties such as; being model-free and fully data-driven differently from MODWT, it can be exposed to mode mixing and overlapping problems and thus it cannot be performed the proper decomposition of the time series [29].

FTS methods also have some advantages i.e. they do not require any statistical assumptions, they can work with incomplete, vagueness, linguistic data and they can easily handle the nonlinear structure of time series. FTS method was firstly proposed by Song and Chissom [30-32]. This method consists of three main steps. The first step, called as fuzzification, is to step of converting classical time series to fuzzy time series. In the second step, the relations between successive fuzzy sets are determined and the last step is the defuzzification which includes converting fuzzy predictions to the classical ones. FTS models can be divided into two categories as fuzzification method used in the first step: i) partitioning universe of discourse into subintervals [26-27] ii) fuzzy clustering [33-35]. The first category is based on partitioning classical time series into the subintervals at the predefined number. In here, each interval corresponds to a fuzzy set and each observation is fuzzified according to the interval it belongs. But, methods in this category do not consider the distribution of time series and assume that all time series have uniform distribution. When real time series are considered, satisfying of this assumption is very difficult. The major advantage of the methods in the second category is to learn the distribution of time series from own and thus is to be more appropriate the constituted fuzzy sets to the time series. Therefore, second fuzzification method is preferred in proposed method. The superiorities of proposed method can be collected under three main titles:

I) According to existing FTS methods based on fuzzy clustering [35-37].
- In the proposed method, time series is decomposed into subseries, each of which reflect the different behavior of the time series and FTS method is constructed for all subseries separately. Thus, proposed method provides predicting a model which can consider the local features of the time series in addition to its global features.
- It is more robust to noise, outlier and unwanted observations. Thus, the negative effects of these observations on prediction performance decrease.
- It analyzes the time series both in time and in frequency domain.

II) According to existing Wavelet Transform based methods [8-20].
- Existing methods assume that the observation measured at t time is only affected by its own lagged-values. But, FTS models investigate the relationship between fuzzy sets defined by classical time series and all observations in the fuzzy sets at t-1 time are used in order to predict an observation at t time. Thus, prediction process is performed with more information.
- Proposed method does not require many initial parameters such as the type of model (AR, MA, ARMA etc.), the network structure (the number of hidden layers, the number of neurons in the hidden layers, learning rate, learning algorithm) etc. It only needs the number of clusters.
- It has the better generalization ability since FTS methods are based on fuzzy sets, which consist of numerous classical time points.

III) According to existing Wavelet Fuzzy Time Series [26,27]
- Existing methods have used partitioning the universe of discourse method in the fuzzification step of FTS models. This kind of the fuzzification generally provides the good performance when the distribution of the time series is uniform. This study has used fuzzy clustering algorithms which

learn the distribution of the time series from the in the fuzzification step. Thus, the obtained fuzzy sets better reflect the behavior of the time series.
-   Existing methods have used DWT for decomposing the time series. In this study, MODWT has been used for the same objective due to some drawbacks of DWT.

This paper is organized into five sections. Section 2 gives material and methods used in this study. In Section 3, general framework of proposed hybrid time series prediction model is presented. Section 4 provides the experimental results including comparisons of prediction performances of traditional FTS (TFTS) model based on fuzzy clustering [33-35] and of proposed model. Last section concludes the study.

2. MATERIAL METHOD

This section consists of three subsections including basic definitions of FTS, FTS models based on fuzzy clustering and brief information about MODWT.

2.1. Basic Definitions of Fuzzy Time Series

The general definitions for FTS are summarized. Let be $U = \{u_1, u_2, \dots, u_c\}$ universe discourse where u_i s are subintervals. A fuzzy set can be defined as follows:

$$f_i = \mu_{f_i}(u_1)/u_1 + \mu_{f_i}(u_2)/u_2 + \dots + \mu_{f_i}(u_c)/u_c \tag{1}$$

where μ_{f_i} is the membership function of fuzzy set f_i, $\mu_{f_i}(u_i): U \rightarrow [0, 1]$ is membership degree of u_i to fuzzy set f_i. Some definitions relating to FTS can be given as follows:

Definition 1: Let be Y_t (t=1, 2,…,n) classical time series, be U is universe discourse of Y_t defined by fuzzy sets f_i. If $F(t)$ consists of f_i, then $F(t)$ is called *fuzzy time series* on Y_t.

Definition 2: Let be ° any arithmetic operator, if F(t) is only affected by one lagged fuzzy time series (F(t-1)), fuzzy relation between F(t) and F(t-1) can be denoted as follows:

$$\begin{aligned} F(t) &= F(t-1)°R(t, t-1) \\ F(t-1) &\rightarrow F(t) \end{aligned} \tag{2}$$

where R denotes fuzzy relation and this relation is called first order fuzzy time series of $F(t)$. If $F(t) = f_i$ and If $F(t-1) = f_j$, then the fuzzy relation can be defined as $f_j \rightarrow f_i$. Where f_j is the left hand side and f_i the right hand side of fuzzy relation.

Definition 3: Fuzzy relations which have the same left-hand sides are grouped. For example, If $f_j \rightarrow f_{i_1}$ and $f_j \rightarrow f_{i_2}$ and ,…, and $f_j \rightarrow f_{i_k}$ then fuzzy relations are reorganized as If $f_j \rightarrow f_{i_1}, f_{i_2}, \dots, f_{i_k}$.

2.2. Fuzzy Time Series Models Based on Fuzzy Clustering

FTS models consist of three main steps as fuzzification, determining of fuzzy relations and defuzzification. Fuzzification step can be performed with two different ways: i) partitioning universe of discourse [26-27] and ii) fuzzy clustering. In this study, FTS models based on three different fuzzy clustering algorithms [33-35] are used and the wavelet-based versions of these FTS models are developed. All of the FTS models used in this study are the first-order FTS.

Fuzzy clustering algorithms used in this study are based on minimizing the following objective function:

$$J(\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{C}) = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^m \; \left\| y_i - c_j \right\|^2 \tag{3}$$

where **U** is the membership degree matrix, **C** is the cluster center matrix, u_{ij} is the membership degree of ith observation to j^{th} cluster, y_i is i^{th} observation of classical time series, n is length of time series, k is number of clusters, m is the fuzziniess index and $\|\quad\|$ is distance measure.

Fuzzy clustering algorithms used in the studies of are Gustafson-Kessel(GK) [36] and Fuzzy C-Means(FCM)[37] and Fuzzy K-Medoids (FKM)[38] clustering algorithms respectively. These algorithms are difference from each other according to the distance measures used and the cluster centers. In order to calculate the distance between the cluster centers and the observations, FKM and FCM use the Euclidian distance measure, while GK uses thecluster-specific Mahalanobis distance. These distance measures are given in Equations (4) and (5) respectively:

$$\|y_i - c_j\| = \sqrt{(y_i - c_j)^2} \tag{4}$$

$$\|y_i - c_j\| = \sqrt{(y_i - c_j)^T \Sigma_i^{-1}(y_i - c_j)} \tag{5}$$

where Σ_i denotes the covariance matrix of jth cluster. Cluster centers are calculated as follows:

$$c_j = \frac{\sum_{i=1}^{n}(u_{ij})^m y_i}{\sum_{i=1}^{n}(u_{ij})^m} \qquad j = 1,2,\dots,k \tag{6}$$

$$c_j = argmin_{1 \le z \le n} \sum_{i=1}^{n} u_{ij}^m \|y_z - y_i\|^2 \quad j = 1,2,\dots,k \tag{7}$$

where FCM and GK use the Equation (6), FKM uses the Equation (7). Lastly, membership degrees for all clustering algorithms are calculated as follows:

$$u_{ij} = \sum_{t=1}^{k} \left(\frac{\|y_i - c_j\|}{\|y_i - c_t\|}\right)^{\frac{-2}{m-1}} \quad i = 1,2,\dots,n, \qquad j = 1,2,\dots,k. \tag{8}$$

Fuzzy clustering algorithms are iterative processes. In the first step, the initial values such as number of clusters(k), fuzziness index(m), etc. are determined. Step 2 contains calculating of matrix of membership degrees (**U**) by using Equation (8). According to new membership degrees, cluster centers(**C**) are computed in Step 3. This process is repeated until distance between two successive cluster centers is smaller than a predetermined termination criteria (ε). After terminating the clustering algorithm, fuzzy sets are determined by using following equation:

$$f_i = f_{\max_{j=1,2,\dots,k} u_{ij}} \quad i=1, 2,\dots,n \tag{9}$$

This equation also performs fuzzification process and in consequence of this process, fuzzy time series (**F**) is obtained.

In the second step of FTS models, fuzzy relations are determined. For example, let be obtained a fuzzy time series as $f_1, f_1, f_2, f_3, f_3, f_5, f_4, f_5$ from fuzzification. In this case, the fuzzy relations between successive fuzzy sets are as follows:

$$f_1 \to f_1 \qquad f_2 \to f_3 \qquad f_3 \to f_3 \qquad f_5 \to f_4 \qquad f_5 \to f_4$$

$$f_1 \to f_2 \qquad\qquad\qquad f_3 \to f_5 \quad .$$

When fuzzy relations are grouped:

$$f_1 \rightarrow f_1, f_2 \qquad f_2 \rightarrow f_3 \quad f_3 \rightarrow f_3, f_5 \quad f_5 \rightarrow f_4 \qquad f_5 \rightarrow f_4 \quad .$$

The last step of the FTS can be summarized as follows:

Let be $F(t-1) = f_i$. Three cases exist for forecasting and defuzzification,
Firstly, fuzzy predictions and forecasts ($\tilde{f}(i)$) are obtained by taking into following three cases,
<u>Case 1:</u> If $f_i \rightarrow f_j$, fuzzy prediction and forecast correspond to f_j,
<u>Case 2:</u> If $f_i \rightarrow f_j, f_k, f_l$ fuzzy prediction and forecast are f_j, f_k, f_l,
<u>Case 3:</u> If $f_i \rightarrow \emptyset$ fuzzy prediction and forecast is f_i.

After fuzzy prediction and forecasts are obtained, these values are transformed into classical values, in other words, crisp predictions and forecasts are obtained. This process is called as defuzzification. Center method is generally used for defuzzification:

<u>Case 1:</u> If $f_i \rightarrow f_j$, classical prediction or forecast ($\breve{y}(i)$) is center of the fuzzy set f_j (c_j),
<u>Case 2:</u> If $f_i \rightarrow f_j, f_k, f_l$ classical prediction or forecast is the arithmetic mean of centers of fuzzy sets f_j, f_k and f_l (($c_j + c_k + c_l$)/3),
<u>Case 3:</u> If $f_i \rightarrow \emptyset$ classical prediction or forecast is the center of fuzzy set f_i (c_i).

2.3. Maximal Overlap Discrete Wavelet Transformation

Wavelet transform (WT) is a data transformation technique used for decomposing a time series into the frequency components and it is based on expressing the time series as linear combination of basis functions, called wavelets. As mentioned in Introduction section, WT has some advantages in time series analysis. First, WT has the ability of analyzing the time series both in time and frequency domain. Thus, time information is not lost [28]. Second, the each of subseries obtained from WT represents the different behavior of the time series. Hence, WT enables the analyzing of the time series in more detail. Third, WT can detect the noise, the outlier and the unwanted data points and can reduce the negative effects of these data points on the modelling process. In time series prediction and forecasting, the most appropriate WT technique is MODWT [39] thanks to its some advantages such as: i) MODWT does not require down-sampling (decimation) process and thus, it can be applied to the time series with any sample size. There is no need that the length of time series is power of two in contrast to DWT, ii) at each level of the decomposition MODWT yields the wavelet coefficients as much as the length of original time series and iii) MODWT coefficients are not affected by the wavelet filter used and the starting point of time series [40].

In fact, MODWT is modified version of the DWT. To construct MODWT, wavelet ($\tilde{h}_{j,l}$) and scaling filter ($\tilde{g}_{j,l}$) are defined by rescaling the filters of DWT as follows:

$$\tilde{h}_{j,l} = \frac{h_{j,l}}{2^{\frac{j}{2}}} \, and \, \tilde{g}_{j,l} = \frac{g_{j,l}}{2^{\frac{j}{2}}} \tag{10}$$

where $h_{j,l}$ and $g_{j,l}$ denote the wavelet and scaling filters of DWT. MODWT wavelet and scaling coefficients are obtained by convolving original time series with these filters. The following equations define the MODWT coefficients [11,41]:

$$W_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{t-l \bmod n} = \sum_{l=0}^{n-1} h_{j,l}^o Y_{t-l \bmod n} \tag{11}$$

$$V_{j,t} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} Y_{t-l \bmod n} = \sum_{l=0}^{n-1} g_{j,l}^o Y_{t-l \bmod n} \ . \tag{12}$$

In these equations, $L_j = (2^j - 1)(L - 1) + 1$, L is the length of filter, n is the length of the time series, $h_{j,l}^o$ and $g_{j,l}^o$ are $\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$ periodized to length n. Equations (11) and (12) also can be written in matrix notation as follows:

$$\boldsymbol{W_j} = \boldsymbol{\omega_j^T Y} \tag{13}$$

$$\boldsymbol{V_{J0}} = \boldsymbol{v_j^T Y} \tag{14}$$

where, J_0 indicates the decomposition level and its maximum value must satisfy following condition:

$$J_0 \le \log_2(n). \tag{15}$$

$\boldsymbol{w_j}$ and $\boldsymbol{v_j}$ also can be given as matrix form:

$$\boldsymbol{\omega_j} = \frac{1}{2^j} \begin{bmatrix} \tilde{h}_{j,0} & \tilde{h}_{j,n-1} & \tilde{h}_{j,n-2} & \tilde{h}_{j,n-3} & \cdots & \tilde{h}_{j,3} & \tilde{h}_{j,2} & \tilde{h}_{j,1} \\ \tilde{h}_{j,1} & \tilde{h}_{j,0} & \tilde{h}_{j,n-1} & \tilde{h}_{j,n-2} & \cdots & \tilde{h}_{j,4} & \tilde{h}_{j,3} & \tilde{h}_{j,2} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \tilde{h}_{j,n-2} & \tilde{h}_{j,n-3} & \tilde{h}_{j,n-4} & \tilde{h}_{j,n-5} & \cdots & \tilde{h}_{j,1} & \tilde{h}_{j,0} & \tilde{h}_{j,n-1} \\ \tilde{h}_{j,n-1} & \tilde{h}_{j,n-2} & \tilde{h}_{j,n-3} & \tilde{h}_{j,n-4} & \cdots & \tilde{h}_{j,2} & \tilde{h}_{j,1} & \tilde{h}_{j,0} \end{bmatrix} \tag{16}$$

$$\boldsymbol{v_j} = \frac{1}{2^j} \begin{bmatrix} \tilde{g}_{j,0} & \tilde{g}_{j,n-1} & \tilde{g}_{j,n-2} & \tilde{g}_{j,n-3} & \cdots & \tilde{g}_{j,3} & \tilde{g}_{j,2} & \tilde{g}_{j,1} \\ \tilde{g}_{j,1} & \tilde{g}_{j,0} & \tilde{g}_{j,n-1} & \tilde{g}_{j,n-2} & \cdots & \tilde{g}_{j,4} & \tilde{g}_{j,3} & \tilde{g}_{j,2} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \tilde{g}_{j,n-2} & \tilde{g}_{j,n-3} & \tilde{g}_{j,n-4} & \tilde{g}_{j,n-5} & \cdots & \tilde{g}_{j,1} & \tilde{g}_{j,0} & \tilde{g}_{j,n-1} \\ \tilde{g}_{j,n-1} & \tilde{g}_{j,n-2} & \tilde{g}_{j,n-3} & \tilde{g}_{j,n-4} & \cdots & \tilde{g}_{j,2} & \tilde{g}_{j,1} & \tilde{g}_{j,0} \end{bmatrix} \ . \tag{17}$$

In order to reconstruct original time series, inverse MODWT defined as below are used:

$$\boldsymbol{Y} = \sum_{j=1}^{J_0} \boldsymbol{\omega_j^T W_j} + \boldsymbol{v_j^T V_{J0}} \tag{18}$$

where $\boldsymbol{D_j} = \boldsymbol{\omega_j^T W_j}$ is j^{th} detail component of the original time series, which capture the details such as noise and random fluctuation and $\boldsymbol{A_{J_0}} = \boldsymbol{v_j^T V_{J0}}$ is the approximation component, which represents the general structure of the time series such as trend and seasonality. From here, reconstructed time series can be rewritten as below:

$$\boldsymbol{Y} = \sum_{j=1}^{J_0} \boldsymbol{D_j} + \boldsymbol{A_{J_0}} \ . \tag{19}$$

In MODWT, the most important parameter that need to be decided beforehand is the type of wavelet filter. There are several types of wavelet filters in the literature such as Haar, Daubechies, Coiflet, Symlet etc. The simplest form of the wavelet filter is Haar and it is defined as follows:

$$\tilde{h}_{j,l} = \begin{cases} \dfrac{1}{2^j} & for\ l = 0, \dots 2^{j-1} - 1 \\ \dfrac{1}{2^j} & for\ l = 2^{j-1}, \dots, 2^j - 1 \end{cases} , \tag{20}$$

$$\tilde{g}_{j,l} = \begin{cases} \dfrac{1}{2^j} & for\ l = 0, \dots 2^{j-} - 1 \\ 0 & otherwise \end{cases} . \tag{21}$$

Haar type wavelet filter is preferred in this study.

3. PROPOSED TIME SERIES PREDICTION MODEL

Proposed FTS model consists of three main steps as given in Figure 1.
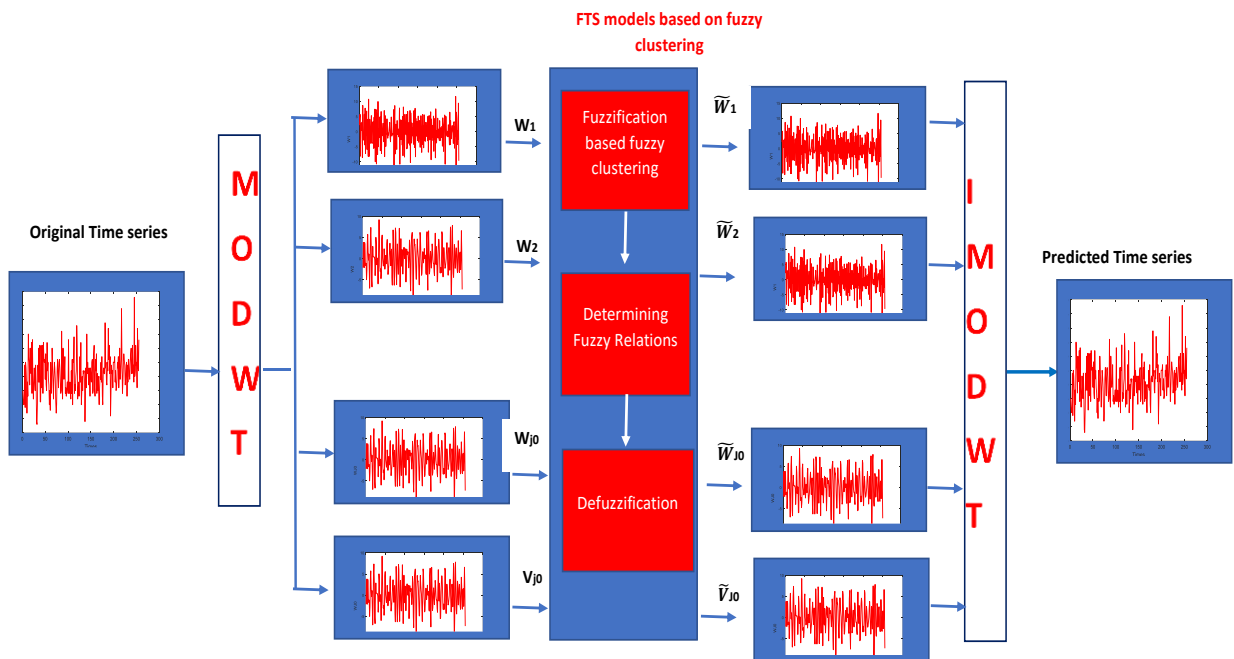

Figure 1. The flowchart of the proposed method

As can be seen in Figure 1, time series firstly is decomposed into sub time series $(W_1, W_2, \dots, W_{J_0}, V_{J_0})$ by using MODWT. In the second step, FTS models based fuzzy clustering are constructed by performing steps of fuzzification, determining of fuzzy relations and defuzzification given in Section 2.2 for each sub time series, separately. In the end of this step, predicted values of sub time series $(\widetilde{W}_1, \widetilde{W}_2, \dots, \widetilde{W}_{J_0}, \widetilde{V}_{J_0})$ are obtained. The last step consists of the reconstruction of predicted time series $(\widetilde{Y})$. For this objective, following equation is used:

$$\widetilde{Y} = \sum_{j=1}^{J_0} \omega_j^T \widetilde{W}_j + v_j^T \widetilde{V}_{J0} \tag{22}$$

where $\widetilde{W}_j s$ (j=1, 2,…, J_0) and $\widetilde{V}_{J0}$ are predicted wavelet and scaling coefficients respectively.

4. EXPERIMENTAL RESULTS

4.1. Data Sets

We conducted experiments on eight real time series described in Table 1.

***Table 1.** The time series used in this study*

Time Series	Description	Length	Time unit	Period
SS	Shampoo Sales [42]	36	Monthly	01/1981-12/1983
MDT	Minimum Daily Temperatures [43]	3650	Daily	1891-1990
DFB	Daily Female Births [44]	365	Daily	01/01/1959-31/12/1959
MS	Monthly Sunspot [44]	2820	Monthly	1749-1983
MT	Mean Temp in Delhi [45]	1463	Daily	01/01/2013-01/01/2017
H	Humidity in Delhi [45]	1463	Daily	01/01/2013-01/01/2017
WS	Wind Speed in Delhi [45]	1463	Daily	01/01/2013-01/01/2017
P	Pressure in Delhi [45]	1463	Daily	01/01/2013-01/01/2017

The line plots of the data sets are given in Figure 2.
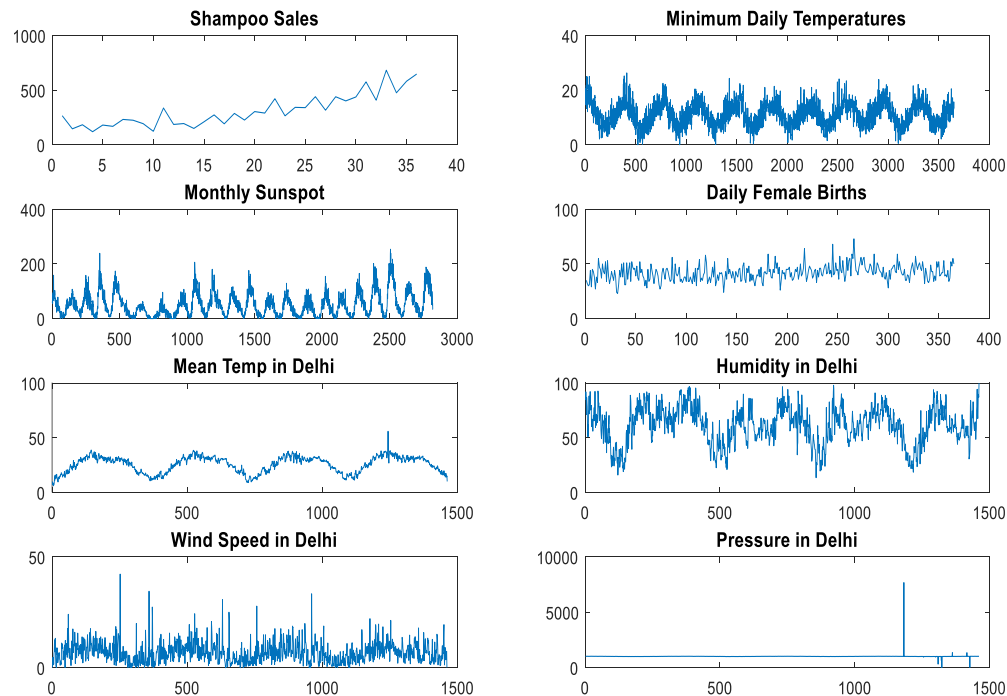
***Figure 2.** Time series used in this study*

As can be seen in Figure 2, it is tried to use time series having different behaviors for the performance comparisons. Some properties of the data sets can be given as follows: the SS time series has the trend, the MDT time series includes the regular seasonality, H and S time series include the varying seasonality, DFT is a more stationary time series when comparing with the others, P data set contains the noise and the

outliers and WS and MT time series include both noise and seasonality. Each time series is divided into two distinct subsets as training and test set. The first 70% of all data sets except from P data set are selected as the training set, the remaining part are selected as test set. The rate of training set is adjusted as 90% in P dataset. The reason of this is to ensure that outliers located in the last time period are included in the training set. Here, training set is used to predict FTS models and test set is used to evaluate of the performance of predicted FTS models. For wavelet-based versions (WFTS), the decomposition level is calculated as $round\ (log_2(n_{test})$ (Where n_{test} corresponds to the length of test set) and *haar* wavelet filter is used.

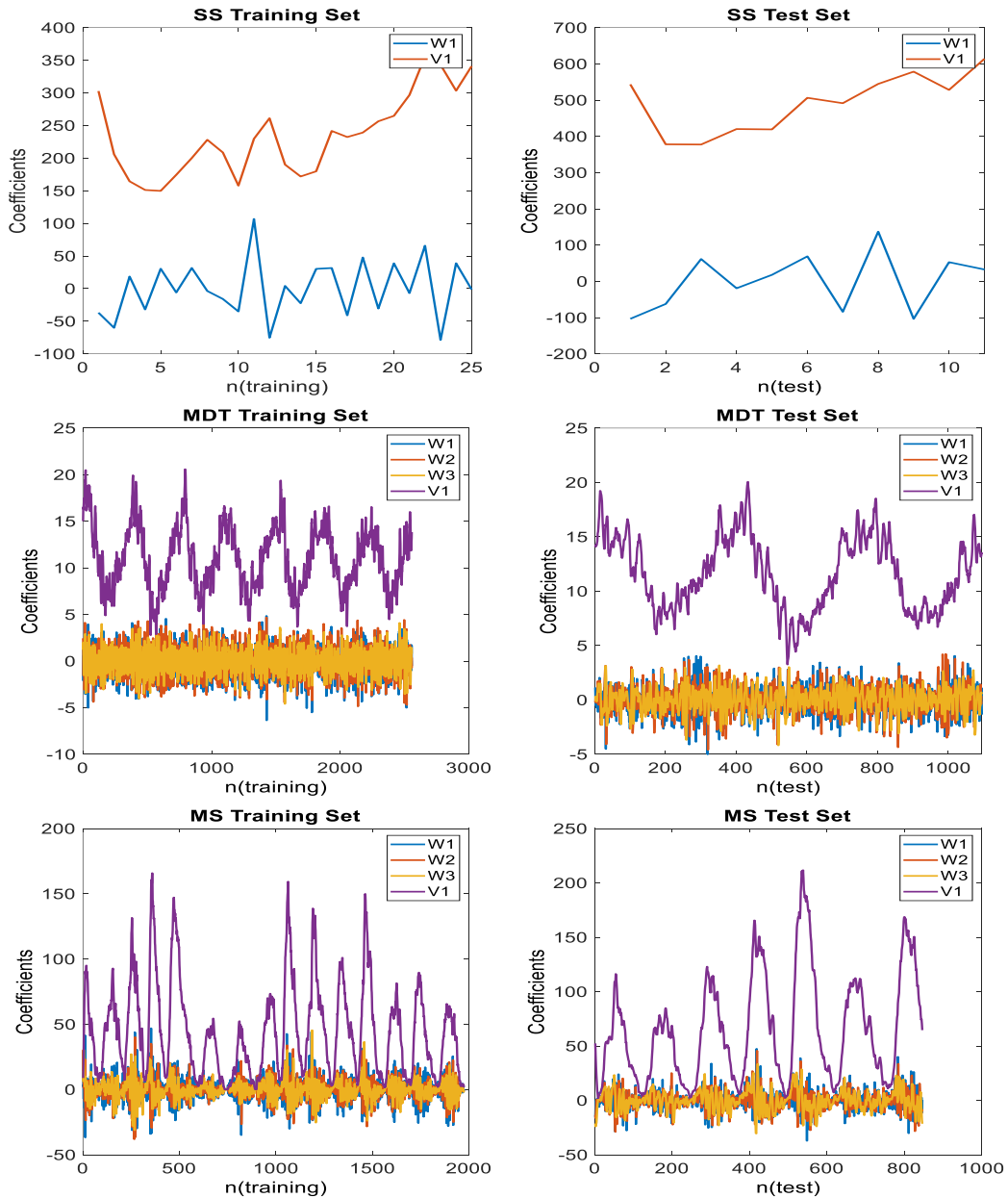The MODWT coefficients are given in Figure 3.

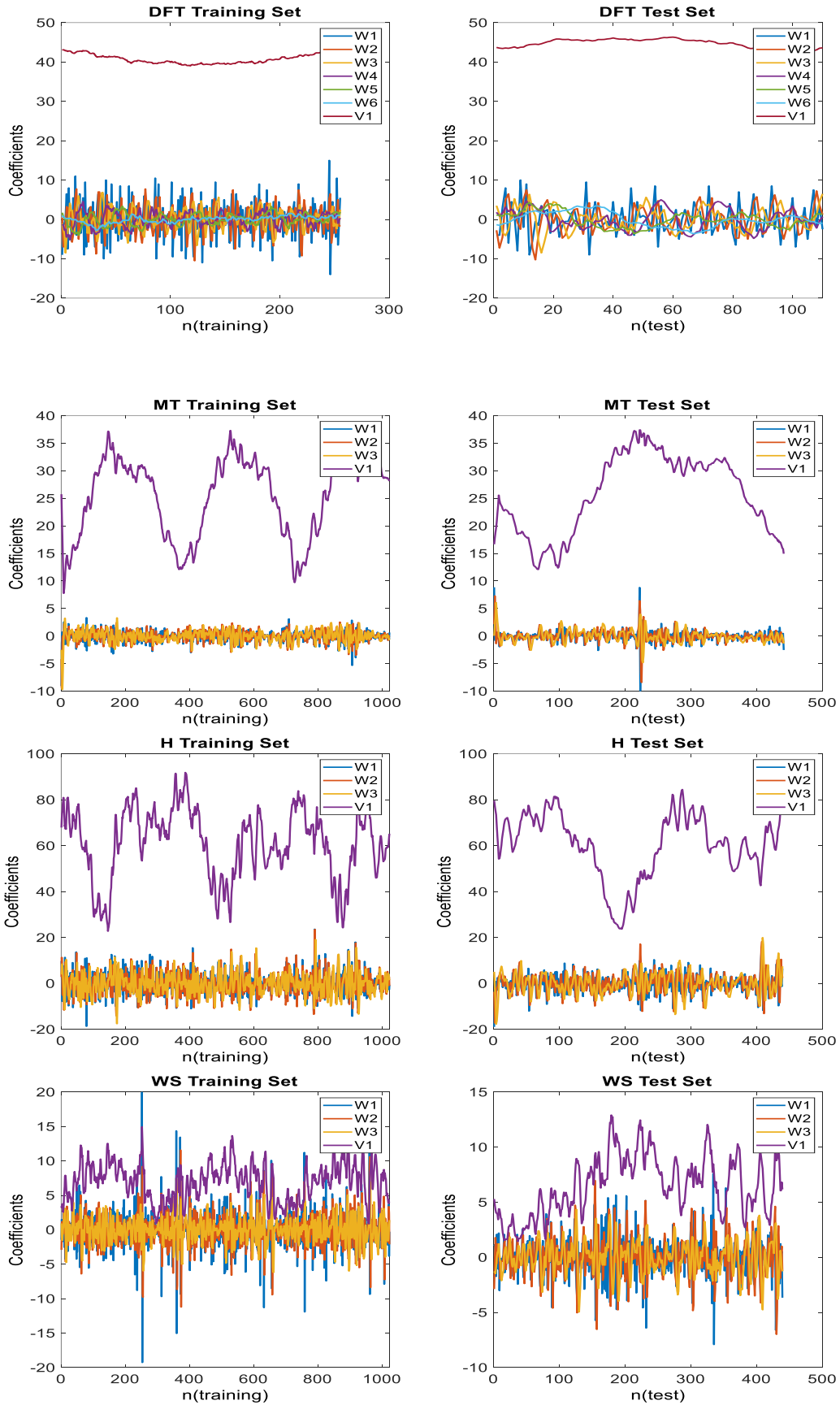***Figure 3.*** *MODWT coefficients for all data sets*

Figure 3. *(Continue)*
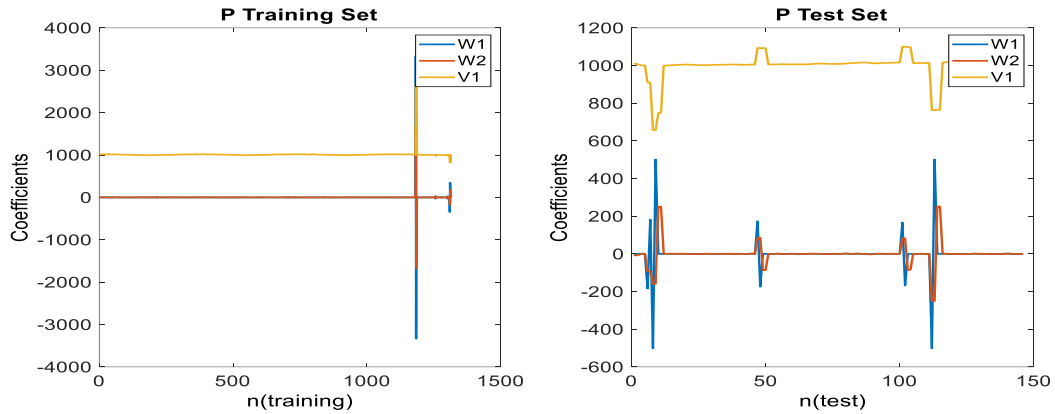
Figure 3. *(Continue)*

4.2. Evaluation Metrics

The mean performances of all FTS models are evaluated by using four goodness of fit measures (GoFs) consisted of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized Root Mean Square Error(NRMSE) and Variance Account For (VAF). Their definitions are expressed as follows:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} x100 \tag{23}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{24}$$

$$NRMSE = \frac{RMSE}{y_{max}} \tag{25}$$

$$VAF = (1 - \frac{var(Y - \hat{Y})}{var(Y)})x100 \tag{26}$$

where y_i is ith actual value, $\hat{y}_i$ is ith predicted value, n is the length of time series, y_{max} is the maximum of actual values, and *var* indicates the variance. The small values of MAE, RMSE and NRMSE and high values of VAF denote the good performance. After GoF measures are calculated for each TFTS and WFTS, improvement percent (IP), indicating how much WFTS provides an improvement are calculated as follows:

$$IP = \left(1 - \frac{GoF_{WFTS}}{GoF_{TFTS}}\right) * 100 . \tag{27}$$

4.3. Comparison Results

In this section, the prediction and forecasting performances of three TFTS models based on FCM (**FF**), GK (**GF**) and FKM (**FKF**) clustering respectively [33-35] and of their wavelet-based versions (**WFF, WGF, WFKF**) are compared. Figures 4 and 5 show the actual and predicted values for training and test sets, respectively.
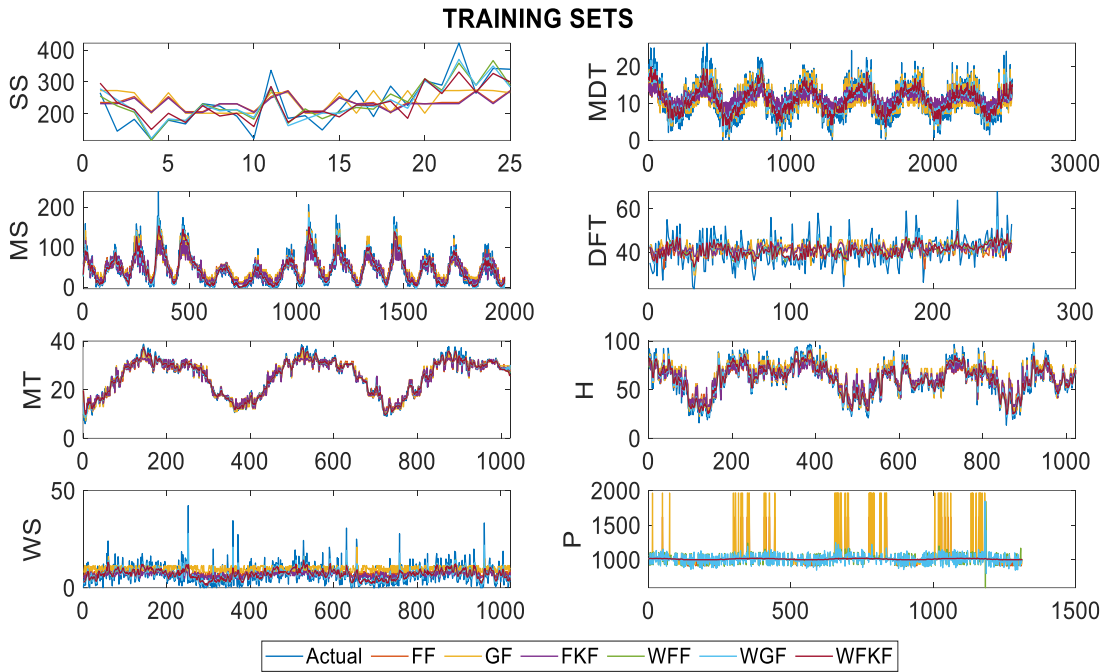
Figure 4. *Actual and predicted values for training sets*
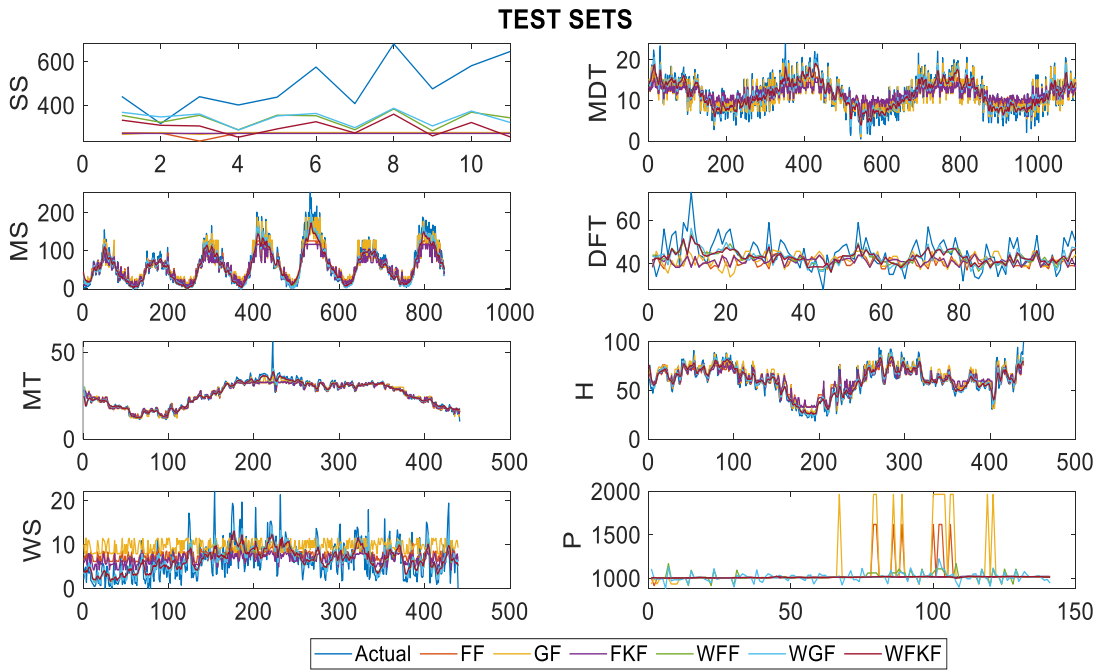

Figure 5. *Actual and predicted values for test sets*

When looking at the Figures 4 and 5, it can be seen that the predicted and forecasting values obtained from all FTS models exhibit the similar behavior with the actual values. The GoF measures calculated with aim of understanding better the difference between the performance of TFTS and WFTS models are given in Tables 2 and 3.

Table 2. *GoF values for training sets*

GoF	Model	SS	MDT	MS	DFB	MT	H	WS	P
	FF	59.98	2.18	13.52	5.37	1.48	6.90	3.39	29.44
	GF	61.28	2.13	14.05	5.49	1.41	7.02	4.50	67.05
	FKF	59.38	2.38	13.04	5.43	1.46	7.09	3.21	2.50
MAE	**WFF**	**26.11**	**1.62**	**8.01**	**4.29**	**0.98**	**4.96**	**2.41**	**26.04**
	WGF	**27.95**	**1.52**	**7.73**	**4.12**	**0.96**	**4.74**	**2.25**	**37.52**
	WFKF	**31.88**	**1.63**	**8.37**	**4.33**	**0.99**	**4.98**	**2.50**	**1.80**
	FF	71.13	2.79	17.68	6.80	1.92	8.75	4.46	121.00
	GF	71.48	2.71	17.95	6.83	1.81	8.78	5.46	244.03
RMSE	FKF	71.13	3.01	18.34	6.85	1.89	8.89	4.43	3.00
	WFF	**33.94**	**2.06**	**11.34**	**5.28**	**1.30**	**6.38**	**3.38**	**58.43**
	WGF	**36.72**	**1.93**	**10.66**	**5.04**	**1.24**	**6.09**	**3.02**	**59.27**
	WFKF	**39.23**	**2.09**	**11.87**	**5.31**	**1.34**	**6.42**	**3.56**	**2.28**
	FF	0.17	0.11	0.07	0.10	0.05	0.09	0.11	0.12
NRMSE	GF	0.17	0.10	0.08	0.10	0.05	0.09	0.13	0.24
	FKF	0.17	0.11	0.08	0.10	0.05	0.09	0.10	0.00
	WFF	**0.08**	**0.08**	**0.05**	**0.08**	**0.03**	**0.07**	**0.08**	**0.06**
	WGF	**0.09**	**0.07**	**0.04**	**0.07**	**0.03**	**0.06**	**0.07**	**0.06**
	WFKF	**0.09**	**0.08**	**0.05**	**0.08**	**0.03**	**0.07**	**0.08**	**0.00**
	FF	10.40	53.34	78.53	7.68	93.34	74.02	14.55	-24763.52
	GF	10.21	56.46	78.23	8.40	94.02	75.86	2.19	-97573.92
VAF	FKF	10.50	46.11	76.73	6.73	93.53	73.37	13.08	92.76
	WFF	**79.62**	**74.56**	**91.05**	**44.39**	**96.92**	**86.09**	**49.39**	**-5730.03**
	WGF	**76.14**	**77.71**	**92.09**	**49.33**	**97.21**	**87.34**	**59.60**	**-6013.45**
	WFKF	**72.83**	**73.92**	**90.18**	**43.67**	**96.73**	**85.93**	**43.90**	**93.91**

When looking at the Table 2, it can be seen that MAE, RMSE and NRMSE values of WFTS models are smaller than those of TFTS models for all training sets. Besides, VAF values obtained from WFTF models are bigger than traditional ones. Thus, it can be said that WFTS models outperform than their traditional equivalents for the training sets. As mentioned before, P data set includes the outliers. FKF model is a method, which is robust against to outlier and noise data points. Therefore, FKF and WFKF models give the similar prediction result for P data set. Table 3 gives the GoF measures for test sets.

Table 3. *GoF values for test sets*

GoF	Model	SS	MDT	MS	DFB	MT	H	WS	P
	FF	221.68	2.16	17.03	6.64	1.55	6.24	3.12	36.41
	GF	217.95	2.04	16.12	6.90	1.40	6.26	4.37	98.50
	FKF	220.37	2.36	18.30	6.30	1.51	6.36	2.89	1.94
MAE	**WFF**	**157.07**	**1.50**	**12.52**	**4.48**	**1.06**	**4.64**	**2.07**	**22.08**
	WGF	**155.16**	**1.41**	**11.49**	**4.20**	**1.04**	**4.34**	**2.02**	**34.10**
	WFKF	**192.91**	**1.50**	**13.11**	**4.38**	**1.05**	**4.63**	**2.13**	**1.87**
	FF	246.44	2.68	19.84	8.80	2.25	8.10	3.87	144.38
	GF	243.27	2.57	19.05	8.75	2.01	8.04	5.16	300.01
RMSE	FKF	245.75	2.91	21.32	8.30	2.19	8.05	3.69	**2.80**
	WFF	**181.82**	**1.93**	**13.54**	**5.61**	**1.59**	**6.02**	**2.71**	**41.74**
	WGF	**180.04**	**1.81**	**12.42**	**5.23**	**1.59**	**5.76**	**2.60**	**50.27**
	WFKF	**218.84**	**1.94**	**14.32**	**5.60**	**1.63**	**6.05**	**2.79**	3.05
	FF	0.36	0.11	0.08	0.12	0.04	0.08	0.18	0.14
	GF	0.36	0.11	0.08	0.12	0.04	0.08	0.23	0.29
NRMSE	FKF	0.36	0.12	0.08	0.11	0.04	0.08	0.17	0.003
	WFF	**0.27**	**0.08**	**0.05**	**0.08**	**0.03**	**0.06**	**0.12**	**0.04**
	WGF	**0.26**	**0.08**	**0.05**	**0.07**	**0.03**	**0.06**	**0.12**	**0.05**
	WFKF	**0.32**	**0.08**	**0.06**	**0.08**	**0.03**	**0.06**	**0.13**	**0.002**
	FF	2.00	55.27	78.74	-17.90	90.91	74.43	18.09	-60342.02
	GF	1.25	58.97	82.66	-28.91	92.57	76.85	-0.36	-250312.31
VAF	FKF	0.00	46.86	75.95	-9.54	91.41	75.23	18.46	**77.52**
	WFF	**26.13**	**76.76**	**90.69**	**45.49**	**95.39**	**85.74**	**55.59**	**-5013.61**

WGF	15.21	79.49	92.43	52.23	95.36	86.94	59.16	-7629.81
WFKF	9.74	76.47	88.61	45.04	95.14	85.61	52.98	71.92

From Table 3, it can be seen that all RMSE, MAE and NRMSE values of WFTS models are smaller than those of TFTS and the all VAF values of WFTS except from P data set are bigger than those of FTS. Thus, it can be easily said that WFTS models provide the best forecasting performances. The arithmetic means of NMRSE values of eight time series are calculated to compare mean success of WFTS and TFTS models. Figure 6 shows the arithmetic mean of NRMSE values.
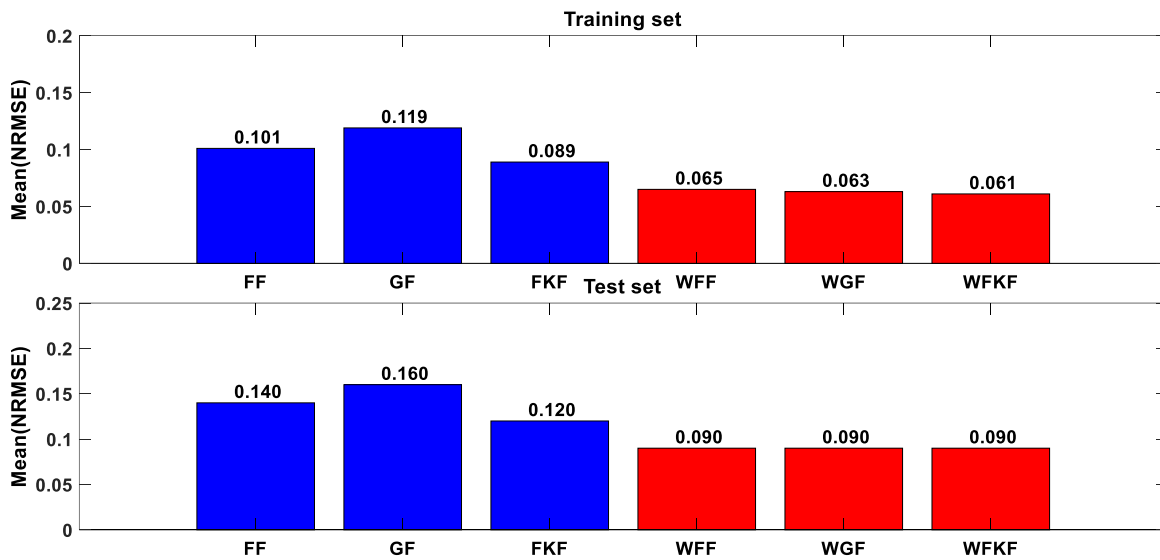

Figure 6. *Arithmetic means of NRMSE values*

When looking at Figure 6, it is seen that the arithmetic means of WFTS models are smaller than those TFTS for both training and test sets. From here, it can be concluded that the WFTS models improve the forecasting and prediction performance. With aim of testing this improvement is statistically significant or not, Wilcoxon Signed-Rank Test is performed. The test is only performed for MAE values since the similar results are also obtained for the other GoF measures. Table 4 gives the Wilcoxon Signed-Rank test results.

Table 4. *Wilcoxon Signed-Rank Test Results*

Data set	Pairs	Negative Rank	Positive Rank	p-value
Training	WFF-FF	8	0	0.012
	WGF-GF	8	0	0.012
	WFKF-FKF	8	0	0.012
Test	WFF-FF	8	0	0.012
	WGF-GF	8	0	0.012
	WFKF-FKF	8	0	0.012

In Table 4, the column of negative rank demonstrates the number of time series, which WFTS give the best performance. According to this, WFTS models outperform the TFTS models for all time series. When looking at the column of p-value, it can be seen that all p-values are smaller than 0.05. This means that the difference between performances of TFTS and of WFTS models are statistically significant and that WFTS models provide the significant improvement in prediction and forecasting. IP values are given in Table 5 for MAE, RMSE and NRMSE.

Table 5. *IP values for each data set*

	GoF	Pair	SS	MDT	MS	DFB	MT	H	WS	P	Mean
Training		WFF/FF	56.47	24.40	40.75	20.11	33.78	28.12	28.91	11.55	**30.51**
		WGF/GF	54.39	29.38	44.98	24.95	31.91	32.48	50.00	44.04	**39.02**
	MAE	WFKF/FKF	46.31	29.33	35.81	20.26	32.19	29.76	22.12	28.00	**30.47**
		WFF/FF	52.28	26.16	35.86	22.35	32.29	27.09	24.22	51.71	**34.00**
		WGF/GF	48.63	28.78	40.61	26.21	31.49	30.64	44.69	75.71	**40.85**
	RMSE	WFKF/FKF	44.85	30.56	35.28	22.48	29.10	27.78	19.64	24.00	**29.21**
		WFF/FF	52.94	27.27	28.57	20.00	40.00	22.22	27.27	50.00	**33.54**
		WGF/GF	47.06	30.00	50.00	30.00	40.00	33.33	46.15	75.00	**43.94**
	NRMSE	WFKF/FKF	47.06	27.27	37.50	20.00	40.00	22.22	20.00	33.33	**30.92**
Test		WFF/FF	29.15	30.56	26.48	32.53	31.61	25.64	33.65	39.36	**31.12**
		WGF/GF	28.81	30.88	28.72	39.13	25.71	30.67	53.78	65.38	**37.89**
	MAE	WFKF/FKF	12.46	36.44	28.36	30.48	30.46	27.20	26.30	3.61	**24.41**
		WFF/FF	26.22	27.99	31.75	36.25	29.33	25.68	29.97	71.09	**34.79**
		WGF/GF	25.99	29.57	34.80	40.23	20.90	28.36	49.61	83.24	**39.09**
	RMSE	WFKF/FKF	10.95	33.33	32.83	32.53	25.57	24.84	24.39	**-8.93**	**21.94**
		WFF/FF	25.00	27.27	37.50	33.33	25.00	25.00	33.33	71.43	**34.73**
		WGF/GF	27.78	27.27	37.50	41.67	25.00	25.00	47.83	82.76	**39.35**
	NRMSE	WFKF/FKF	11.11	33.33	25.00	27.27	25.00	25.00	23.53	33.33	**25.45**

As can be seen from Table 5, proposed method provides an improvement at least 29.21% in prediction and at least 21.94% in forecasting. It gives the highest improvement in GKF and the lowest in FKF. But, according to RMSE, it is found that FKF is more successful in forecasting of P time series since IP is ne negative number.

5. CONCLUSIONS

In this study, a hybrid time series prediction method is proposed based on decomposing the time series into subseries by MODWT and predicting an FTS model for each subseries separately. The major advantage of proposed model is that it has the ability to include the detail behavior of the time series to modeling process and it learns the fuzzy sets from time series itself. Thus, it is aimed that the performances of TFTS models based on fuzzy clustering are improved. To validate the efficiency of proposed method, three TFTS models (FF, GF, FKF) and their proposed versions (WFF, WGF, WFKF) are compared in terms of performance of prediction and forecasting. For performance comparisons, four GoF measures (MAE, RMSE, NRMSE and VAF) and eight real time series are used. According to the results of the comparisons, it is concluded that

- According to IP values, proposed method provides an improvement at least 29.21% in prediction and at least 21.94% in forecasting
- According to Wilcoxon Signed-Rank test results, performance of proposed method is statistically significant in both the prediction and forecasting performances of TFTS models
- Proposed method provides the highest improvement in GKF, the lowest in FKF.
- Proposed method is also successful in the time series, having seasonality and containing outlier and noise.

In this study, simple form of fuzzy time series models is studied. In the future works, high-order version of proposed method and FTS models based on other transformation techniques will be developed.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1]   Box, G.E.P., Jenkins, G.M., Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, USA, (1970).

[2]   Topuz, B.K., Bozoglu, M., Baser, U., Eroglu, N. A., "Forecasting of apricot production of Turkey by using Box-Jenkins method", Turkish Journal of Forecasting. 2(2): 20-26, (2018).

[3]   Mithiya, D., Datta, L., Mandal, K., "Time series analysis and forecasting of oilseeds production in India: using autoregressive integrated moving average and group method of data handling – neural network", Asian Journal of Agricultural Extension, Economics & Sociology, 30(2): 1-14, (2019).

[4]   Galeshchuk, S., "Neural Networks performance in exchange rate prediction", Neurocomputing. 172: 446-452, (2016).

[5]   Bas, E., Egrioglu, E., Aladag, C.H., Yolcu, U., "Fuzzy time series network used to forecast linear and nonlinear time series", Applied Intelligence, 43: 343-355, (2015).

[6]   Akdeniz, E., Egrioglu, E., Bas, E., Yolcu, U., "An ARMA type pi-sigma artificial neural network for nonlinear time series forecasting", Journal of Artificial Intelligence and Soft Computing Research, 8(2): 121-132, (2018).

[7]   Jiang, P., Dong, Q., Li, P., "A novel high-order weighted fuzzy time series model and its application in nonlinear time series prediction", Applied Soft Computing, 55: 44-62, (2017).

[8]   Yong, N.K., Awang, N., "Wavelet-based time series model to improve the forecast accuracy of PM10 concentrations in Peninsular Malaysia", Environmental Monitoring and Assessment, 191(64): 1-12, (2019).

[9]   Wadi, S.A., Alsaraireh, A.A., "Industrial data forecasting using discrete wavelet transform", Italian Journal of Pure and Applied Mathematics. 40: 607-614, (2018).

[10]   Md-Khair, N.Q.N., Samsudin, R., Shabri, A., "Forecasting crude oil prices using discrete wavelet transform with autoregressive integrated moving average and least square support vector machine combination approach", International Journal on Advanced Science, Engineering and Information Technology. 7(4-2): 1553-1561, (2017).

[11]   Zhu, L., Wang, Y., Fan, Q., "MODWT-ARMA model for time series prediction", Applied Mathematical Modelling, 38: 1859-1865, (2014).

[12]   Kalteh, A.M., "Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform", Computers and Geosciences, 54: 1–8, (2013).

[13]   Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., "Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models", Journal of Hydrolgy, 54: 1-8, (2014).

[14] Lahmiri, S., "Wavelet low- and high-frequency components as features for predicting stock prices with backpropagation neural networks", Journal of King Saud University – Computer and Information Sciences", 26: 218-227, (2014).

[15] Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., "Artificial neural network forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation", Atmospheric Environment 107: 118-128, (2015).

[16] Pradhan, P.P., Subudhi, P.P., "Wind speed forecasting based on wavelet transformation and recurrent neural network", International Journal of Numerical Modelling, (2019).   DOI: https://doi.org/10.1002/jnm.2670

[17] Seo, Y., Kim, S., Kisi, O., Singh, V., "Daily water level forecasting using wavelet decomposition and artificial intelligence techniques", Journal of Hydrology, 520: 224-243, (2015).

[18] Parmar K., Bhardwaj, R., "River waver prediction modeling using neural networks, fuzzy and wavelet coupled model", Water Resources Management, 29: 17-33, (2014).

[19] Seo, Y., Choi, Y., Choi, J., "River stage modeling by combining maximal overlap discrete wavelet transform, support vector machines and genetic algorithm", Water, 9(7): 525, (2017).

[20] Yaslan, Y., Bican, B., "Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting", Measurement, 103: 52-61, (2017).

[21] Liu, Z., Liu, J., "A robust time series prediction method based on empirical mode decomposition and high-order fuzzy cognitive maps", Knowledge-Based Systems, 203: 106105, (2020).

[22] Yang, H-F., Chen, Y.P.P., "Hybrid deep learning and empirical model decomposition for tiem series applications", Expert Systems and Applications, 120: 128-138, (2019).

[23] Yang, H-F., Chen, Y.P.P., "Representation learning with extreme learning machines and empirical mode decomposition for wind speed forecasting methods", Artificial Intelligence, 277: 103176, (2019).

[24] Chen, M-Y., Chen, B.T., "Online fuzzy time series analysis based on entropy discretization and fast fourier transform", Applied Soft Computing, 14(B): 156-166, (2014).

[25] Halliday, J.R., Dorrell, D.G., Wood, R.A., "An application of the fast fourier transform to the short-term prediction of sea wave behavior", Renewable Enegry, 36(6): 1685-1692, (2011).

[26] Basakın, E.E., Ekmekcioğlu, Ö., Özger M., Çelik, A., "Prediction of Turkey wheat yield by wavelet fuzzy time series and gray prediction methods", Türkiye Tarımsal Araştırmalar Dergisi, 7(3): 246-252, (2020).

[27] Başakın, E.E., Özger, M., "Montly river discharge prediction by wavelet fuzzy time series method", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 29(1): 17-35, (2021).

[28] Chen, C.C., Tsui, F.R., "Comparing different wavelet transform on removing electrocardiogram baseline wanders and special trends", BMC Medical Informatics and Decision Making, 343, (2020).

[29] Özel, P., Akan, A., Yılmaz, B., "Noise-Assisted Multivariate Empirical Mode Decompotision based Emotion Recognition", Electrica, 18(2): 263-274, (2018).

[30] Song, Q., Chissom, B.S., "Fuzzy time series and its models", Fuzzy Set Systems, 54: 269-277, (1993a).

[31] Song, Q., Chissom, B.S., "Forecasting enrollments with fuzzy time series-Part I", Fuzzy Set Systems 54: 1-9, (1993b).

[32] Song, Q., Chissom, B.S., "Forecasting enrollments with fuzzy time series-Part II", Fuzzy Set Systems, 62: 1-8, (1994).

[33] Li, S. T., Cheng, Y. C., Lin, S. Y., "A FCM-Based deterministic forecasting model for fuzzy time series", Computers and Mathematics with Applications, 56: 3052–3063, (2008).

[34] Egrioglu, E., Aladag, C. H., Yolcu, U., Uslu, V. R., Erilli, N. A., "Fuzzy time series forecasting method based on Gustafson-Kessel fuzzy clustering", Expert Systems with Applications, 38: 10355-10357, (2011).

[35] Guler Dincer, N., Akkus, O., "A new fuzzy clustering based on robust clustering for forecasting of air pollution", Ecological Informatics, 43: 157-164, (2018).

[36] Gustafson, E., Kessel, W., "Fuzzy clustering with a fuzzy covariance matrix", IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes, San Diego, USA, (1979).

[37] Bezdek, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, (1981).

[38] Krishnapuram, R., Joshi, A., Yi, L., "A Fuzzy relative of the k-medoids algorithm with application to document and snippet clustering", Proceedings IEEE International Conference on Fuzzy Systems. Seoul, South Korea, (1999).

[39] Percival, D.B., Walden, A.T.," Wavelet Methods for Time Series Analysis", Cambridge University Press, (2000).

[40] Elayouty, A.S.M., "Time and frequency domain statistical methods for high-frequency time series", PhD thesis, University of Glasgow, (2017).

[41] https://faculty.washington.edu/dbp/s530/PDFs/05-MODWT-2018.pdf. Access Date: 12.08.2020

[42] Makridakis, S.G., Wheelwright, S.C., Hyndman, R.J., Forecasting: Methods and Applications, John Wiley & Sons: New York, (1998).

[43] https://datamarket.com/data/license/0/default-open-license.html. Access Date: 12.08.2020

[44] https://machinelearningmastery.com/time-series-datasets-for-machine-learning/. Access Date: 13.08.2020

[45] https://www.kaggle.com. Access Date: 12.08.2020