

Annual Forecasting Using a Hybrid Approach

Qais Mustafa Abdulqader

Duhok Polytechnic University, Technical College of Petroleum and Mineral Sciences, Zakho, Iraq
qais.mustafa@dpu.edu.krd

Abstract In this paper, we used a hybrid method based on wavelet transforms and ARIMA models and applied on the time series annual data of rain precipitation in the Province of Erbil-Iraq in millimeters. A sample size has been taken during the period 1970 - 2014. We intended to obtain the ability to explain how the hybrid method can be useful when making a forecast of time series and how the quality of forecasting can be enhanced through applying it on actual data and comparing the classical ARIMA method and our suggested method depending on some statistical criteria. Results of the study proved an advantage of the statistical hybrid method and showed that the forecast error could be reduced when applying Wavelet-ARIMA technique and this helps to give the enhancement of forecasting of the classical model. In addition, it was found that out of wavelet families, Daubechies wavelet of order two using fixed form thresholding with soft function is very suitable when de-noising the data and performed better than the others. The annual rainfall in Erbil in the coming years will be close to 370 millimeters.

Keywords: ARIMA, De-noising, Forecasting, Time series, Wavelet transforms
2000 MSC No: 97K80, 65T60, 37M10.

1 Introduction

Rainfall forecasting is one of the most challenging objects. Many algorithms have been developed and proposed but still an accurate prediction of rainfall is very difficult. (Tantaneet et al., 2005), presented in their study a new procedure for predicting rainfall and depending on a combination of wavelet analysis and conventional autoregressive AR model. The research showed that the wavelet autoregressive model procedure gives a better prediction of annual rainfall than the classical AR model. (Al-Safawi et al., 2009) have estimated the autoregressive model using wave shrink. The results showed that the suitable model using classical ARIMA method is AR(6) and this model has improved when using wave shrink technique and especially when using Haar wavelet with a soft threshold to forecast the quantity of the annual rainfall in Erbil city for the period 1992-2007. (Al-Shakarchy, 2010) applied the factor analysis for forecasting two series representing rain rates and relative humidity in Mosul province. Results showed that the suitable model for the two series is ARIMA(0,0,1) and ARIMA(1,0,0) respectively. (Ali, 2013) used ARIMA method for analyzing and forecasting of Baghdad rainfall. It is found that the seasonal model SARIMA(2,1,3) \times (0,1,1) is the best model and according to this model, rainfall forecast for the coming years was also prepared and presenting and showing similar trend and range of the real data. (Venkata Ramana et al., 2013) searched to obtain a good model for monthly prediction of rainfall data by using hybrid technique consisting of the wavelet technique with artificial neural network ANN. The results of the analysis showed that the performances of the obtained models are more efficient than the ANN models. (Shoba and Shobha, 2014) have made an analysis of various algorithms of data mining used for rainfall prediction model. The study showed that sometimes when certain algorithms are combined,

they perform better and are more effective. (Eni and Adeyeye, 2015) applied seasonal ARIMA method for building a suitable model and to forecast the rainfall data in Warri Town, Nigeria. Results showed that seasonal model ARIMA (1, 1, 1) (0, 1, 1) is adequate depending on some statistical criteria.

Recently, (Shafaei et al., 2016) offered some techniques for testing their capability of predicting the monthly precipitation such as wavelet analysis WA, seasonally mixed model SARIMA and ANN method. The study concluded that searching for the effect of decomposition level on model performance, it was indicated that going from 2 to 3 decomposition levels increased the correlation between observed and estimated data, but no significant difference was found between predictions from 2 and 3 level models. (Ramesh Reddy et al., 2017) applied ARIMA model to forecast the monthly mean rainfall of coastal Andhra -India. They found that the best model for fitting data is ARIMA (5,0,0)(2,0,0) depending on some performance criteria. (Ashley et al., 2017) applied DCT presenting the discrete cosine transform and DWT presenting discrete wavelet transform to make a reduction in the 5 dimensionalities of rainfall time series observations. The conclusions of the research demonstrated that the DWT has the superiority to the DCT and best preserves and characterizes the observed rainfall records of the data.

From the above-suggested methods, we observe that most of these approaches and methods are applying to forecast the short period. This paper offers a new technique for forecasting the long-range of annual rainfall data. In another word, it mainly deals with combining wavelet transformation with classical ARIMA methodology for modeling of annual rain precipitation based on the available data. The procedure of this paper is prepared as the following: First, we provide brief explanations of ARIMA methodology and wavelet transformation and then we offer the hybrid method. Next, we deal with application on real data. Finally, we present some conclusions of the study.

2 ARIMA Methodology, Wavelet Transformation, and Hybrid Method

2.1 ARIMA Methodology

Box-Jenkins suggested an approach for analyzing time series data including an identification of the model, parameters estimation, diagnostic checking for the suggested model, and using the model for forecasting. ARIMA model is a mixed model which depends on parameters p , d , q representing a combination of autoregressive order part (AR); the degree of difference involved and the moving average order part (MA) respectively. The model becomes popular by (Box et al., 1970) and can be well explained through the mathematical formula:

$$F_t = \sum_{i=1}^p \phi_i x_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

Here, p represents a non-seasonal autoregressive order, q is a moving average order of the non-seasonal, ϕ_i are called coefficients of autoregressive, θ_j are coefficients of moving average and ε_t is a random error. If the data are not stationary, then the difference of first or second order has to be taken. For obtaining a convenient model, we depend on two functions called ACF as Autocorrelation Function and PACF as Partial Autocorrelation Function. The pattern of both functions plot provides us an idea towards which one of the specified model could be the best for fitting and appropriate for making a prediction and depending on some statistical performance. Also, in this study, we will apply the Portmanteau test statistic (i.e. Box-Pierce) for the purpose of randomness of time series. We refer to (Makridakis et al., 1998) for more details.

2.2 Wavelet Transformation

A wavelet transformation is a proceeding subject, very efficient, and effective in the field of processing the signal that has been very interest after developing the theory of wavelet methodology

(Grossman and Morlet, 1984). applications of wavelet analysis have increased in many fields such as in edge detection, image compression, optical engineering, and the applications of time series as alternate to the classical Fourier transformation in local maintain, not involving cyclic and multi-scaled phenomenon. Wavelets can give the specific locality of any changes in the dynamical patterns of the sequence, while the transformations of Fourier focus essentially on their frequency and this is the major difference between wavelets analysis and Fourier analysis. in addition, the transformation of Fourier supposes unlimited length signals, while the transformation of wavelet can be used to any form and any size of time series data, even when these time series are not identically sampled (Antonios and Constantine, 2003). Generally, wavelet transforms can be applied for seeking, reducing the noise and filtering time series data which help and also support forecasting and other analysis of the experiment. The formula of wavelet transform can be presented as the following:

$$f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \varphi\left(\frac{t-b}{a}\right) dt \quad (2)$$

Here, $\varphi(t)$ represents the essential wavelet with efficient length (t) that is commonly much shorter than the target time series $f(t)$, 'a' represents the scale factor or dilation that specifies the information of characteristic frequency so that its variation yields increase to a spectrum and 'b' represents the translation of time information so, its difference displays the 'sliding' of the wavelet over $f(t)$ (Burrus et al., 1998).

2.3 Hybrid Method

The concept of the suggested method is based on combining ARIMA methodology with wavelet transforms. As the wavelet approach can be easily used for signal analysis, this study used the approach to decompose the details (which are small differences) from the approximations (which represents the important part) of data. In wavelet analysis, the approximations are the high-scale and limited frequency components of the signal, and the details represent the limited-scale and high-frequency components (Fugal, 2009). The process is done by applying discrete wavelet transform DWT because the data of the study are recorded in discrete time. The procedure of hybrid method can be expressed in fig.1

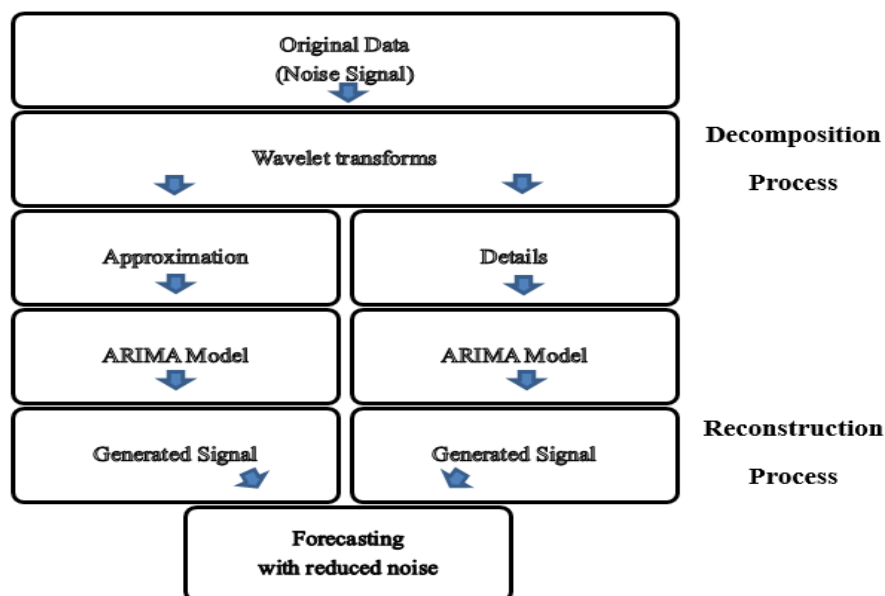


Fig. 1: The process of hybrid method

3 Application

3.1 Information About Erbil City

Erbil which is the Kurdish central is the capital city of Kurdistan Region in Iraq. The city is located between (36°12'17"N 44°20'33"E). It is located about 350 kilometres north of Baghdad. The climate of Erbil is very hot in summer and very cold and wet in winters. There is more rainfall in the winter than in the summer in Erbil. The average total of receiving rain of the city is between 300-400 millimetres annually. The city represents the managerial centre of Erbil province. It is bounded from the north by Turkey and nearby Duhok Province, from the east by Iran and near to Sulaymaniyah Province, from the south, is close to Kirkuk province, and near to Mosul province from the west (Wahab and Khayyat, 2014).

3.2 Application Using ARIMA Methodology

The variable used in the analysis represents the annual data of rain precipitation in Erbil province in Kurdistan Region of Iraq (in millimeters) and represents taking (45) observations as sample size during the period 1970 - 2014 which is shown in table1. The data were obtained from the General Directorate of Meteorology and Seismic Monitoring in Erbil province.

Table 1: Annually data on rain precipitation from 1970 to 2014

Year	Amount of Rain	Year	Amount of Rain
1970	255.4	1993	601.6
1971	448.2	1994	583.0
1972	406.4	1995	494.4
1973	261.5	1996	418.9
1974	547.5	1997	441.6
1975	417.2	1998	337.2
1976	452.3	1999	229.2
1977	347.2	2000	272.3
1978	380.1	2001	330.9
1979	375.6	2002	361.5
1980	321.5	2003	587.7
1981	141.8	2004	255.6
1982	444.1	2005	297.5
1983	178.3	2006	514.6
1984	43.9	2007	273.4
1985	463.9	2008	410.7
1986	154.0	2009	411.0
1987	235.9	2010	359.6
1988	626.9	2011	301.6
1989	367.3	2012	366.4
1990	332.0	2013	345.2
1991	344.1	2014	385.2
1992	694.0		

Fig.2 shows the plots of time series of rain data for Erbil city. Depending on Box-Jenkins procedure, the first step to do is identification through employing the ACF and PACF plots which are clear in fig. 3.

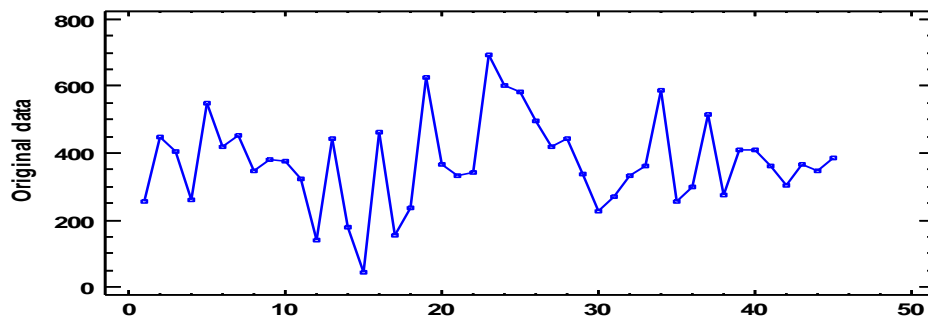


Fig. 2: Time series plot of rain data in Erbil province from 1970 to 2014

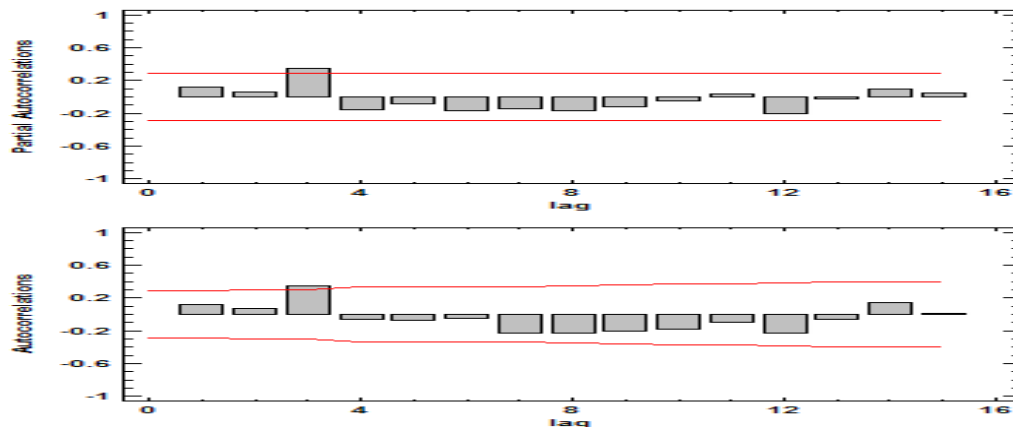


Fig. 3: ACF and PACF of rain data

Depending on PACF and PACF plots and checking for stationarity in mean and variance, the appropriate model for the respected series is identified as ARIMA(2,1,0) after well consideration of modelling and fitting and depending on two performance measures such as RMSE as root mean square error and MAE as mean absolute error. The estimated model is shown in table2.

Table 2: Estimation of ARIMA(2,1,0)

Parameter	Estimates	Std. Error	t-ratio	P-value
AR(1)	-0.72091	0.129125	-5.58304	0.000002
AR(2)	-0.540025	0.128616	-4.19875	0.000136

After getting the estimation of the ARIMA (2,1,0) model, we should look for getting randomness. Fig. 4 offers the residuals pattern and stability of ACF and PACF inside the intervals using classical ARIMA (2,1,0).

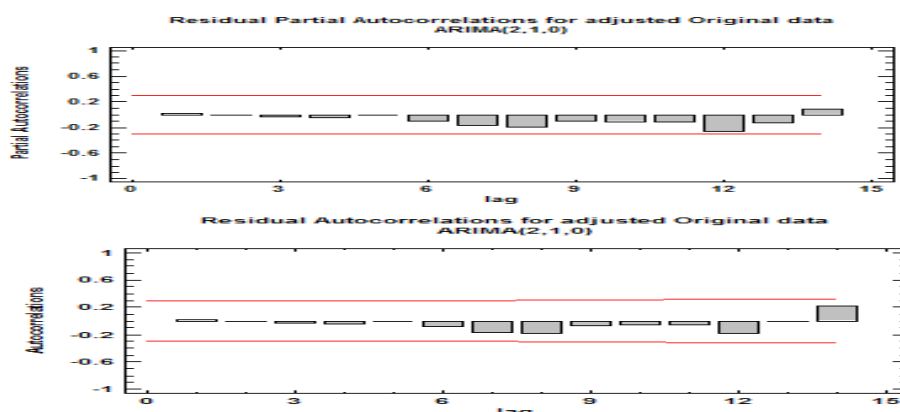


Fig. 4: ACF and PACF of residuals using ARIMA(2,1,0) on series data.

From Figure 4, it is concluded that the time series is random because there is no significant appear from the coefficients of ACF and PACF. Concerning the randomness of residuals, we did a test using a Portmanteau test, which has been mentioned in theoretical part. The value of the test was (7.326) comparing to the P-value (0.835) indicates that the hypothesis cannot be rejected at the 95% or higher confidence level and concluding that the series is random.

3.3 Application Using a Hybrid Method

In this part, the conversion of original data from time domain to frequency domain has been done to make filtration. Fig. 5 shows applying Daubechies wavelet with multiresolution of five levels for the rain precipitation for 45 values as sequential observations, denoting s as a signal and it means the summation of signal approximation and its details, a_5 is an approximation at level 5 and the details levels are d_5 ; d_4 ; d_3 ; d_2 ; and d_1 respectively.

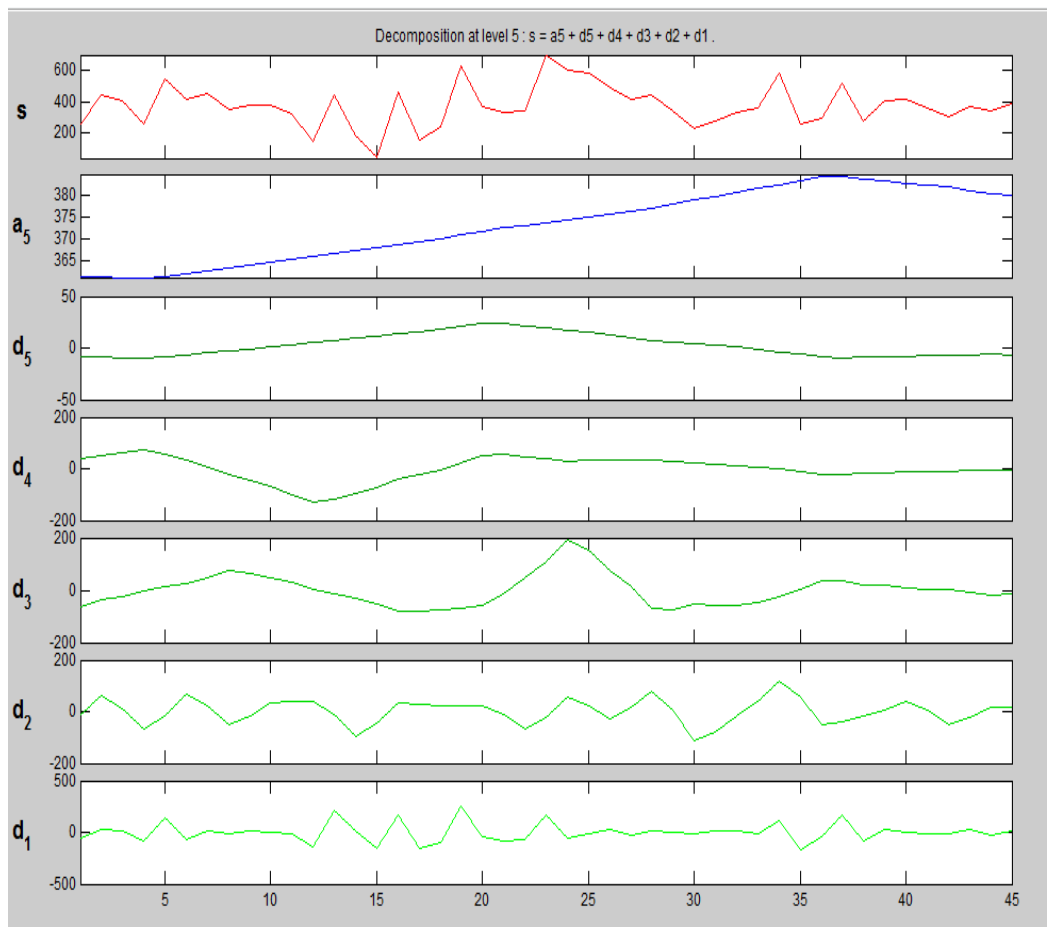


Fig. 5: Daubechies wavelet of the rain precipitation using multiresolution of five levels.

The real data of rain precipitation were reduced from noise using wavelet denoising procedure (using the software MATLAB, version 2013) with Daubechies wavelet family from order 2 to order 5 as shown in fig. 6. It should be noted that after making many empirical experiments, it has been found that the performance of Daubechies wavelet was better than others in terms of de-noising the rain data. Fig. 7 shows the real and de-noised signals by applying the Daubechies wavelet with Fixed Form Threshold (Patil and Raskar, 2015).

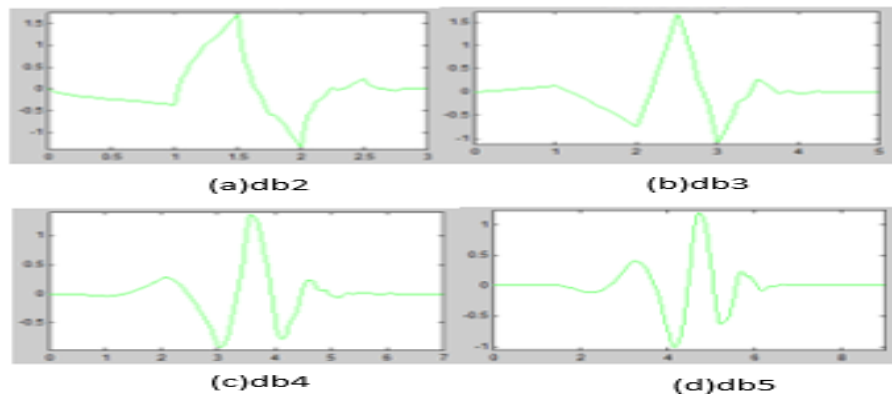


Fig. 6: Daubechies wavelet of order 2,3,4, and 5

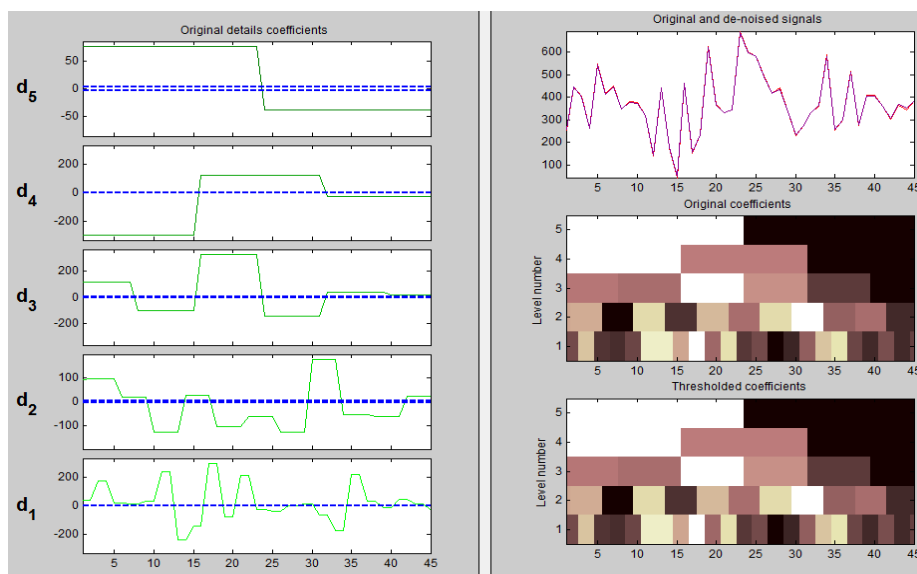


Fig. 7: The original and de-noised signals using Daubechies wavelet with Fixed Form Threshold.

The data were analysed using five levels of multiresolution for the selected wavelet, and then de-noised using Fixed Form Threshold and depending on soft thresholding. After that, the filtered series was modelled again using ARIMA methodology and compared with classical method depending on the values of forecasting criteria. Table 3 presents the comparison of the two methods using performance values of the two indicators for selecting an optimal model.

Table 3: The comparison of the performance measures for the original data model using ARIMA methodology and hybrid method.

Method	Kind	RMSE	MAE
ARIMA Method Original (raw) data	ARIMA(2,1,0)	133.937	106.565
	Daubechies(2)	131.380	104.143
Hybrid	Daubechies(3)	131.555	104.553
Fixed Form	Daubechies(4)	131.593	104.411
De-noised data	Daubechies(5)	131.706	104.546

From Table 3, we observe that the best model for the original data was ARIMA(2,1,0). However, when the hybrid method applied to the original data the errors of the forecasting have decreased for all wavelet orders and the new models have been enhanced depending on the forecasting measures. To make a comparison of the two procedures, we observe that the maximum reduction is when applying Fixed Form Thresholding and using Daubechies wavelet of order 2 (i.e.; the good reduction in RMSE and MAE from 133.937 to 131.380 and from 106.565 to 104.143, respectively). Fig. 8 presents the original and filtered data using Daubechies wavelet of order 2.

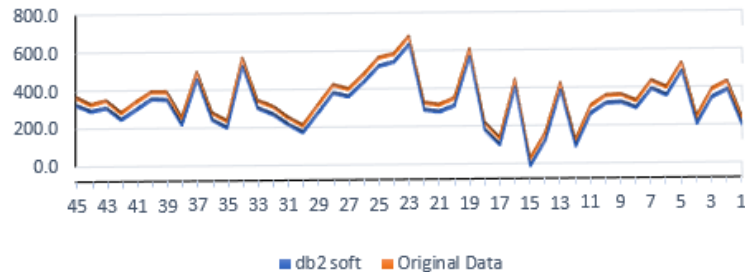


Fig. 8: The original and filtered signals using Daubechies wavelet of order 2

The forecast values of our hybrid method are presented in table 4 which shows the forecasting for the next years starting from 2015 up to 2030 of the annual rain precipitation (in millimeters) of Erbil province – Iraq.

Table 4: Forecast values of the annual rain of Erbil province-Iraq using hybrid method

Period	Forecast
2015	367.8
2016	360.3
2017	373.5
2018	368.1
2019	364.9
2020	370.1
2021	368.1
2022	366.7
2023	368.8
2024	368.0
2025	367.4
2026	368.3
2027	368.0
2028	367.7
2029	368.1
2030	368.0

4 Conclusions

In this research, we offered a new technique as hybrid method for enhancing the Box-Jenkins ARIMA analysis when forecasting time series data. Indeed, we concluded that:

- [1] The appropriate model for forecasting using classical Box – Jenkins method was ARIMA(2,1,0).
- [2] The classical model has been enhanced and improved when making filtration of the data and using Daubechies wavelets orders from 1 to 5 and among them, the Daubechies wavelet of order 2 gave results better than others.
- [3] Depending on our hybrid method to forecast for the coming years, the Erbil city will receive an average total rainfall of 360-370 millimeters annually.

References

- [1] S.M. Ali, Time series analysis of Baghdad rainfall using ARIMA method, Iraqi Journal of Science, 54(2013), 1136-1142.
- [2] S. Al-Safawi, T. Ali & M. Badal, Estimation AR(p) model using wave shrink, Second Scientific Conference of Mathematics - Statistics and Informatics, University of Mosul, (2009), 274-299.
- [3] D.H. Al-Shakarchy, Using factor analysis to forecast of time series with an application on two series rain rates and relative humidity in Mosul city, Tikrit Journal of Administrative and Economic Sciences, 6(2010), 93-108.
- [4] A. Antonios & E.V Constantine, Wavelet exploratory analysis of the FTSE ALL SHARE index, In Proceedings of the 2nd WSEAS international conference on non-linear analysis, Non-linear systems and Chaos, Athens, (2003)
- [5] W. Ashley, J. P. Walker, D. E. Robertson & V. R.N. Pauwels, A Comparison of the discrete cosine and wavelet transforms for hydrologic model input data reduction, Journal of Hydrology and Earth System Sciences, 6(2017), 1-23, <https://doi.org/10.5194/hess-2017-26>
- [6] G. Box, G. Jenkins & G. Reinsel, Time series analysis: Forecasting and control, third edition, Prentice-Hall International Inc., New Jersey, USA, (2008).
- [7] C. Burrus, R. Gopinath & H. Guo, Introduction to wavelet and wavelet transforms, Prentice Hall, New Jersey, USA, (1998).
- [8] D. Eni & F. Adeyeye, Seasonal ARIMA modeling and forecasting of rainfall in Warri Town, Nigeria, Journal of Geoscience and Environment Protection, 3(6)(2015), 91-98, <https://doi.org/10.4236/gep.2015.36015>
- [9] D. Fugal, Conceptual wavelets in digital signal processing, Space and Signals Technologies LLC, San Diego, California, (2009).
- [10] A. Grossman & J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape, SIAM Journal of Mathematical Analysis, 15(4)(1984), 723-736, <https://doi.org/10.1137/0515056>
- [11] S. Makridakis, S. Wheelwright & R. Hyndman, Forecasting methods and applications, Third edition, Wiley& Sons, Inc, New York, (1998).
- [12] P. L. Patil & V. B. Raskar, Image denoising with wavelet thresholding method for different level of decomposition, International Journal of Engineering Research and General Science, 3 (2015), 1092-1099.

- [13] J. C. Ramesh Reddy, T. Ganesh, M. Venkateswaran & P. Reddy, Forecasting of monthly mean rainfall in Coastal Andhra, *International Journal of Statistics and Applications*, 7 (2017), 197-204.
- [14] M. Shafaei, J. Adamowski, A. Fakheri-Fard, Y. Dinpashoh & K. Adamowski, A wavelet-SARIMA-ANN hybrid model for precipitation forecasting, *Journal of Water and Land Development*, 28(1)(2016), 27-36, <https://doi.org/10.1515/jwld-2016-0003>
- [15] G. Shoba & G. Shobha, Rainfall prediction using data mining techniques: A survey, *International Journal of Engineering and Computer Science*, 3(2014), 6206-6211.
- [16] S. Tantanee, S. Patamatammakul, T. Oki, V. Sriboonlue & T. Prempre, Coupled wavelet-autoregressive model for annual rainfall prediction, *Journal of Environmental Hydrology*, 13(2005), 1-8.
- [17] R. Venkata Ramana, S. Krishna, R. Kumar & N. G. Pandey, Monthly rainfall prediction using wavelet neural network analysis, Springer, *Water Resource Manage*, 27(10)(2013), 3697–3711, <https://doi.org/10.1007/s11269-013-0374-4>
- [18] S. Wahab & A. Khayyat, Modeling the suitability analysis to establish new fire stations in Erbil City using the analytic hierarchy process and geographic information systems, *Journal of Remote Sensing and GIS*, 2(2014),1-10.