مجلة العلوم البحثة والتطبيقية
Journal of Pure & Applied Sciences
www.Suj.sebhau.edu.ly ISSN 2521-9200
Received 17/05/2019 Revised 23/08/2019 Published online 11/12/2019

Comparison of hybrid models in temperature prediction in the Libyan city of Sirt

*Muhammad Baay Mohammed, Abdslam K. Suliman and Nuri Omer. Ali
Department of Statistics, faculty of Science, University of Sebha, Libya
*Corresponding author: moh.mohammed1@sebhau.edu.ly

Abstract Temperature forecasting is one of the most important factors in climate studies such as agriculture, vegetation, water resources and tourism. Given the importance of temperature forecasting we believe that there is a real need to develop appropriate solutions for reading the future in terms of accuracy, quality and reliability. This paper compares several advance methods of prediction; including (EMD/EEMD/SEMD) with ARIMA model using temperature recorded data in the Libyan city of Sirte from 1945 to 2010. Predicting the next 10 years tells us that the best of these methods is (SEMD with ARIMA model).

Keywords: Accuracy, Forecasting, Temperature, EMD, EEMD, SEMD.

مقارنة النماذج الهجينة في التنبؤ بدرجات الحرارة في مدينة سرت الليبية

*محمد باي علي حسين محمد و عبدالسلام كامل سليمان و نوري عمر علي
قسم الإحصاء–كلية العلوم–جامعة سبها، ليبيا
*للمراسلة: moh.mohammed1@sebhau.edu.ly

الملخص يعد التنبؤ بدرجات الحرارة أحد أهم العوامل في دراسات المناخ مثل الزراعة والغطاء النباتي وموارد المياه والسياحة. بالنظر إلى أهمية التنبؤ بدرجات الحرارة، نعتقد أن هناك حاجة حقيقية لتطوير حلول مناسبة لقراءة المستقبل بأكثر دقة وجودة وموثوقية. هذه الورقة تقارن عدة طرق متقدمة للتنبؤ، وتشمل (EMD / EEMD / SEMD) مع نموذج ARIMA باستخدام بيانات درجات الحرارة المسجلة في مدينة سرت الليبية للفترة من 1945 إلى 2010. التنبؤ للسنوات العشر المقبلة يخبرنا أن أفضل هذه الطرق هي (SEMD مع نموذج ARIMA).

الكلمات المفتاحية: الدقة، التنبؤ، درجات الحرارةEMD.،EMD، SEMD.

1. Introduction

Prediction is defined as estimating and making assumptions about future events using special techniques over different time periods. The importance of forecasting lies in most areas of life and most economic institutions and international institutions that there is a fear of future events, especially in cases of profit and loss and increasing decreasing for some natural phenomena. For this, it comes the role of prediction in putting some confidence and to bring the audacity to take future steps. The prediction ensures to a large extent the efficiency and effectiveness of the organization and is able to know the needs of the organization in the short and long term, gives a fairly clear picture of the future direction of the institution. However, it is not possible to say that there is a technique or method of prediction that it is effective only if a set of conditions, including accuracy and provide the necessary data and time to collect information and provide the material and human resources necessary for the prediction process has been cleared. There are many ways to do the prediction process. These methods are called hybrid methods because they use a hybrid method between the experimental analysis models and the ARIMA model. These methods are similar in terms of the implementation method because they dismantle the original unstable time series into a number of stable time series called IMFs, but they differ in content and calculations performed within the process and in terms of the result that can be obtained. On this basis, after the time series is broken up into several series, the ARIMA model can be applied to each IMF and then make a prediction and see more clearly what will happen in the future.

2. Methodology

In this section we will look at the explanation used in the prediction from the EMD method to the EEMD to the last methods used SEMD ,by demonstrating each method individually and linking each method to the ARIMA model, which analyses stable time series.

2.1. Non-stationary time series

As defend in [1], non-stationary time series contains a trend, in which the observations will not come down to zero except for very large values of the lag. This is because an observation on one side of the overall mean tends to be followed by a large number of further observations on the same side of the mean because of trend. A typical non-stationary time series are shown in figure 1.

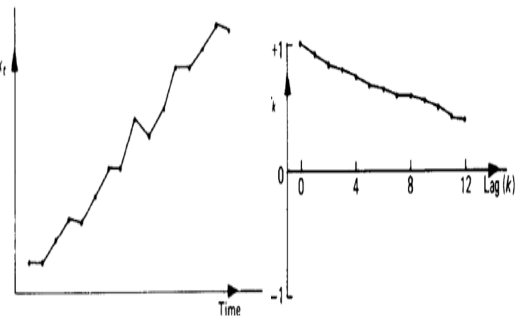


Figure (1) Non-stationary time series

2.2. Empirical Mode Decomposition[EMD]

Empirical Mode Decomposition (EMD) is an iterative procedure which decomposes signal into a set of oscillatory components, called Intrinsic Mode Functions (IMFs).[2]

Two condition need to be satisfied for a component to be considered an IMF. First: total zero crossings and the total extrema in the whole data set should be equal or vary by at most one. Second: the mean value of envelope from maxima and minima should be equal to zero at any interval of the component.[3]

EMD can decompose a signal into IMFs. EMD is a sifting process with the goal of decomposing the signal into narrow band signals. No zero crossing the middle of extrema, camwood make wiped out. The EMD calculation hence recognizes sign oscillations toward a. nearby level and separates the information under generally non-overlapping. Time scale segments. It breaks down An sign $x(t)$ under its. Part IMFs complying with two properties

1) a IMF need special case extremum between two resulting zero crossings, i. E. Those number from claiming nearby minima. What's more maxima contrasts at most Eventually Tom's perusing you quit offering on that one.

2) a IMF need An mean quality from claiming zero. Note that the second state infers that an IMF is. Stationary which simplifies its examination. At an IMF might have. Abundancy tweak Also likewise evolving recurrence.

Those EMD Algorithm: those filtering procedure could a chance to be summarized in the accompanying algorithm. Break down An information set. $X(t)$ under IMFs $x_n(t)$ What's more An residuum $r(t)$ such-and-such those sign. Might a chance to be spoke to Likewise.

The implementation of the EMD is mainly composed of the following steps[4]

1. Let $\{x(t) \in X: t = 1, 2, \dots, N\}$ denote the original average time series. All the local extremes of $x(t)$ are identified and all the maxima and minima are connected by a cubic spline line to form the upper envelope $x_{max}(t)$ and lower envelope $x_{min}(t)$. However, the spline fitting method has a serious problem at the end point where the cubic spline can have a wide swing. In order to deal with this problem, extended the original time series by adding characteristic waves at the ends which are defined by the two consecutive extrema for both their frequency and amplitude of the added waves. This method has been proved to be able to confine the large swings successfully. In this study, the time series based on the EMD package in software R.
2. The mean of the upper envelope and lower envelope $m(t)$ is calculated by Equation (1),
$$m(t) = \frac{x_{max}(t)+x_{min}(t)}{2} \quad (1)$$
3. Subtract $m(t)$ from the original time series $x(t)$ to obtain the component $h(t)$ as shown in Equation (2)
$$h(t) = x(t) - m(t) \quad (2)$$

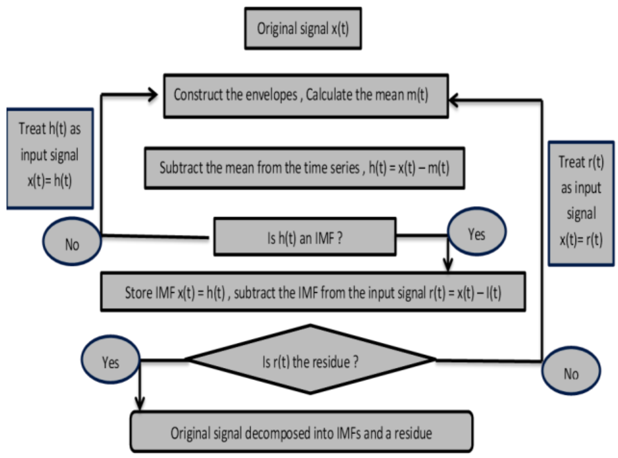
Check a time series $h(t)$ satisfies two requirements of IMF. If not, substitute $h(t)$ for $x(t)$ to repeat steps 1–3 until $h(t)$ meets the conditions of IMF, when $h(t)$ is IMF $I_1(t)$. Subtract $I_1(t)$ from the original time series $x(t)$ to get the residual $r_1(t)$, as in Equation (3).

$$r_1(t) = x(t) - I_1(t) \quad (3)$$

Then, regard $r_1(t)$ as the up-dating original time series and repeat steps 1–3 to obtain IMF $I_1(t)$, $I_2(t), \dots, I_n(t)$ and finally get the residual series $r_n(t)$ which is a monotonic function or a function with only one extreme value from which no more IMF can be obtained. The original time series can be then expressed by Equation (4)

$$x(t) = \sum_n x_n(t) + r(t) \quad (4)$$

By the process above, the original time series can be decomposed into several simple IMFs and a residual sequence,



2.3. Ensemble Empirical Mode Decomposition (EEMD)

the EEMD[5]method to deal with such mode mixing problems. The true IMF components are defined as the mean of an ensemble of trials, each of which is composed of a signal and a white noise of finite amplitude. The additional white noise would distribute the time-frequency space uniformly with several components of different scales When the signal is added to the uniform white background, the signals with different scales are automatically projected onto proper scales of reference established by the white noise in the background. Therefore, this approach greatly eliminates the problem of mode mixing . Therefore, each IMF component reflects characteristic of the corresponding time scale of the original data. The implementation of EEMD can be briefly described as following

1. Set the ensemble number and amplitude of white noise added sequence.
2. Add a set of white noise to the original data with the determinate amplitude.
3. Decompose the time sequence with the added white noise in the ensemble into IMFs by EMD
4. Repeat steps 2 and 3 until all the time series in the ensemble have been decomposed. Every time a new white noise sequence is added, the final mean of the corresponding IMFs in the ensemble are the true IMFs.

2.4. Statistical Empirical Mode Decomposition (SEMD)

The (SEMD) is a function performs empirical mode decomposition using spline smoothing not interpolation for sifting process. The smoothing parameter is automatically determined by cross-validation .

2.5.Autoregressive Integrated Moving Average (ARIMA)

model can be expressed by the Equation

$$x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + y_t - (\beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_q y_{t-q})$$

where c is constant, αi (i = 0, 1, . . . , p) is the autoregressive coefficient, βi(i = 0, 1, . . . , q) is the moving average coefficient; {εt} is a white noise process which is denoted as $\varepsilon t \sim N(0;\sigma^2)$, p represents the lag order of the autoregressive processes. q represents the lag order of the moving average processes, d represents the d-th difference

3. Measurements of study

The Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are used to estimate the performance of hybrid models, as defined in these Equations

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{x_i - \hat{x}_i}{x_i}\right| \times 100$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i|$$

4. Case Study:

Sirte enjoys a unique location in the middle of Libya and hosts a large part of the southern coast of the Mediterranean Sea for a length of 400 kilometres and an area of the city of Sirte 69000 square kilometres. The coastal areas of the city of Sirte are characterized by moderate Mediterranean weather for most seasons of the year. The areas away from the coast are characterized by being very hot summer and very cold winter and figure (2) shows the geographical location of the city of Sirte

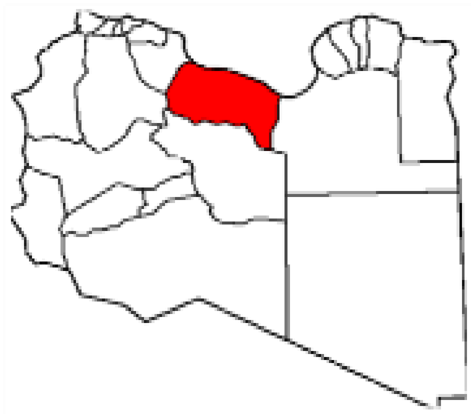


figure (2) shows the geographical location of the city of Sirte

In this study, the data used for temperature were recorded in Sirte during the period 1945 - 2010. Figure (3) shows the time series of temperatures.

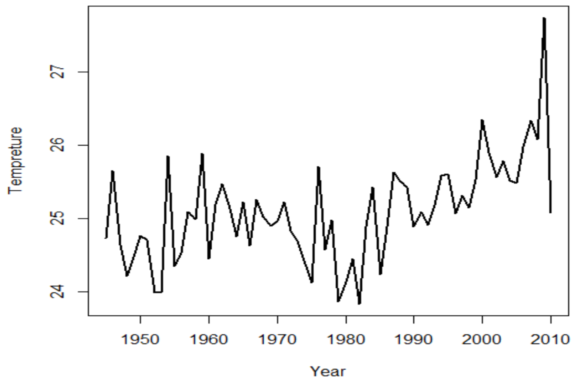
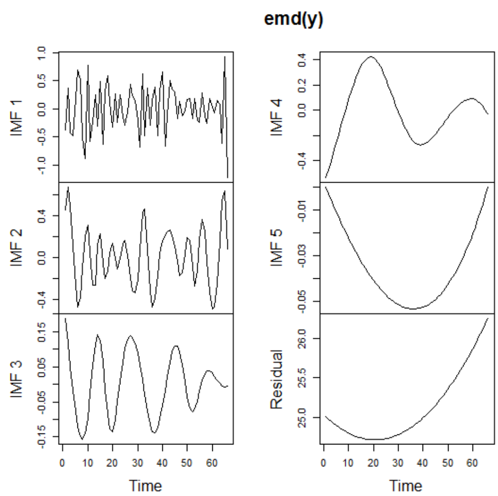


Figure (3) shows the time series of temperatures

Obviously, this series is unstable. We have applied the three methods (EMD \\ EEMD \\ SEMD) to analyse and disassemble the original series to make it appear in the stability image as shown in Figures (4), (5) and (6).



Fig(4) IMF of EMD

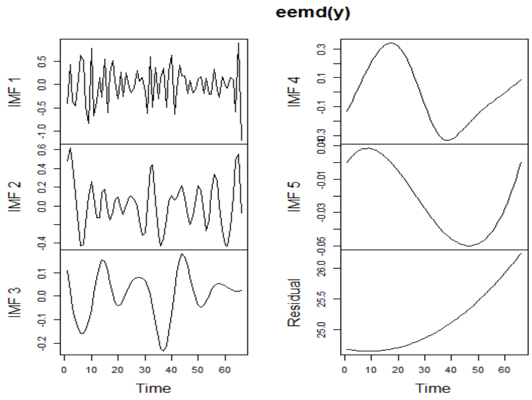


Fig (5) IMF of EEMD

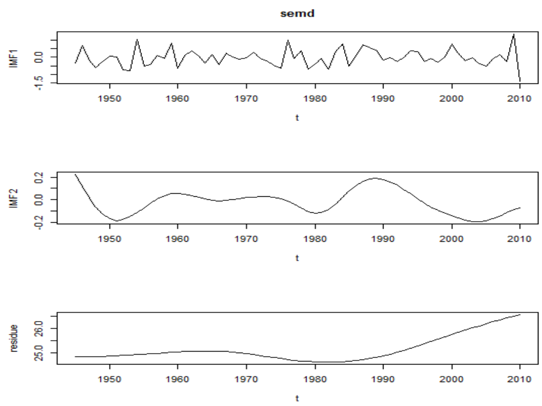


Fig (6) IMF of SEMD

4.1. RESULT

The three models whose series are shown in Figures (3), (4) and (5) are applied ARIMA model for the purpose of obtaining future forecasts and we have taken 10 years to predict the temperature of the city of Sirte

Table (1) shows the statistical accuracy of the predictions by different models, indicated by MSE, MAE and MAPE.

MODEL	TEN YEARS		
	EMD-ARIMA	EEMD-ARIMA	SEMD-ARIMA
MSE	**0.8079**	**0.6345**	**0.4595**
MAE	**0. 5758**	**0.6911**	**0.4526**
MAPE	**2.1562**	**2.6747**	**1.7167**

Table(1) the results

4.2. Discussion and Conclusions

In our study of temperature prediction in Sirte for the next ten years, we have noticed that the SEMD WITH ARIMA method has the lowest values in all three measures MSE, MAE and MAPE used in the study. So we can say that the SEMD with ARIMA method is the best of the three methods used when predicting a future 10 years. For further investigation one could try other number of (h) .

5. References

[1]- C. Chatfield, (1996). Network, (fifth edition.) [online]. Available at: https://book.onepdf.us/time-series-chatfield.pdf

[2]- N. E. Huang, Z. Shen, S. R. Long, (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis.

[3]- Amar. A, Guennoun. Z, (2012). Contribution of wavelet transformation and empirical mode decomposition to measurement of U.S core inflation.

[4]- M. Ismail, A. Awajan, (2017). A new hybrid approach EMD-EXP for short-term forecasting of daily stock market time series data.

[5]- W. Zhi-Yu, Q. Jun, L. Fang-Fang, Hybrid Models Combining EMD/EEMD and ARIMA for Long-Term Streamflow Forecasting.