Comparative Study of Wavelet-SARIMA and EMD-SARIMA for Forecasting Daily Temperature Series

Luwam Ghide, Siyuan Wei, Yiming Ding*

Center for Mathematical Sciences and Department of Mathematics,

Wuhan University of Technology, Wuhan 430070, China

*Corresponding author: dingym@whut.edu.cn

ABSTRACT. This paper aims to find a forecasting model based on the comparative study of wavelet-ARIMA and EMD-ARIMA models and residual-based model selection technique for temperature time series. Time series analysis is essential in studying temperature data for investigating the variation and predicting the future trend, in which we can control the changes and make good decisions. And most important is to understand the trend in the series with time. This study applied hybridized models of wavelet transform and empirical mode decomposition with seasonal autoregressive integrated moving average (SARIMA), which combines two models to get better accuracy, for forecasting daily average temperature time series data in the central region of Eritrea, Asmara. Daily data was collected for 30 years, from January 1, 1991, to December 31, 2020. The study compares WT-SARIMA and EMD-SARIMA models to find a well fit and better forecasting model. Model selection techniques are essential for time series analysis to determine which model best fits our data. AIC and BIC are the most used methods in model selection. This paper uses an additional method based on the residual series. In estimating accurate parameters, the structure of the residual sequence had a lot of connection, in which a stationary residual depict an accurate estimation. From this perspective, a nonstationarity measurement of the residual series is used for model selection. The relative performance is based on the predictive capability of sample forecasts assessed. The results indicate that the hybridized wavelet-SARIMA model is more effective than the other models, and MATLAB soft-wire is used for this analysis.

1.Introduction

Climate change is one of the most critical issues in our time. Earth's climate has changed throughout history and threatens the lives and livelihood of billions of people. Earth-orbiting satellites and additional technological advances have enabled scientists to see the big picture by collecting different information about our planet and its climate on a global scale. The collected data over many years show the signals of climate change. Research conducted in 2018 [1] says that climate change is the most severe environmental problem. It directly affects human activities and property. Precipitation, melting ice, drought, poverty, river drying, and reducing the number and distribution of groundwater resources are some of the climate change signals. Temperature is one of the main climate change elements. Increasing or decreasing in temperature will affect the weather pattern in the lives of plants and animals. Understanding and predicting the future courses of temperature quantities based on the historical time series data is fundamental in climate change to know the pattern and increase in the frequency and magnitude of extreme events [2]–[5]. In addition, a good forecast with minimum error support for future preparation and good decision-making. Most time series forecasting methods are based on analyzing the past observed data by assuming that the past patterns in the data can be used to forecast future events.

The autoregressive integrated moving average (ARIMA) model is one of the most popular time series modelling methods. The main aim of the ARIMA model is to cautiously and rigorously study the past observations of a time series to develop a suitable model which can predict future value. Over the past years, the ARIMA model has been widely used in temperature time series forecasting. For annual temperature prediction in Libya, [6] study showed that the linear ARIMA model and the quadratic ARIMA model had the best performance in making short-term predictions. The study [32] used the ARIMA model for 50 year time period (1955-2005) in the south of Iran, and the ARIMA model was selected as the optimal model for temperature data. Furthermore, [7] used the seasonal autoregressive integrated moving average (SARIMA) model to forecast the monthly mean of the maximum surface air temperature of India, and [8] used SARIMA for monthly mean temperature in Nanjing, China. In all these studies, the selected models had higher prediction accuracy. But a disadvantage of the ARIMA model, it assumes that data are stationary and has a limited ability to capture nonstationarity and nonlinear data. ARIMA model cannot handle noisy time series data without preprocessing it. So removing the noise before applying the ARIMA model makes it easier in the modelling process to capture the features of the actual series. Different

papers have introduced the hybridized approach (combination of other methods to model to get more accurate results) of decomposition models and ARIMA model to get a well-fitted model with good prediction performance of nonstationary series, [9]–[13]. Decomposition methods decompose the main time series into stationary subcomponents, which help us to see the detailed structure of the series and improve the model's prediction performance by removing noisy components. Wavelet transform (WT) and Empirical mode decomposition (EMD) based ARIMA model achieve better prediction accuracy than direct applying ARIMA model to our time series.

A comparative study in Bangladesh [9] used wavelet decomposition with the ARIMA and artificial neural network models for temperature prediction. In the study, the hybrid WT-ARIMA model result was effective. This technique supports the ARIMA model to fit the data structure, but it also has progressed in forecasting performance [10]–[12][34]. Wavelet analysis has beautiful and deep mathematical properties, making them a well-adapted tool for different data types. It has much attention in signal processing since its theoretical development in 1984 by Grossman & Morlet. The studies of climate changes using wavelet analysis have received much consideration and applied in detecting climate signals [13]–[15]. In theory and applications, wavelet transforms enormously relates to Fourier Transform. Still, the advantage of using WT is that it offers a simultaneous localization of output in frequency and time domain with faster computation. The most significant benefit is separating fine details in the signal using small wavelets [16]. As a result, several applied fields are making use of wavelets such as[17], [18], [33].

Empirical Mode Decomposition (EMD) is a data-adaptive time-frequency representation technique proposed by Huang. It requires only that the decomposed component consists of a simple intrinsic mode of oscillations [19], which makes it quite different from WT. This method is most suitable for nonlinear and nonstationary time series data. The EMD methodology is based on a sifting process, which identifies local maxima and minima and results in intrinsic mode functions (IMFs). In detail, it decomposes the time series into a finite sum of intrinsic mode functions [20]. EMD decomposes a nonstationary signal into several stationary series components. And its combination with the ARIMA model improves prediction performance than traditional [21]. The hybrid EMD-ARIMA model shows effective long-term and short-term prediction [22], [23].

After denoising the original series with EMD and wavelet analyzer for the applied ARIMA model, we must get a suitable parameters estimation technique to build a well-fitted model, which is the main problem in model selection. A good model selecting process will balance the goodness of

model fit with simplicity. In most studies, AIC and BIC are used to estimate the parameters of the ARIMA (p, d, q) model. Another effective method of finding a well fit model is based on the residual structure of the model [24]. Accurate parameter estimation makes its residual sequence stationary, which means a well fit model makes a stationary residual series. From this perspective, [25] used the stationarity of estimated residual series to calibrate inaccurate estimation in the case of data with complex noise. A nonstationary residual series shows a trend in the sequence that the model does not capture, which lead to under-fitting. And the stationarity of residual series can be measured by the nonstationary measure (NS) technique proposed by Ding et al. [26], [27], [28]. NS is constructed based on Shannon entropy and ergodic theory, the level of nonstationarity of a data series and the value of NS is in the interval [0, 1]. But the more the NS value is smaller, the more stationary the series is. From this perspective, the Nonstationary measure is used in this paper to check the NS value of residual in both the denoising and model selection process. This method has been used in different studies, such as [29] using NS measure to select forecasting models for irrigation water consumption and [30] using decomposition of noise and trend of a series based EMD and nonstationary measure.

This paper aims to find a model that can predict all data features. And the main problems are the limitation of the ARIMA model to nonstationary data and model selection techniques. This paper examines the daily mean temperature's statistical properties. We constructed a predictive model to forecast the daily mean temperature for long term prediction, using a comparative study of the seasonal ARIMA model based on decomposition methods. Wavelet transforms (WT) and Empirical mode decomposition (EMD) remove noise from the temperature series and Seasonal autoregressive integrated moving-average (SARIMA) model applied to the denoised series. And nonstationary measures are used in model construction in all processes. Model performance is tested in three years and one year and a half ahead forecasting.

2.   Study area and data collection

Temperature time-series data is used in this study. Our study aims to build a well fit model representing the series and forecasting the future temperature trend. The study area is in Eritrea, and one city is chosen, Asmara. We chose Asmara as the study location to represent Eritrea's high land climatic condition as it is the central area of the region. Daily average temperature data of 30 years collected from January 1, 1991, to December 31, 2020, from the website-

https://power.larc.nasa.gov/data-access-viewer/. Collected time series data has strong seasonality, irregular patterns and trends. We do not have any missing data. Eritrea found in East Africa or the Horn of Africa, between 12° and 18° north, and 36° and 44° east. It borders the Red Sea, Ethiopia, Djibouti, and Sudan. Asmara is the capital city of Eritrea, home to almost 1 million people, and it is the largest and most populated city. Asmara has a high temperature in the spring from April to June and again in September reaches an averagely of 27.8°c. It went to the lowest temperature from November to February, on averagely, to 21.6°c. The Horn Africa region has an unpredictable change in climate, which causes frequent droughts. So as Eritrea is part of this region, the climate signals and future prediction are essential. For a daily time series Y_t with a sample size of N, the statistical description such as mean, standard deviation, minimum and maximum observation and trend analysis with Mann–Kendall (M–K) test are used in this study, Table 1.

3. Method

3.1. Autoregressive Integrated Moving-average

  The main point of time series analysis using the ARIMA model is to cautiously analyze and rigorously process the past observation to develop a suitable model that describes the series's inherent structure. It brings to light the possibility of explaining the data to facilitate prediction, monitoring, and control of the data. Autoregressive Integrated Moving Average (ARIMA) models are one approach for this process. ARIMA is a prevalent model for time series prediction; it has been widely used since Box and Jenkins initially proposed it. ARIMA model is a combination of three different models, the auto-regressive (AR), integrated (I), and moving average (MA). The general equation for ARIMA (p, d, q) model is:

$$\phi_p(L)(1-L)^d Y_t = c + \theta_q(L)\varepsilon_t \qquad (3.1)$$

where $\phi_p(L)$ represent the AR and $\theta_q(L)$ the MA part of the model with ε_t error[2], [9]. The AR model is predicted Y_t as the linear combination of past values of the variable. BA uses the linear combination of past forecasting error of a variable. ARIMA model combines these models with differencing in nonstationary time series. For a series of data with seasonal behavior seasonal ARIMA, SARIMA(p, d, q)(P, D, Q)$_s$ models are appropriate to use to handle seasonality. It extends the ARIMA model, explicitly supporting data analysis with a seasonal component.

$$\Phi_P(L^s)\phi_p(L)(1-L^s)^D(1-L)^d Y_t = c + \Theta_Q(L^s)\theta_q(L)\,\varepsilon_t \qquad (3.2)$$

3.2. Wavelet analysis

Wavelet analysis is a mathematical model that reveals nonstationary time series data information in the time and frequency domain. That is why it is suitable for the nonstationary and seasonal temperature series. The main advantage of wavelet transform is that it decomposes a signal into different frequencies at a different resolution and is used as an alternative to Fourier transform. The basic wavelet function called "mother wavelet" is defined as:

$$\psi_{a,b}(y) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad for \ \ a, b \ \epsilon R; a \neq 0 \tag{3.3}$$

the parameter a is called the scaling parameter and measure the degree of compression, while b the translation parameter moves the function in the time location to analyze the signal. The transformation process is implemented using a multi-resolution decomposition technique. It divides the original series Y_t into different domain components, the detail (high frequency) series $d_{j,t}$ and the smooth (approximate) series $A_{L,t}$ using a high pass filter and low pass filter, respectively. There exist various choices of wavelet basis functions such as Haar, Daubechies, Symlet, Meyer, biorthogonal wavelet, etc. A detailed explanation of wavelet transform functions is in [11], [13]. In our study, the wavelet family was selected based on the NS measure of the residual between denoised series and original series.

When conducting wavelet analysis, selecting the optimal number of decomposition levels is one of the keys to determining the performance of a model in the wavelet domain. Based on [9], the number of decomposition levels is selected using the following formula:

$$L = int[log(N)] \tag{3.4}$$

where L is the number of decomposition levels, N is time-series length. The original time-series data Y_t will be decomposed into its detailed $d_{j,t}$ and approximate (smooth) series $A_{L,t}$. And with the wavelet de-nosing method, the series was reconstructed by removing the high-frequency component.

$$Y_t = A_{L,t} + \sum_{j=1}^{L} d_{j,t} \ \ for \ \ t=1,2,3....N. \tag{3.5}$$

3.3. Empirical mode decomposition

EMD decomposes time series into components in which all are in a time domain and has the same length as the original series, where varying frequency in time is preserved, which is generally hidden in Fourier transform. The overall result of the decomposition is to remove the highest frequencies or noise from a series successively. EMD can extract valuable information from the original data series and has received

more attention in signal denoising. But it also has a disadvantage of mode mixed in the resulting IMFs, which means similar time scales are broken down into different components.

Generally, EMD decomposes a noisy signal into different IMFs plus residuals. The decomposed signal is written as:

$$Y_t = \sum_{i=1}^{l} h_{i,t} + r_t \qquad (3.6)$$

where $h_{i,t}$ indicates the ith IMF and r_t represents the residual of the signal Y_t. In the EMD, each IMF function must satisfy the following conditions:

1. The number of maximum and minimum extrema and the number of zero crossings must either be equal or differ at most by one.

2. At any point in the IMF series, the mean of the envelope defined by the local maxima and the local minima is zero.

From the second condition, we can understand that IMF is stationary, which simplifies the analysis, but IMF can have variable amplitude and frequency. An iterative procedure of EMD called the "sifting process" is adopted to extract the separate IMF components. The detailed guidelines of the process can refer it in [31]. The residual signal contains information about the lower frequency components. The sifting process continues until the final residue is constant or a monotonic function, a function with only one maximum and minima, in which IMF cannot be obtained. The denoising series is constructed by choosing a low-frequency component that has more relation with the actual series. Selecting components was based on the NS value of the error between the Denoised and original data. After adding the low-frequency component to reconstruct the new series, find the error by subtracting it from the training set.

Step1: let Y_t be the original time series, decompose the series with EMD decomposition procedure

to $\quad Y_t \rightarrow imf_1, imf_2, \dots, imf_n, residual$

Step2: construct the Denoised series Y_t^d by adding low-frequency components

$$Y_t^d = residual + \sum_{n}^{n-i} imf_n, \text{ for } i=1,2,\dots,n-1. \qquad (3.7)$$

Step3: calculate the difference series between original and denoised data

$$\varepsilon_t = Y_t - Y_t^d, \text{ t=1, 2,..., n} \qquad (3.8)$$

Step4: If the Ns value of the difference series is $NS(\varepsilon_t) \leq 0.05$ then Y_t^d is the new Denoised series. If the value is bigger, go to step 2. and $i=i+1$. Continue the process until we get a stationary residual series.

3.4. Hybrid decomposition models with ARIMA model

The noise component affects the forecasting results' accuracy significantly because ARIMA model cannot handle nonstationary data. To solve this problem, wavelet and EMD were chosen as denoising methods and used the seasonal ARIMA model to control the series's seasonal behavior. The first step is to de-noise the series with EMD and wavelet analysis. In each method, the NS value of residual between the denoised series and the actual series is checked. The removed component must be stationary to consider noise or a series without trend patterns. The advantage of using NS measurement is to check if there is any trend in the removed component, which makes the NS value bigger. The threshold we select is 0.05, and a higher value is considered a nonstationary residual.

Figure 1: Model construction flow chart
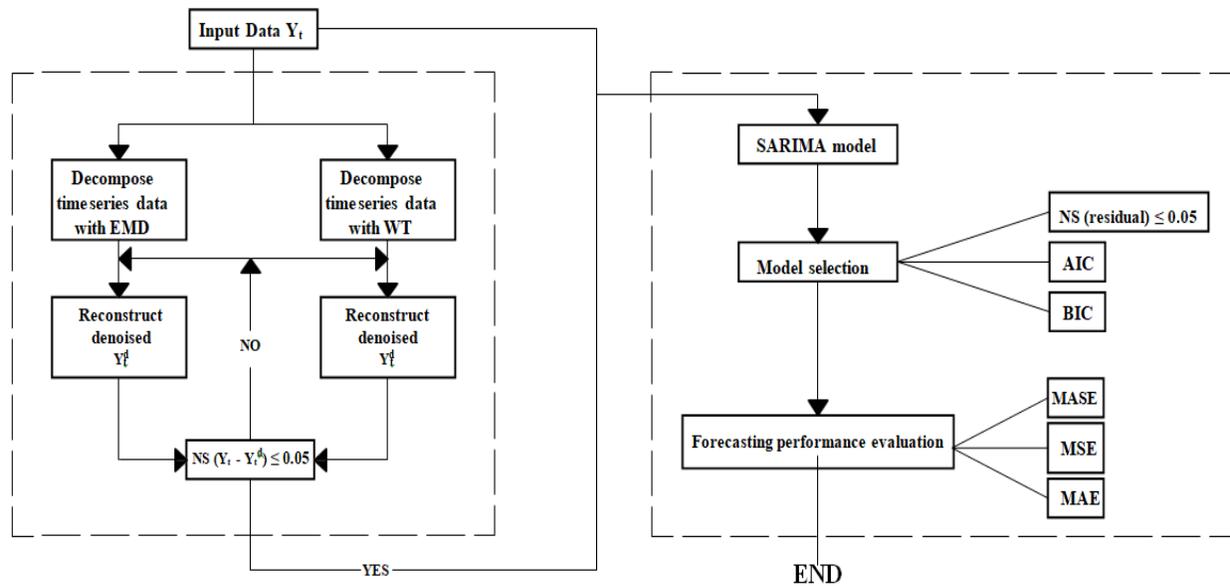
Note: Y_t and Y_t^d donate the actual and denoised time series respectively with time $t = 1, 2, ..., N$

3.5.   Model selection and forecasting performance evaluation

Model selection is an essential task of selecting a well-fitted model from a candidate set of models. A good model selection technique finds a model that shows the data's features for data points collected under random noise and predicts future points without any available future information. AIC and BIC are model selection criteria's, which are the most used techniques in ARIMA model selection. Test how well the model fits the time series. Another more effective method can be used based on the residual characteristics. By checking if the estimated residual is

close to noise? Let Y_t been a time series data collected somewhere in the past, and the relation we need between the collected series of data and the actual model f which represents it is written as:

$$Y_t = f(t) + \varepsilon_t \qquad (3.9)$$

where ε_t is identically and independently distributed noise. If $\hat{f}$ is the estimated function of f, then the estimated residual $\hat{\varepsilon}_t$ of the model at each time $t=\{t=1, 2, ..., N\}$ is:

$$\hat{\varepsilon}_t = (y_t - f(t)) + (f(t) - \hat{f}(t)) \qquad (3.10)$$

where y_t denote the observation at time t. Let's rewrite the equation as:

$$\hat{\varepsilon}_t = y_t - \hat{f}(t) \qquad (3.11)$$

The estimated residual will be close to random noise if the estimated model correctly fits the actual time series. But if any trend appears in the residual series, it shows that the model fitted poorly [24][25]. From this perspective, model selection based on the residual structure is more effective. For checking residual series behaving random, a nonstationary measure can be taken to the residual series. An accurate estimation had a stationary residual sequence. The NS measure proposed by Ding et al. is constructed based on ergodic theory and Shannon entropy and is sensitive to nonstationary signals. The value of NS is from [0,1]. The bigger NS value represents the higher level of nonstationarity. In this study, 0.05 is taken as the threshold to check if the residual series ε_t is stationary. But measuring the residual stationarity is not enough. Because even the residual is stationary, the model may over-fit the data and read the noisy component with actual patterns of the series. We use AIC, BIC, and NS comparatively in model selection and choose models that give better performance to solve such problems.

In the forecasting process, for each selected model the forecasting performance was assessed using the mean absolute error (MAE), the root means square error (RMSE), and the mean absolute scaled error.

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t| \qquad (3.12)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2} \qquad (3.13)$$

where Y_t testing set and $\hat{Y}_t$ forecasted data series with n length series. MAE is the mean absolute difference between the testing set and forecasted data. RMSE, the square root of the average square difference of test set and forecasted value. In both, the minimum value is a better result. The formula gives for MASE:

$$MASE = \frac{MAE}{Q} \qquad (3.14)$$

where Q is the scaling statistic computed on the training set, for non-seasonal time series, the Q value is the mean absolute error of the one-step naïve forecasting method.

$$Q = \frac{1}{n-1}\sum_{j=1}^{n}|Y_j - Y_{j-1}| \qquad (3.15)$$

and for seasonal series:

$$Q = \frac{1}{n-m}\sum_{j=m+1}^{n}|Y_j - Y_{j-m}| \qquad (3.16)$$

where m is the season period and n the size of the training set, if the MASE result is less than one, it means a better forecasting value than the naïve forecast, and if the result is greater than one, it shows worse forecasting. A model with a minimum value in all tests is considered better performance.

4. Result and Discussion

4.1.   Data Division

  The average temperature time series data used in this study were daily collected data of 30 years in Asmara, Eritrea. A nonparametric trend analysis is used in our research. The seasonal Mann-Kendall (M-K) test shows an increasing trend in the series as the *p-value* is <0.0001. Table 1. demonstrates the statistical description of the data and the type of trend in the series. Sen's slope refers to the slope of the trend, which is 0.013. As we see the structure of our data next step was to find a model that captures the trend and features of the series. And categorizing the series into groups is essential because different sets are needed for training data, while another set is used as testing data. Training data is manipulated to train and develop the forecasting model.

In contrast, testing data is treated as future data that must be used to compare the fitted model's predictability accuracy. It gives the ability to compare the effectiveness of different models of prediction. The partitioning way of the series is essential, which could affect the performance of the forecasting result of the models, and several aspects such as the data type and the size of the available data should be taken into consideration. In this research, the training set was chosen from January 1, 1991- December 31, 2017, 90% of the data, and the test set from January 1, 2018 - December 31, 2020, 10%. From the Fig.2 of the actual data graph, it is noticeable that the temperature series had seasonality and instabilities.

Table 1: Descriptive statistics and trend analysis of the daily temperature data

Observations	missing data	Minimum	Maximum	Mean	Std. deviation
10957	0	13.795 °c	29.320 °c	22.707 °c	2.285
		Mann-Kendall trend test			
Kendall's tau	Sen's slope	p-value	alpha, α		
0.140	0.031	<0.0001	0.05	There is a trend in the series	

Figure 2: Original series graph
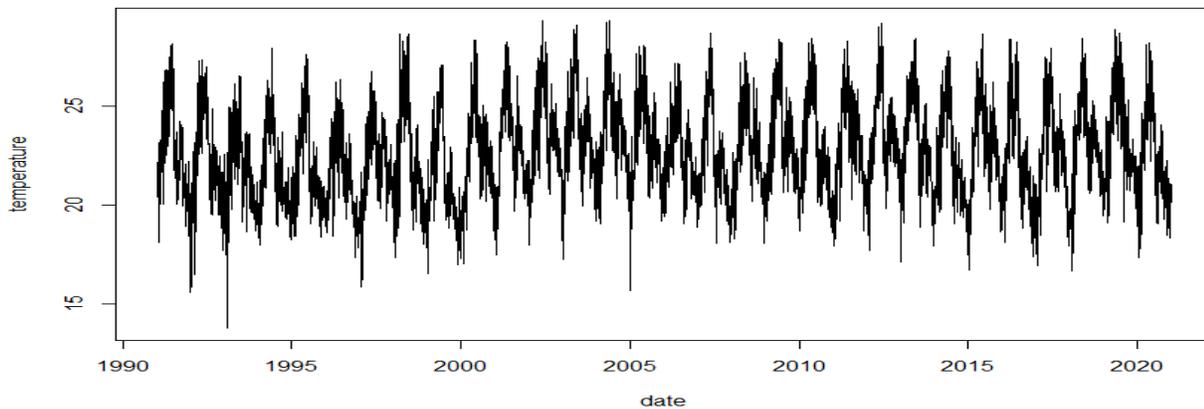
4.2. Data Decomposition

Two decomposition methods are used in denoising the temperature time series, and the first step is to decompose the noisy series into different components. In the wavelet transform procedure, the decomposition level is chosen based on the log value of the time series size de. Based on Eqn. (3.4). the optimal number of decomposition for this series is four levels. Many researchers have used this method to choose the decomposition level and stated that the data series could be expressed helpfully and more meticulously. Good forecasting results can be achieved for nonstationary time series [10]. In this study, the Haar family, with decomposition level 4, was chosen compared to other families of wavelet transformation because it presented relevant smooth

and depending on the residual series between the actual series and denoised series data. Residual series tested for NS measurement and Haar family Denoising series were suitable methods. Applying to the time series Y_t results in 5 series, a_4 the approximate series and d_4, d_3, d_2 and d_1 are the detail, with residual NS value of the training set Y_t and denoised series Y_t^d:

$$\varepsilon_t = Y_t - Y_t^d; \; NS(\varepsilon_t) = 0.02$$

Figure 3: Wavelet decomposed component
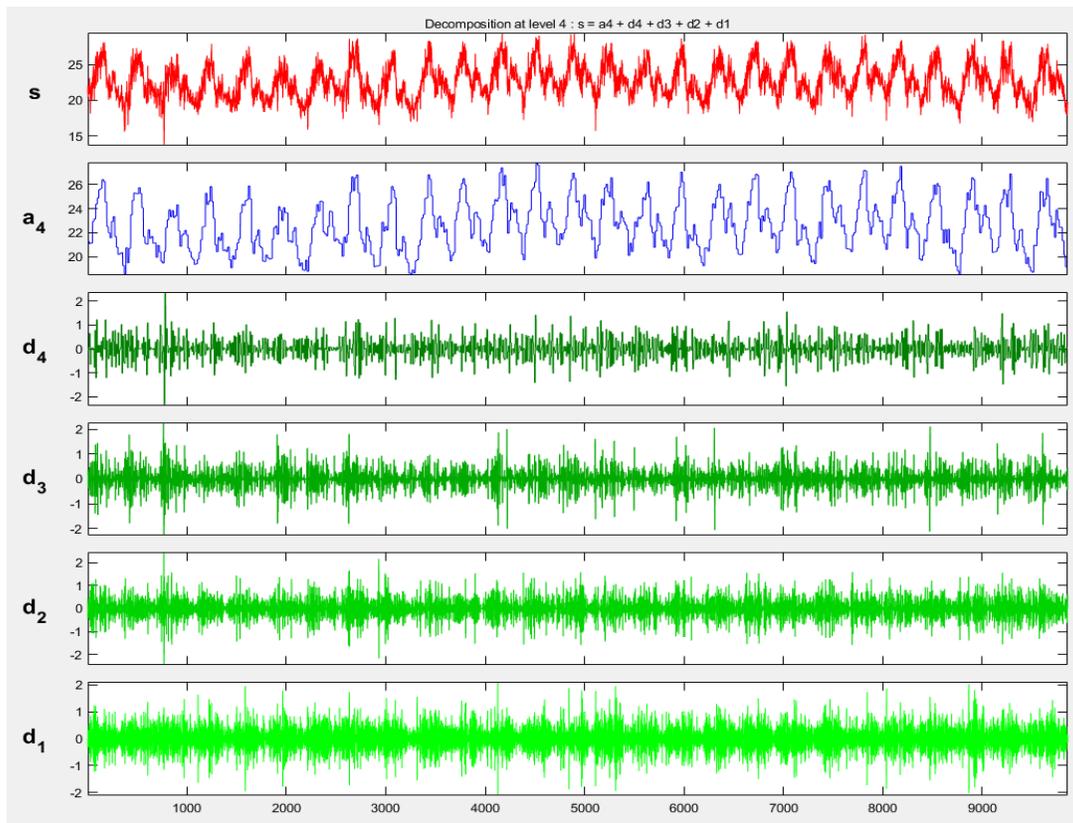
Note: d$_1$, d$_2$, d$_3$ and d$_4$ are the detail component and a$_4$ approximate component

The second method used for denoising is EMD. The training set Y_t decomposed into residual and 7 IMF components, as shown in Fig.4. The original series is decomposed into relatively stationary series components. The high-frequency component is extracted in the first few IMFs, such as imf1, imf2 showing the high variation in the series. And the low-frequency components are extracted in the last IMFs reveals the seasonal behavior and the trend of the series. A denoised series was built by adding residual components and the low-frequency IMFs, starting from residual and IMF7 to higher IMF step-by-step. Each time we add one IMF, considering it as the new series, check the

NS value of the error ε_t between the new series and training set data and the NS value less or equal to 0.05 consider being stationary. Continue the process until we get a stationary series ε_t and a smaller NS value. In this study, the reconstructed series was:

$$Y_t^d = \{residual + imf_7 + imf_6 + imf_5 + imf_4\},$$

with smallest NS value $\varepsilon_t = Y_t - Y_t^d$; $NS(\varepsilon_t) = 0.035$.

Figure 4: EMD decomposed component
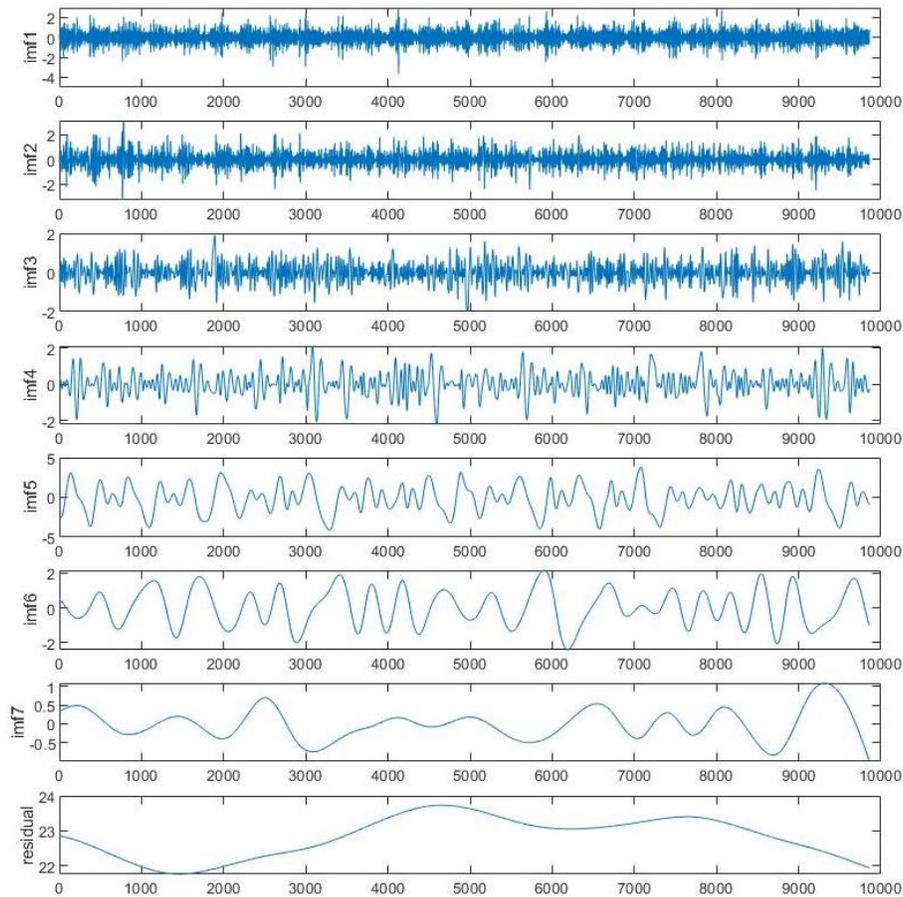
The denoised series from both WT and EMD with the original series are given in Fig 5. Denoised series must have same trend and seasonality as the original series. The wavelet denoised series except the noise component it had similar trend and seasonal patterns as the original series. But in the case of EMD denoised series there is showing some different patterns which is out of the original series structure.

Figure 5: Denoised series from each decomposition method Vs training set data
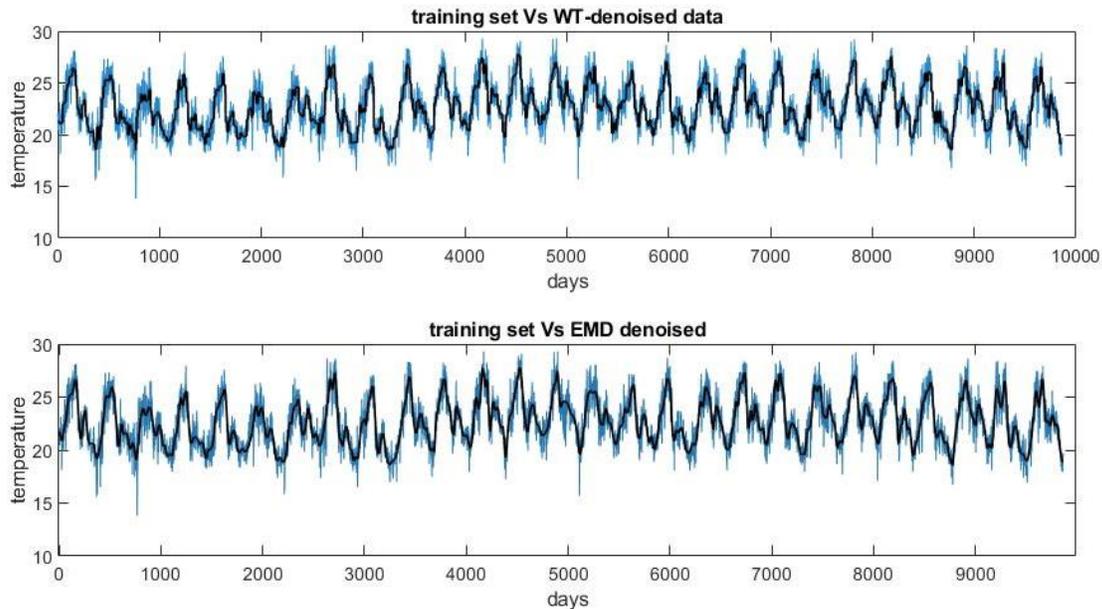
4.3. ARIMA model construction

The time-series data is stationary if the ACF graph cuts off rapidly, but if the ACF graph dies down too slow, the series is considered nonstationary. In Fig.6, it is noticeable that the ACF graph drops gradually; thus, the training set of the actual data is considered nonstationary and nonlinear. In this study, decomposition methods were added to remove the noise component of the series, and as the decomposed components illustrate, the original series has seasonal behavior. Therefore, the ARIMA model will use seasonal differencing to handle the seasonal behavior of the series. SARIMA applied to the reconstructed denoised time series. We have two different methods; the EMD based seasonal ARIMA (EMD-SARIMA) and wavelet-based seasonal ARIMA (WT-SARIMA) are the hybrids models. And seasonal ARIMA model was applied directly to the original series to compare the improvement of the hybrid methods.

Figure 6: ACF and PACF graph of temperature series data (training set)
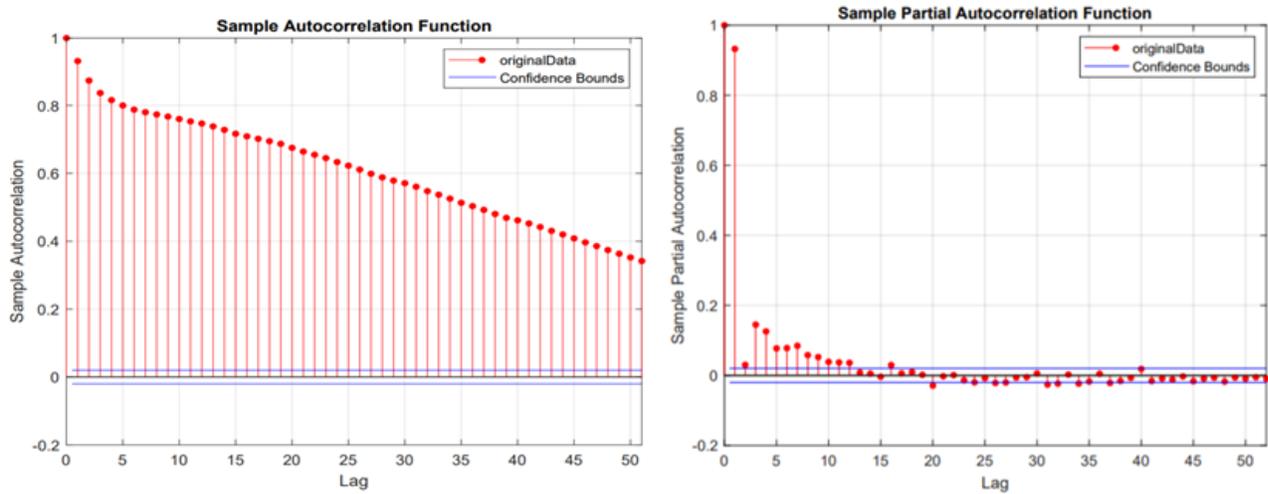
As the proposed procedure, the original and denoised data will be treated as input to develop their respective models. As mentioned in section II, all procedures will be applied to the series with the assistance of MATLAB software. The model parameters are chosen based on the AIC, BIC, and NS values residual series between the training set and model predicted data. In AIC and BIC criteria, a model with the minimum result shows it fits the data well. On the other hand, measuring the NS value to the residual, the difference between model prediction and actual observation (training set), will tell us if there is ignored or not captured information of the original data. It will handle the under-fitting problem. So a well fit model has a smaller value in those three methods. But in some models, The result of three model selection criteria may not reduce proportionally. In this situation model is selected by balancing these three criteria to find the minimum value of AIC and BIC and stationary residual series where the residual NS value must be less or equal to 0.05.

Results of AIC, BIC and NS of sample models from WT and EMD based SARIMA are listed in Table 2. The table shows that AIC, BIC result and NS value for WT-SARIMA hybrid models do not have proportional change. In the model with smaller AIC, BIC has a bigger residual NS value and vice versa. Comparatively, SARIMA(3,0,2)(1,1,1)$_{365}$ has a better result. SARIMA(2,0,1)(1,1,1)$_{365}$ and SARIMA(3,0,2)(1,1,1)$_{365}$ selected to test the forecasting performance examine the effect of residual structure in model performance.

SARIMA(2,0,1)(1,1,1)$_{365}$ has the minimum AIC and BIC value, but as shown in Table 2., the NS value of its residual is greater than 0.05, which implies its residual series is not stationary. While SARIMA(3,0,2)(1,1,1)$_{365}$ was chosen based on the minimum value of the three test. Similarly SARIMA(3,1,2)(1,1,1) and SARIMA(3,1,1)(0,1,1) selected from direct applied models on training set without decomposition methods. But still, for all series, the NS value is bigger than 0.05. Applying the SARIMA model directly to the nonstationary series had a weak performance. A bigger NS value, as we say it early, there is a trend in the residual series because the model has not captured every structure of the series. The other hybrid models were EMD based seasonal ARIMA (EMD-SARIMA). Unlike WT-SARIMA in the EMD denoised series, the result of three model selection methods was proportional, so we chose one model with a minimum value in all the tests, SARIMA(7,0,0)(1,1,0)$_{365}$. Table 3. illustrate all the result of the selected models, and all hybrid models have stationary residual series, which show how well they fit the training set.

Table 2: Sample models for AIC, BIC and NS value of WT-SARIMA models and EMD-SARIMA

model	AIC	BIC	NS
WT-SARIMA			
SARIMA(2,0,1)(1,1,1)$_{365}$	3.2844e+03	3.327e+03	0.066
SARIMA(3,0,2)(0,1,1)$_{365}$	3.3928e+03	3.4432e+03	0.057
SARIMA(3,0,2)(1,1,1)$_{365}$	3.3589e+03	3.4165e+03	0.039
SARIMA(2,0,1)(0,1,1)$_{365}$	3.4762e+03	3.5122e+03	0.024
SARIMA(2,0,1)(1,1,0)$_{365}$	5.7783e+03	5.8143e+03	0.015
EMD-SARIMA			
SARIMA(7,0,0)(1,1,0)$_{365}$	-1.4393e+05	-1.4387e+05	0.038
SARIMA(8,0,0)(1,1,0)$_{365}$	-1.4394e+05	-1.4386e+05	0.038
SARIMA(7,0,0)(0,1,0)$_{365}$	-1.4108 e+05	-1.4102e+05	0.038

Table 3: Selected models from WT-SARIMA, EMD-SARIMA and SARIMA applied to original data

Model		NS	AIC	BIC
Actual data	SARIMA(3,1,2)(1,1,1)	0.72	2.4365e+04	2.4422e+04
	SARIMA(3,1,1)(0,1,1)	0.54	2.4533e+04	2.4570e+04
Wavelet	SARIMA(2,0,1)(1,1,1)$_{365}$	0.066	3.2844e+03	3.327e+03
	SARIMA(3,0,2)(1,1,1)$_{365}$	0.039	3.3589e+03	3.4165e+03
EMD	SARIMA(7,0,0)(1,1,0)$_{365}$	0.038	-1.4393e+05	-1.4387e+05

4.4. Effectiveness Evaluation result

The MAE, MASE, and RMSE results of testing data for all hybrid forecasting models and SARIMA are exhibited in Fig 7, Table 4 and 5. depict the graph of the forecasted result of chosen three models. Smaller error testing result indicates a better forecasting accuracy approach. And Tables 4 and 5. showing the WT-SARIMA(3,0,2)(1,1,1)$_{365}$ model has the best performance in both 1095 and 546 days prediction. It gives a smooth series and captures the original series seasonal behavior. However, EMD- SARIMA model has weak performance. Referring to both tables, it is clear that the MAE, MASE and RMSE result of EMD based SARIMA model is bigger than the results of all other methods. And Fig 7. Depicts the graph also shows that it weakly fit the original time series data.

On the other hand, in WT-SARIMA and directly applied models, we chose two models based on minimum AIC and BIC values with bigger NS measures of residual series and models with smaller NS values of the residual. The forecasting performance improved in both SARIMA and WT-SARIMA models selected based on residual test, which has smaller NS results. In the modelling process, sometimes the model over-fit the series by coping with the noisy part or under-fits the series when the model ignores some information or features of the actual series. That gives a wrong picture of the original time series structure. For this reason, we have to choose appropriate model selection methods. AIC and BIC are most used methods in model selection, but as we can see the results models chosen base on minimum value of both criteria are having nonstationary residual series. In other word, there is some information which is not captured by the model, using NS measure in the residual help to control this problem. And applying these three criteria improves the accuracy of the prediction. Following all performance evaluation outcomes, we can deduce that the proposed hybrid model, wavelet transformation and SARIMA models can give a superior and improved prediction outcome than the straightforward application of the ARIMA model and others.

Table 4: forecasting performance of models for 1095 days

Model		MASE	rank	RMSE	rank	MAE	rank
Actual data	SARIMA(3,1,2)(1,1,1)	0.844633	4	1.51558	4	1.201135	4
	SARIMA(3,1,1)(0,1,1)	0.822906	3	1.47866	3	1.170238	3
wavelet	SARIMA(2,0,1)(1,1,1)$_{365}$	0.761913	2	1.359894	2	1.0835	2
	SARIMA(3,0,2)(1,1,1)$_{365}$	0.735675	1	1.302188	1	1.046187	1
EMD	SARIMA(7,0,0)(1,1,0)$_{365}$	0.890858	5	1.569255	5	1.266871	5

Table 5: 546 days forecasting test

Model		MASE	rank	RMSE	rank	MAE	rank
Actual data	SARIMA(3,1,2)(1,1,1)	0.906336	4	1.579976	4	1.288881	4
	SARIMA(3,1,1)(0,1,1)	0.872747	3	1.526071	3	1.241115	3
wavelet	SARIMA(2,0,1)(1,1,1)$_{365}$	0.791927	2	1.382071	2	1.126183	2
	SARIMA(3,0,2)(1,1,1)$_{365}$	0.743171	1	1.314283	1	1.056848	1
EMD	SARIMA(7,0,0)(1,1,0)$_{365}$	0.982199	5	1.704065	5	1.396765	5

In this work, forecasting performance is tested for long-term predictions. Frist evaluation was for 10% of the original data, which is 1095days, and the performance of all models was checked with residual test Table 4. The prediction performance was evaluated for 546 days (year and a half) results are in Table 5. And the hybrid models of wavelet transformation and SARIMA had the smallest value in all tests. In general, the WT-SARIMA model has the highest performance in this study and captures all critical features of the original series. EMD- SARIMA has weak performance in this study. In our opinion, this could be because EMD has a weak ability to read large size data and mix information in different components. In the EMD denoising process, all the original series features were not captured. As we can see in Fig 7. the EMD-SARIMA predicted series graph has some different patterns from the original series.

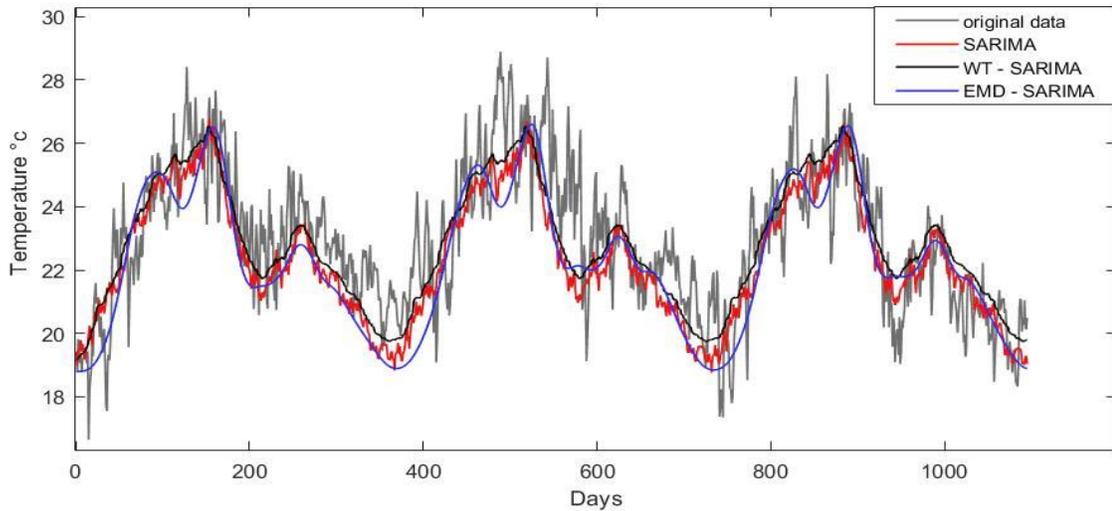Figure 7: 1095 days forecasting series Vs testing set

Table 6: Parameter estimation of WT-SARIMA $(3,0,2)(1,1,1)_{365}$

	Coef.	Std. error	T-statistic	P-value
AR{1}	-0.45845	0.020271	-22.616	2.98E-113
AR{2}	0.43318	0.011512	37.629	7.1269e-310
AR{3}	0.90625	0.018038	50.241	0
SAR{365}	-0.1488	0.006316	-23.575	6.88E-123
MA{1}	1.4175	0.023856	59.417	0
MA{2}	0.91483	0.022793	40.136	0
SMA{365}	-0.8159	0.003962	-205.93	0

From Eqn.(3. 2). the equation of WT- SARIMA$(3,0,2)(1,1,1)_{365}$ model with estimated parameters, L lag operator and ε_t the error at t is defined as:

$$(1 - \Phi_{365}L^{365})(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)(1 - L^{365})Y_t = (1 + \Theta_{365}L^{365})(1 + \theta_1 L + \theta_2 L^2)\varepsilon_t$$

The estimated coefficients value, standard error, t-statistic and p-value of the model are given in Table 6. Since all P-values are less than 0.05, the results are statistically significant.

5.   Conclusion

Eritrea is located in the area where climate change is a major issue. And the climate of this country is affected by topography, which divide the country into high land and low land. Most of the population lives in the high land region and more than 70% of the population life depends on agriculture. Therefore, studies conducting on climate change are important to understand the trend and future prediction. As temperature is one element of climate change in this study we focus on finding a model for long term prediction of temperature in the city Asmara. This study assessed the capability of prediction of SARIMA based two different decomposition methods and the characteristics of the temperature data. The result from Mann-Kendall test show increasing trend in the series with Kendall's tau 0.014 and Sen's slope 0.031. The wavelet transform and empirical mode decomposition were used to denoise the temperature series and SARIMA applied to the denoised series. And wavelet transform base SARIMA model give better prediction results. The selected WT-SARIMA model can now forecast future daily temperature values. Due to the fundamental importance of forecast accuracy, a test should be performed to verify the forecasting accuracy by comparing the forecast values with observational values in the testing set. The test methods can also support avoiding under-fitting or over-fitting. And as we see in the 1095 and 546 days forecasting result tables, it had a good performance. The temperature time series had seasonal behaviors, trends, and irregularities. Appling ARIMA models directly without preprocessing will not give accurate results. As we see in this paper and much other research, using decomposition methods with the ARIMA model provides better performance. Using the hybrid models of WT with SARIMA models had shown improvement in the accuracy of the results we had. Our study supports using hybrid models instead of directly applying the ARIMA model to nonstationary and nonlinear data. In this study, wavelet transformation with the seasonal autoregressive integrated moving average model was best to choose for our time series data. But as the result and graph of the forecasting test elucidate the hybrid model base, EMD has weak performance. Fig 8. shows that the model has some different seasonal patterns. The other important point we focused on is the model selection problem. Using of Nonstationary measurement in model selection shows good performance compare to AIC and BIC in this study. NS measure helps us check if the residual series of the model has a trend or behaves nonrandom, which indicate if there is removed information or trend from the series and treated as noise. If there is any trend in the residual, the NS value gets bigger. NS measure results can show how well

the models capture all the information and structure of the series. As we see in the model construction and forecasting evaluation, models chosen based on the NS value has higher performance.   As we are going for minimum AIC and BIC, the model's residual is showing nonstationarity. And this is one drawback of AIC and BIC we see in this study. Measuring the NS value of the residual is practical to find a well-fitted model.

Conflicts of Interest: The author(s) declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] M.G. Ghebrezgabher, X. Yang, Spatio-Temporal Assessment of Climate Change in Eritrea Based on Precipitation and Temperature Variables, World Wide J. Multidiscip. Res. Develop. 4 (2018), 1–10.

[2]]M. Murat, I. Malinowska, M. Gos, J. Krzyszczak, Forecasting Daily Meteorological Time Series Using ARIMA and Regression Models, Int. Agrophys. 32 (2018), 253–264. https://doi.org/10.1515/intag-2017-0007.

[3] D.B. Lobell, A. Sibley, J. Ivan Ortiz-Monasterio, Extreme Heat Effects on Wheat Senescence in India, Nat. Clim. Change. 2 (2012), 186–189. https://doi.org/10.1038/nclimate1356.

[4] M.A. Semenov, P.R. Shewry, Modelling Predicts That Heat Stress, Not Drought, Will Increase Vulnerability of Wheat in Europe, Sci. Rep. 1 (2011), 66. https://doi.org/10.1038/srep00066.

[5] N. Subash, A.K. Sikka, Trend Analysis of Rainfall and Temperature and Its Relationship Over India, Theor. Appl. Climatol. 117 (2014), 449–462. https://doi.org/10.1007/s00704-013-1015-9.

[6] E. El-Mallah, S. Elsharkawy, Time-Series Modeling and Short Term Prediction of Annual Temperature Trend on Coast Libya Using the Box-Jenkins ARIMA Model, Adv. Res. 6 (2016), 1–11. https://doi.org/10.9734/AIR/2016/24175.

[7] K. Anitha Kumari, N. Kumar Boiroju, P.R. Reddy, Forecasting of Monthly Mean of Maximum Surface Air Temperature in India, Int. J. Stat. Math. 9 (2014), 14–19.

[8] P. Chen, A. Niu, D. Liu, W. Jiang, B. Ma, Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing, IOP Conf. Ser.: Mater. Sci. Eng. 394 (2018), 052024. https://doi.org/10.1088/1757-899X/394/5/052024.

[9] A.H. Nury, K. Hasan, Md.J.B. Alam, Comparative Study of Wavelet-ARIMA and Wavelet-ANN Models for Temperature Time Series Data in Northeastern Bangladesh, J. King Saud Univ. - Sci. 29 (2017), 47–61. https://doi.org/10.1016/j.jksus.2015.12.002.

[10] N.Q.N. Md-Khair, R. Samsudin, A. Shabri, Wavelet Transform and Autoregressive Integrated Moving Average Combination Approach in Crude Oil Prices Forecasting, Int. J. Innov. Comput. 8 (2018), 41-48. https://doi.org/10.11113/ijic.v8n2.177.

[11] A. Ben Mabrouk, N. Ben Abdallah, Z. Dhifaoui, Wavelet Decomposition and Autoregressive Model for Time Series Prediction, Appl. Math. Comput. 199 (2008), 334–340. https://doi.org/10.1016/j.amc.2007.09.067.

[12] T. Kriechbaumer, A. Angus, D. Parsons, M. Rivas Casado, An Improved Wavelet–ARIMA Approach for Forecasting Metal Prices, Resources Policy. 39 (2014), 32–41. https://doi.org/10.1016/j.resourpol.2013.10.005.

[13] S.A.A. Karim, M.T. Ismail, M.K. Hasan, J. Sulaiman, Denoising the Temperature Data Using Wavelet Transform, Appl. Math. Sci. 7 (2013), 5821–5830. https://doi.org/10.12988/ams.2013.38450.

[14] A. Araghi, M. Mousavi Baygi, J. Adamowski, J. Malard, D. Nalley, S.M. Hasheminia, Using Wavelet Transforms to Estimate Surface Temperature Trends and Dominant Periodicities in Iran Based on Gridded Reanalysis Data, Atmospheric Res. 155 (2015), 52–72. https://doi.org/10.1016/j.atmosres.2014.11.016.

[15] K.-M. Lau, H. Weng, Climate Signal Detection Using Wavelet Transform: How to Make a Time Series Sing, Bull. Amer. Meteor. Soc. 76 (1995), 2391–2402. https://doi.org/10.1175/1520-0477(1995)076<2391:CSDUWT>2.0.CO;2.

[16] M. Sifuzzaman, M. Islam, M. Ali, Application of Wavelet Transform and its Advantages Compared to Fourier Transform, J. Phys. Sci. 13 (2009), 121–134.

[17] N. Al Bassam, V. Ramachandran, S. Eratt Parameswaran, Wavelet Theory and Application in Communication and Signal Processing, in: S. Mohammady (Ed.), Wavelet Theory, IntechOpen, 2021. https://doi.org/10.5772/intechopen.95047.

[18] G. Bousaleh, F. Hassoun, T. Ibrahim, Application of Wavelet Transform in the field of Electromagnetic Compatibility and power quality of industrial systems, in: 2009 International Conference on Advances in Computational Tools for Engineering Applications, IEEE, Beirut, Lebanon, 2009: pp. 284–289. https://doi.org/10.1109/ACTEA.2009.5227896.

[19] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis, Proc. R. Soc. Lond. A. 454 (1998), 903–995. https://doi.org/10.1098/rspa.1998.0193.

[20] M. Kedadouche, M. Thomas, A. Tahan, A Comparative Study Between Empirical Wavelet Transforms and Empirical Mode Decomposition Methods: Application to bearing defect diagnosis, Mech. Syst. Signal

Process. 81 (2016), 88–107. https://doi.org/10.1016/j.ymssp.2016.02.049.

[21] H. Wang, L. Liu, Z. (Sean) Qian, H. Wei, S. Dong, Empirical Mode Decomposition–Autoregressive Integrated Moving Average: Hybrid Short-Term Traffic Speed Prediction Model, Transport. Res. Record. 2460 (2014), 66–76. https://doi.org/10.3141/2460-08.

[22] Z.Y. Wang, J. Qiu, F.F. Li, Hybrid Models Combining EMD/EEMD and ARIMA for Long-Term Streamflow Forecasting, Water. 10 (2018), 853. https://doi.org/10.3390/w10070853.

[23] Y. Zhou, M. Huang, Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model, Microelectron. Reliab. 65 (2016), 265–273. https://doi.org/10.1016/j.microrel.2016.07.151.

[24] Q. Tan, H. Jiang, Y. Ding, Model Selection Method Based on Maximal Information Coefficient of Residuals, Acta Math. Sci. 34 (2014), 579–592. https://doi.org/10.1016/S0252-9602(14)60031-X.

[25] Z. Zhou, Z. Zhou, L. Wu, Calibration for Parameter Estimation of Signals with Complex Noise via Nonstationarity Measure, Complexity. 2021 (2021), 8840757. https://doi.org/10.1155/2021/8840757.

[26] Y. Ding, W. Fan, Q. Tan, K. Wu, Y. Zou, Nonstationarity Measure of Data Stream (in Chinese), Acta Math. Sci. 30A (2010), 1364–1376.

[27] K. Wu, Nonstationarity of Stock Returns, in: M. Bohner, Y. Ding, O. Došlý (Eds.), Difference Equations, Discrete Dynamical Systems and Applications, Springer International Publishing, Cham, 2015: pp. 153–165. https://doi.org/10.1007/978-3-319-24747-2_12.

[28] Q. Tan, The Nonstationarity Measure of Time Series and Its Application (in Chinese), Ph.D. thesis, University of Chinese Academy of Sciences, Beijing, China, 2013.

[29] J. Lan, Q. Tan, and Q. Dong, Selection of Forecasting Models For Irrigation Water Consumption Based on Nonstationarity Measure (in Chinese), Eng. J. Wuhan Univ. 47 (2014), 721–725.

[30] Q. Tan, L. Wu, B. Li, Decomposition of Noise and Trend Based on EMD and Nonstationarity Measure (in Chinese), Acta Math. Sci. 36A (2016), 783–794.

[31] S. Maheshwari and A. Kumar, Empirical Mode Decomposition: Theory & Applications, Int. J. Electron. Electric. Eng. 7 (2014), 873–878.

[32] B. Yadollah, F. Gharib, B. Ali, A Study and Prediction of Annual Temperature in Shiraz Using ARIMA Model, Geograph. Space, 12 (2012), 127-144.

[33] N. Al Bassam, V. Ramachandran, S. Eratt Parameswaran, Wavelet Theory and Application in Communication and Signal Processing, in: S. Mohammady (Ed.), Wavelet Theory, IntechOpen, 2021. https://doi.org/10.5772/intechopen.95047.

[34] Y. Wei, J. Wang, C. Wang, Network Traffic Prediction Based on Wavelet Transform and Season ARIMA Model, in: D. Liu, H. Zhang, M. Polycarpou, C. Alippi, H. He (Eds.), Advances in Neural Networks – ISNN 2011, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 152–159. https://doi.org/10.1007/978-3-642-21111-9_17.