Al-Rafidain Journal of Computer Sciences and Mathematics (RJCSM)

www.csmj.uomosul.edu.iq

Dealing with Outliers in ARMA Time Series Analysis Using Hampel Filter and Wavelet Analysis

Heyam A. A. Hayawi[1], Jwana R. Qadir[2], Taha H. Ali[3] and Saif R. Ahmed[4]

[1] Department of Statistics and Informatics, College of Computer Science and Mathematics, Mosul University- Mosul, Iraq
[2] Department of Medical Laboratory, College of Science, Raparin University- Ranya, Kurdistan Region, Iraq
[3] Department of Statistics, College of Administration and Economics, Salahaddin University- Erbil, Iraq
[4] Ministry of Planning, Authority of Statistics & Geographic Information Systems, Nineveh Statistics Office, Mosul, Iraq
Email: he.hayawi@uomosul.edu.iq[1], Jwana.Rostam93@gmail.com[2], taha.ali@su.edu.krd[3] and saiframz525@gmail.com[4]

Article information

Article history:
Received 07 February ,2025
Revised 20 April ,2025
Accepted 07 May ,2025
Published 26 June ,2025

Keywords:
Time Series,
ARMA Model,
Outliers,
Hampel Filter,
Wavelets.

Correspondence:
Heyam A. A. Hayawi
Email:
he.hayawi@uomosul.edu.iq

Abstract

Outliers affect the accuracy of the estimated parameters of ARMA time series models which can be handled by the Hampel filter. In this article, wavelet shrinkage is proposed to handle outliers of ARMA models by using wavelet (Daubechies for order 4, Symlets for order 1, and Dmey) with a universal threshold method and applying a soft threshold. To compare the efficiency of the proposed method and the traditional method (Hampel filter), the mean square error, Akaike and Bayes information criteria were calculated for simulated and real data (The wind speed series data). The proposed method addresses the problem of outliers and provides estimated parameters for ARMA models with higher efficiency than the traditional method.

DOI: 10.33899/csmj.2025.157319.1173, ©Authors, 2025, College of Computer Science and Mathematics, University of Mosul, Iraq.

1. Introduction

Time series include all phenomena that change depending on the change of time. This change is either regular at points in time equidistant from each other, such as annual population growth or irregular change, such as changes in the volume of production, and any time series is characterized by the fact that its data are arranged relative to time and successive observations are usually not independent,[13][19] i.e. dependent on each other. This lack of independence will be exploited in reaching reliable predictions, A time series is defined mathematically as a semantic relationship between the value of the phenomenon under study and time, and the time series is usually of two types, either an intermittent or continuous time series, Time series analysis is the process of separating its components from each other and to determine the impact of each of these components on the values of the phenomenon under study [3][11]. Time series are characterized by the presence of some common models in them, most notably the self-regression model and moving averages developed by both Box and Jenkins, [7] where they assumed that part of the series is self-regression, and the other part is moving averages and merging these two models with the model of self-regression and moving averages. One of the most important problems that time series suffer from is the instability caused sometimes by the presence of extreme values or so-called abnormal values, which have a significant and obvious impact on the process of analyzing the time series, these abnormal values are often caused by a defect in the data collection process or the presence of some unexpected events that significantly affect the analysis of the time series and thus the predictive values resulting from the analysis process,[22] the best way to get rid of the problems that occur during the analysis and prediction of the presence of abnormal values is to get rid of these values or and filters The use of these methods does not affect the data or the process of analysis and prediction of the phenomenon under study[15][16]. In this article, wavelet shrinkage is proposed to handle outliers of ARMA (Autoregressive and Moving Average) models by using wavelet (Daubechies for order 4, Symlets for order 1, and Dmey) with a universal threshold method and applying a soft threshold.

2. ARMA Model

Both Box and Jenkins presented a method in 1976 called self-regression models and mixed moving averages, and these models were designed to be used in forecasting and assuming that the time series is part of self-regression and the other part is moving averages to get from us the general model of the time series, which is symbolized by ARMA and can be calculated from the following equation [4][8][21]:

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (1)$$

Since:p, q: represent the rank of the model, $\emptyset_j, \theta_i$: Model parameters, a_t: Represent random variables that are not related to each other, called white noise, and have an average of zero and variance σ_a^2 [20].

3. Outliers

Researchers sometimes face a set of statistical problems, some of which may be obvious and others unclear, so the researcher finds himself in need of new methods that enable him to organize the course of the experiment by making the resulting error as small as possible and at the same time get an unbiased estimate of the amount he is looking for the idea of studying Outliers began with simple ideas based on intuition and guessing [6][23].

Outliers are defined as those observations that seem illogical and show a significant deviation from the other components of the sample in which that observation was found [2][14]. It was stated by Barnett that the Outliers observation in a set of data is an observation that seems illogical when compared to the rest of the data set, the Outlier's values have been defined by many researchers, but all definitions are It boils down to one concept, which is that an Outliers viewing is a viewing that is inconsistent with the rest of the views [18].

The researchers pointed out that it is important to examine the data to get rid of the influence of Outliers before entering statistical analysis, as statistical analysis in all practical respects depends mainly on the selection of a set of data and the purification of these data from Outliers, which constitute a clear deviation from the rest of the observations [11].

4. Hampel Filter

The humble filter is considered a statistical tool or means to detect and get rid of Outliers in the data set, and it is also a method through which the method of deviation is improved on the traditional standard because it has a high ability to deal with data that are not distributed normally, and This filter is easy to use in various fields such as time series, signal analysis and others [12]. The Hampel identifier is a robust statistical filter using median absolute deviation (MAD), for a given window size k around t, compute:

1. Local Median: $m_t = median(y_{t-k}, \dots, y_t, \dots, y_{t+k})$
2. Median Absolute Deviation (MAD): $\sigma_{mad} = median(|y_{t-i} - m_t|) \ , \ \ i = -k, \dots, k$
3. Threshold for Outliers: $|y_t - m_t| > 3\sigma_{mad}$, If true replace y_t with m_t .

Since MAD is more robust than standard deviation, this removes spiky outliers without distorting smooth trends.

5. Wavelets

Wavelet contraction is a method to remove noise and reduce noise in signals can be reduced using wavelet shrinkage, a technique that involves thresholding wavelet coefficients [24]. The wavelet deflation method was introduced for generalizing curve estimation problems. There are multiple compelling reasons to employ wavelet contraction in the estimation function.

5.1. Daubechies Wavelet

In the year (1992), researcher Ingrid Daubechies (DB), who is famous for her work with wavelets, named this wavelet after her. It is generated from a group of wavelets to improve the properties of a frequency field [17]. One of the features of this wavelet is the smoothness we have given it by using the smallest possible number of parameters, it is by (DN).

5.2. Symlets Wavelet

The researcher . Daubechies proposed the samelt wave, which is an orthogonal wave approaching symmetry, through which some modifications were made to the Dupuis family , as the symmetry is increased while the simplicity of the wave remains [9][1]. This symmetry is useful because it reduces the noise in data reconstruction. Symlets have compact support and orthogonal.

5.3. Dmey Wavelet

The Dmey wave is a modified version of the Daubechies wave, but with limited support and defined separately from the rest of the Daubechies family. It has similar characteristics to Daubechies waves in terms of compact support, but it has a different shape that makes it suitable for some applications, such as signal processing and image analysis. [5]. The number of vanishing moments directly affects its ability to capture polynomial trends within the data, thereby enhancing its suitability for various tasks such as noise reduction, compression and feature extraction. The effectiveness of DME waves is evidenced by case studies in signal reconstruction and noise reduction.

6. Proposed Method

wavelet analysis used in handling outliers in time series models and estimating ARMA model parameters depends on discrete wavelet transform (Daubechies, Symlets, and Dmey) to obtain approximation and detail coefficients [10]:

$$y_t = \sum_{i=1}^{L} A_i + \sum_{j=1}^{L} D_j \tag{2}$$

A_i represents the low-pass filter at level i (or the approximation coefficients).
D_j represents the high-pass filter at level j (or the detail coefficients).

The next step is to apply a thresholding operation to the detail coefficients to suppress the outliers. Outliers typically have small coefficients in the wavelet domain, so we shrink or remove coefficients below a certain threshold. The soft thresholding technique (Threshold rule) is applied to the detail coefficients:

$$D_j \to wthresh\left(D_j', s', \lambda\right) \tag{3}$$

$wthresh$ applies soft thresholding and λ is the threshold value, determined using thresholding universal based on Donoho's method:

$$\lambda = \sigma\sqrt{2Log(n)} \tag{4}$$

Where n is the length of the time series (the number of time points) and σ is the estimated standard deviation for the wavelet coefficients. Specifically, it is computed as the median of the absolute values of the wavelet coefficients at the last level, normalized by 0.6745 to make it consistent with an unbiased estimate of the normal distribution standard deviation.

$$\sigma = \frac{Median(|Coeffs[L]|)}{0.6745} \tag{5}$$

After thresholding the detail coefficients, the time series data is reconstructed using the inverse discrete wavelet transformation. This reconstructs the time series data processed from outliers:

$$yw_t = Waveletrec(Coeffs, L, Wavelet) \tag{6}$$

$Coeffs$ contains the approximation and threshold detail coefficients, L is the level of decomposition, and *Wavelet* is the same wavelet used in the decomposition (Daubechies, Symlets, and Dmey). This inverse discrete wavelet transformation reconstructs the time series data by combining the approximations, which are kept intact and the modified details, which have been threshold to handle outliers.

Using maximum likelihood estimation (MLE) to estimate the parameters models, based on the input data (yw_t). This might involve using methods like Yule-Walker for the AR parameters and a simple moving average for the MA parameters.

The likelihood function L(θ) is typically given by:

$$L(\theta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left(-\frac{\hat{\epsilon}_t^2}{2\hat{\sigma}^2}\right) \tag{7}$$

$$\hat{\epsilon}_t = yw_t - \hat{y}w_t \tag{8}$$

$\hat{\epsilon}_t$ is the residual (error term) at the time t and $\hat{\sigma}^2$ is the estimated variance of the residuals. The goal is to find the values of parameters θ that maximize the likelihood function, i.e., minimize the negative log-likelihood. This is generally achieved through iterative optimization techniques like Newton-Raphson. After optimization, the estimated values of the AR and MA parameters, the constant (mean) and the residual variance $\hat{\sigma}^2$. These estimates are used to define the final ARMA model.

7. Simulation Study

The time series is simulated by specifying the AR (1 and 2) and MA (1 and 2) coefficients and using the ARMA function in MATLAB with different sample generations (200, 300, and 400), Using the MATLAB program. **Table 1** shows the assumed coefficient values for the simulation models.

Table 1. The Assumed Coefficient Values for the Simulation Models

Model	AR		MA	
ARMA (1, 1)	0.7		0.5	
ARMA (2, 1)	0.7	-0.3	0.5	
ARMA (1, 2)	0.7		0.5	-0.4
ARMA (2, 2)	0.7	-0.3	0.5	-0.4

Outliers are typically defined as data points that deviate significantly from the rest of the dataset. In the time series, outliers can occur for various reasons such as data entry errors, sudden changes in the system being measured, or external disturbances. In this step, artificial outliers are injected at specific time points to simulate noise or contamination in the data. Outliers are added by identifying specific indicators in the time series and replacing their values with values that are significantly larger or smaller than the expected range. Applies techniques of Hampel filter (with a window size of 5 and threshold parameter of 3) and wavelet shrinkage to remove the outliers. To simulate the first experiment, the generated time series data with the estimated models without using a filter, Hampel filter and the proposed method (Daubechies4, Symlets1, and Dmey) were plotted as in **Figures (1-3).** Wavelet decomposition is used for denoising, using the Daubechies wavelet (DB4), Symlets1, and Demy wavelet at level 4. A universal threshold is calculated for soft thresholding of the detail coefficients, and the time series data for ARMA (2, 1) is reconstructed using inverse wavelet transform:

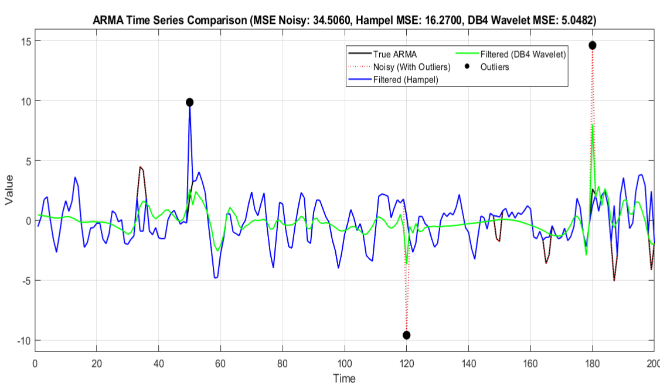

Figure 1. ARMA (2, 1) Models for the first simulation experiment (DB4)

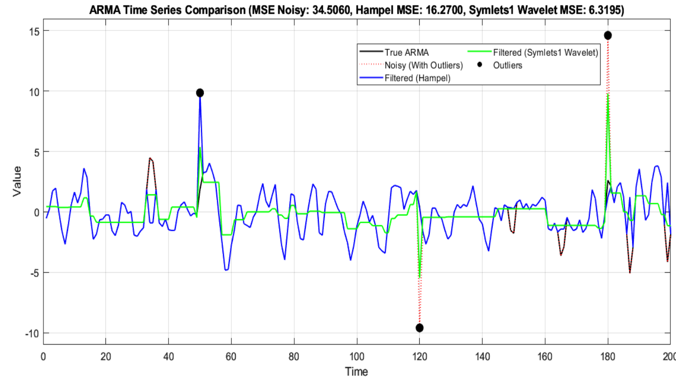

Figure 2. ARMA (2, 1) Models for the first simulation experiment (Symlets1)

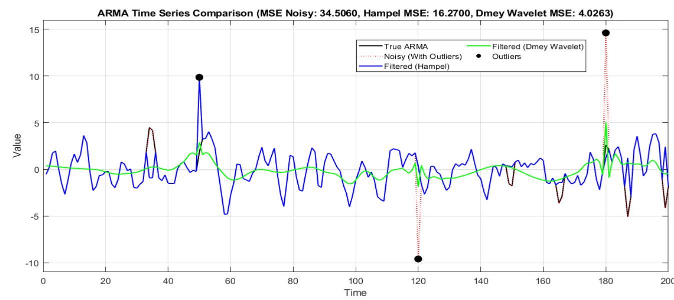

Figure 3. ARMA (2, 1) Models for the first simulation experiment (Dmey)

Figures (1-3) show that True ARMA Series (Black Line): This is the original, clean ARMA (2,1) time series that was simulated using the specified AR and MA coefficients. It represents the "true" underlying process without any noise or outliers. In the plot, it's shown as a black line. This serves as the baseline for comparing all other time series. Noisy series with outliers (Red Dashed Line): This line represents the noisy time series that was generated by injecting outliers at specific positions in the true ARMA time series. The outliers are large, with extreme values added to the series at indices 50, 120, and 180, The outliers injected into the noisy series are highlighted as black dots on the plot. These dots correspond to the positions where large spikes were added in the noisy time series (indices 50, 120, and 180). Visually identify these outliers by looking for the large deviations from the rest of the series at these points. The red dashed line in the plot indicates this noisy series. The outliers can be seen as the spikes in the red line at the mentioned positions. Hampel filtered series (Blue Line): After applying the Hampel filter to the noisy series, this line shows the result of removing the outliers. The Hampel filter smooths the noisy time series by replacing the outliers with median values within a moving window. The blue line represents this filtered series. See that the sharp spikes from the outliers are reduced or removed, providing a cleaner series, closer to the true ARMA process. Wavelet filtered Series (Green Line for DB4, Symlets,1 and Dmey): This line represents the wavelet denoised time series. The wavelet denoising method uses the Daubechies wavelet to decompose the series into different frequency components, then apply thresholding to remove noise in the high-frequency components. The green line shows the wavelet-filtered series, which should also smooth out the outliers while preserving the underlying structure of the time series. The Hampel and wavelet filtering methods aim to reduce the impact of the outliers. The Hampel filter typically replaces outliers with the median, leading to a smooth series where the outliers are no longer visible. The wavelet filter achieves a similar goal but through a different approach by manipulating the wavelet coefficients. The title of the plot includes the MSE for the three models: The MSE for the ARMA model is estimated from the noisy data (with outliers) equal to (34.506). The MSE for the ARMA model was estimated from the Hampel-filtered data equal to (16.27). The MSE for the ARMA model was estimated from wavelet-filtered data (DB4, Symlets1, and Dmey) equal to (5.0482, 6.3195, and 4.0263) respectively. These values help quantify how well the different filtering techniques (Hampel and wavelet) performed in removing the outliers and improving the model's prediction accuracy. These results suggest that wavelet denoising is a more robust method for handling noisy data and outliers in time series modelling, particularly when the goal is to estimate ARMA parameters accurately. Subsequently, ARMA (1, 1), ARMA (2, 1), ARMA (1, 2), and ARMA (2,2) models are fitted to the noisy, Hampel-filtered, and wavelet-filtered data. The model performance is evaluated based on the AIC, BIC, and MSE average for repeated experiments (1000) times in **Tables 2-5.**

Table 2. ARMA (1, 1) Model Performance Comparison

Method	*n*	AIC	BIC	MSE
Outliers	200	818.5659	831.7592	31.2903
Hampel		611.9258	625.1191	17.6929
DB4 Wavelet		551.2921	564.4854	9.6881
Symlets Wavelet		571.7301	584.9234	8.2813
Dmey Wavelet		377.5651	390.7583	7.5245
Outliers	300	1159.900	1174.700	26.5747
Hampel		903.9810	918.7961	17.7113
DB4 Wavelet		747.9420	762.7571	8.9144
Symlets Wavelet		769.8217	784.6368	7.4290
Dmey Wavelet		600.4529	615.2680	7.3415
Outliers	400	1489.000	1505.000	24.4836
Hampel		1201.900	1217.800	17.8232
DB4 Wavelet		920.6017	936.5676	8.5394
Symlets Wavelet		951.7056	967.6714	7.0717
Dmey Wavelet		777.5241	793.4899	7.3916

Table 3. ARMA (2, 1) Model Performance Comparison

Method	*n*	AIC	BIC	MSE
Outliers	200	818.9055	835.3971	20.8607
Hampel		629.1936	645.6852	9.7609
DB4 Wavelet		498.7126	515.2042	4.0387
Symlets Wavelet		491.2683	507.7599	3.7017
Dmey Wavelet		274.0129	290.5044	3.2383
Outliers	300	1163.500	1182.100	16.9520
Hampel		934.5751	953.0941	9.7401
DB4 Wavelet		676.9350	695.4539	3.7244
Symlets Wavelet		639.0566	657.5755	3.3775
Dmey Wavelet		470.3434	488.8623	3.2227
Outliers	400	1496.300	1516.200	15.1671
Hampel		1240.100	1260.000	9.7325
DB4 Wavelet		831.1457	851.1031	3.5665
Symlets Wavelet		764.2793	784.2366	3.2492
Dmey Wavelet		613.8565	633.8138	3.2071

Table 4. ARMA (1, 2) Model Performance Comparison

Method	*n*	AIC	BIC	MSE
Outliers	200	818.7992	835.2908	21.1054
Hampel		620.1938	636.6854	9.9009
DB4 Wavelet		461.8118	478.3034	4.6615
Symlets Wavelet		504.8405	521.3321	4.4504
Dmey Wavelet		187.6904	204.1820	6.1634
Outliers	300	1164.800	1183.300	17.2154
Hampel		924.6426	943.1615	10.0036
DB4 Wavelet		608.9759	627.4948	4.2792
Symlets Wavelet		659.3379	677.8568	4.0041
Dmey Wavelet		345.8474	364.3663	4.8564
Outliers	400	1500.500	1520.400	15.4805
Hampel		1228.500	1248.500	10.0435
DB4 Wavelet		729.2719	749.2293	4.1067
Symlets Wavelet		792.9371	812.8944	3.8313
Dmey Wavelet		446.8885	466.8458	4.2054

Table 5. ARMA (2, 2) Model Performance Comparison

Method	*n*	AIC	BIC	MSE
Outliers	200	823.9167	843.7066	19.2616
Hampel		648.8131	668.6030	8.4786
DB4 Wavelet		402.0685	421.8584	3.0710
Symlets Wavelet		424.1386	443.9285	3.0273

Dmey Wavelet		123.4819	143.2718	3.0023
Outliers	300	1174.400	1196.700	15.5211
Hampel		970.5687	992.7914	8.5225
DB4 Wavelet		530.1244	552.3471	2.9119
Symlets Wavelet		530.7518	552.9745	2.8452
Dmey Wavelet		261.8567	284.0793	2.7992
Outliers	400	1515.000	1538.900	13.7899
Hampel		1284.900	1308.800	8.4618
DB4 Wavelet		634.0158	657.9646	2.8360
Symlets Wavelet		611.2408	635.1896	2.7771
Dmey Wavelet		325.1483	349.0970	2.7402

Tables 2-5 show that the performance of ARMA models under different data filtering methods (Outliers, Hampel, DB4 Wavelet, Symlets1 Wavelet, and Dmey Wavelet) has been systematically evaluated across a range of configurations (ARMA (1, 1), ARMA (2, 1), ARMA (1, 2), and ARMA (2, 2)) and varying sample sizes (200, 300, 400). AIC, BIC, and MSE are criteria used in time series modelling to assess the goodness of fit, model complexity, and prediction error, respectively.

1. Wavelet Analysis versus Outliers and Hampel method

The results across all configurations consistently demonstrate the superiority of wavelet-based filtering methods, particularly Dmey Wavelet, over both Outliers and Hampel methods in terms of predictive performance (MSE), as well as model complexity (AIC and BIC). This finding is particularly notable in high-dimensional time series data, where traditional filtering techniques like Hampel are often less effective at removing outliers while preserving relevant time series data characteristics.

2. Impact of Sample Size

Increased sample size (n = 400) consistently leads to a reduction in MSE for all methods, which is typical in statistical modelling. As more data is available, the model can better capture the underlying time series dynamics, improving generalizability. This effect is particularly evident with the wavelet analysis, where the decrease in MSE is sharper, further highlighting their effectiveness in handling larger datasets. This trend is not as pronounced with the traditional methods, where performance stagnates at higher sample sizes, likely due to their less adaptive nature than wavelet analysis techniques.

3. Model Configuration (ARMA (1, 1), ARMA (2, 1), ARMA (1, 2), ARMA (2, 2))

The performance of wavelet methods is notably consistent across different ARMA configurations, although slight differences emerge depending on the complexity of the model. ARMA (1, 1), this simplest configuration benefits from the filtering methods, especially Dmey Wavelet, which achieves the lowest MSE across all sample sizes. This suggests that even basic ARMA models can benefit from dealing with outliers provided by wavelet shrinkage. The ARMA (2, 1) model, which introduces an additional autoregressive term, further emphasizes the importance of wavelet filtering. The results indicate that wavelet methods (Dmey, Symlets, and DB4) excel in capturing the additional autoregressive structure compared to the traditional methods. With an additional moving average component, the ARMA (1, 2) model still shows strong results with wavelet methods. However, there is a slight dip in performance for Dmey Wavelet compared to simpler ARMA configurations, possibly due to the increased complexity of the model, which may require additional tuning or a more nuanced decomposition of the data. The most complex model in the study, ARMA (2, 2), benefits most from the wavelet methods, with Dmey Wavelet showing the best results. This is particularly striking given the relatively large MSE for traditional methods like Outliers and Hampel. As the complexity of the model increases, the need for precise outlier reduction becomes more pronounced, explaining the higher efficacy of wavelet methods in this case.

4. Model Selection Criteria (AIC and BIC)

AIC and BIC serve as crucial model selection tools in time series analysis. Both criteria balance goodness of fit with model complexity, penalizing overly complex models that do not provide sufficient improvement in fit.

In this study, Dmey Wavelet again shows the lowest values for both AIC and BIC across most sample sizes, particularly in the more complex models (ARMA (2, 1) and ARMA (2, 2)). This suggests that wavelet methods not only improve predictive accuracy (as seen in the MSE values) but also result in more parsimonious models, which is essential in avoiding overfitting. The superiority of Dmey Wavelet in AIC and BIC implies that it provides a better balance between fitness and complexity compared to the traditional methods.

8. Real Data

Climatic data were obtained from the Ministry of Agriculture, the agricultural meteorological center of Nineveh Governorate, and the Mosul station located at Longitude E 43.16 and latitude 36.33. These data were collected with the help of the Remote Sensing Center at the University of Mosul, which contributed significantly to the provision of data for the period between 2013-2022. The data were real and monitored by the meteorological station and represented average wind speed; the average wind speed is measured in m/s, meaning meter/second.

The wind speed time series data were tested for stationarity and were not stationary on the mean, so the second difference was taken. To balance overfitting and underfitting (By testing the significance of the estimated parameters of the different models) and to minimize AIC and BIC, the ARIMA (1, 2, 2) model(Autoregressive Integrated Moving Average) was chosen as the best fit for the data.

Figures (4-6) show that True ARIMA Series (Black Line): This is the original, ARIMA (1, 2, 2) time series. It represents the "true" underlying process with outliers. In the plot, it's shown as a black line. This serves as the baseline for comparing all other time series. Visually identify these outliers by looking for the large deviations from the rest of the series at these points. Hampel filtered series (Blue Line): After applying the Hampel filter to the noisy series, this line shows the result of removing the outliers. The Hampel filter smooths the noisy time series by replacing the outliers with median values within a moving window. The blue line represents this filtered series. See that the sharp spikes from the outliers are reduced or removed, providing a cleaner series, closer to the true ARIMA process. Wavelet filtered Series (Green Line for DB4, Symlets,1 and Dmey): This line represents the wavelet denoised time series. The green line shows the wavelet-filtered series, which should also smooth out the outliers while preserving the underlying structure of the time series.

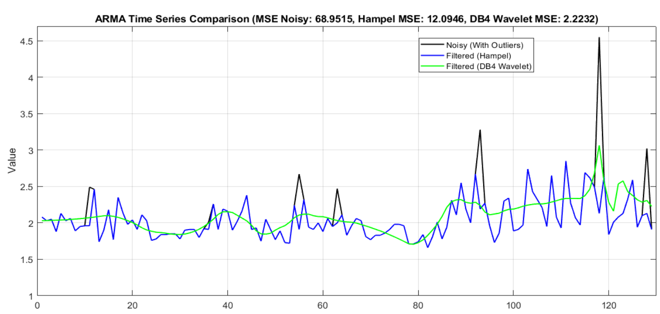

Figure 4. ARMA (1, 2, 2) Models for Wind Speed Time Series Data (DB4)

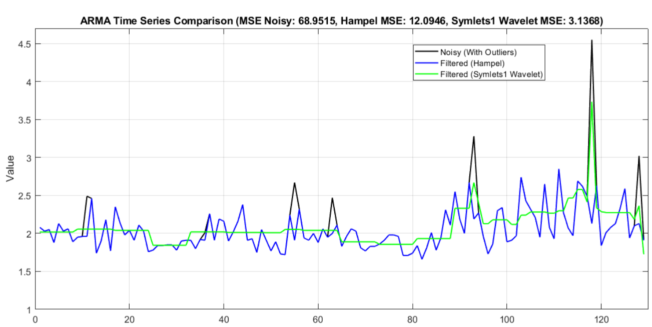

Figure 5. ARMA (1, 2) Models for Wind Speed Time Series Data (Symlets1)

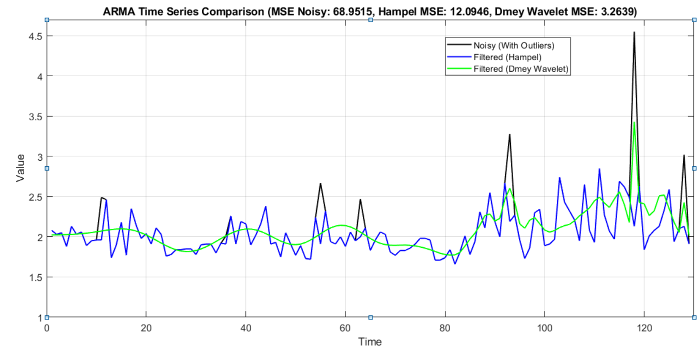

Figure 6. ARMA (1, 2) Models for Wind Speed Time Series Data (Dmey)

The title of the plot includes the MSE for the three models: The MSE for the ARIMA model is estimated from the noisy data (with outliers) equal to (68.9515). The MSE for the ARIMA model was estimated from the Hampel-filtered data equal to (12.0946). The MSE for the ARIMA model was estimated from wavelet-filtered data (DB4, Symlets1, and Dmey) equal to (2.2232, 3.1368, and 3.2639) respectively. These results suggest that wavelet denoising is a more robust method for handling noisy data and outliers in time series modelling, particularly when the goal is to estimate ARIMA parameters accurately and DB4 Wavelet was the best.

Table 6 presents the results of an estimation process for the parameters of two-time series models. The methods employed are Hampel (classical method) and DB4 Wavelet (best-proposed method), which are used to estimate coefficients in an ARIMA Model. The coefficients estimated include the constant, AR {1}, MA {1} and MA {2}. The statistical significance of these coefficients is assessed using the t-statistics and p-values, which offer insights into the model's fitness and the relevance of individual terms.

Hampel Method: The constant term has an estimated value of 0.1256, with a t-statistic of 1.2064 and a p-value of 0.228. This p-value exceeds the conventional significance threshold of 0.05, indicating that the constant term is not statistically significant at the (0.05) level. This suggests that, for the Hampel method, the constant term does not provide a significant contribution to the model's predictive power. DB4 Wavelet Method: The constant term in the DB4 Wavelet method is 0.6913, with a t-statistic of 5.5610 and a p-value of 0.000. The extremely low p-value indicates that the constant term is highly statistically significant, meaning it plays an essential role in the model's formulation.

Hampel Method: The AR {1} term has a value of 0.9386, with a t-statistic of 18.524 and a p-value of 0.000. The low p-value suggests that the AR {1} term is statistically significant. The coefficient indicates that the previous point has a strong positive relationship with the current value of the series, with a high degree of persistence. DB4 Wavelet Method: The AR {1} coefficient is 0.6683, with a t-statistic of 13.495 and a p-value of 0.000, also suggesting that the AR {1} term is highly significant.

Hampel Method: The MA {1} term is estimated as -0.7562, with a t-statistic of -6.7006 and a p-value of 0.000. The negative coefficient, combined with the very low p-value, suggests that there is a significant negative correlation between the residuals at lag 1. This means that shocks or errors in the previous time step tend to have an inverse effect on the current value. DB4 Wavelet Method: The MA {1} coefficient is 0.8138, with a t-statistic of 16.647 and a p-value of 0.000, indicating a statistically significant positive correlation between residuals at lag 1. The positive sign implies that the residual error from the previous period has a reinforcing effect on the current value, as opposed to the negative relationship in the Hampel method.

Hampel Method: The MA {2} term has an estimated value of -0.0238, with a t-statistic of -0.2764 and a p-value of 0.782. The p-value being well above 0.05 indicates that the MA {2} term is not statistically significant. This suggests that the second lag of residuals does not meaningfully contribute to explaining the time series data when using the Hampel method. DB4 Wavelet Method: The MA {2} term is 0.5740, with a t-statistic of 14.100 and a p-value of 0.000. In contrast to the Hampel method, the DB4 Wavelet method finds the MA {2} term to be statistically significant. The positive coefficient indicates that there is a significant effect from the second lag of residual errors, implying that the model benefits from capturing additional error structure at this lag.

Table 6. Testing the significance of the estimated parameters

Coefficient	Method	Value	Standard Error	t-statistics	p-value
Constant	Hampel	0.1256	0.1041	1.2064	0.228
AR {1}		0.9386	0.0507	18.524	0.000
MA {1}		-0.7562	0.1129	-6.7006	0.000
MA {2)		-0.0238	0.0860	-0.2764	0.782
Constant	DB4 Wavelet	0.6913	0.1243	5.5610	0.000
AR {1}		0.6683	0.0495	13.495	0.000
MA {1}		0.8138	0.0489	16.647	0.000
MA {2)		0.5740	0.0407	14.100	0.000

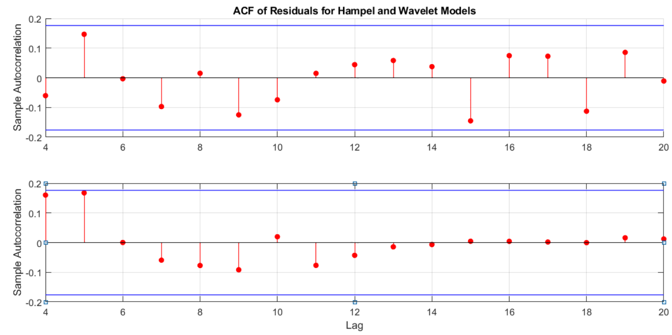

Figure 7. ACF of Residuals for Hampel and DB4 Wavelet Models

The ACF plot in **Figure 7**, all values close to zero, with most of the bars falling within confidence intervals. This would indicate that the model has successfully captured the underlying time series structure.

Conclusion

This analysis underscores the importance of advanced filtering techniques like wavelets in time series modelling. Across various ARMA model configurations, Dmey Wavelet consistently emerges as the most effective method to remove outliers, leading to lower MSE, AIC, and BIC values. The results suggest that wavelet methods, particularly Dmey Wavelet, should be strongly considered for future time series forecasting tasks, especially those involving outliers. For wind speed time series data, both methods (Hampel and proposed method) exhibit strong parameter significance, the DB4 Wavelet method appears to be the better-performing model in terms of both statistical significance and prediction accuracy, making it a more reliable choice for forecasting time series data.

Acknowledgement

The authors would express they're thanks to the College of Computer Science and Mathematics, University of Mosul to support this report.

Conflict of interest

None.

References

[1] Ali, T. H., Raza, M. S., & Abdulqader, Q. M. (2024). VAR TIME SERIES ANALYSIS USING WAVELET SHRINKAGE WITH APPLICATION. Science Journal of University of Zakho, 12(3), 345–355. https://doi.org/10.25271/sjuoz.2024.12.3.1304

[2] Ali, T. H., Saleh, D., Mustafa Abdulqader, Q., & Omer Ahmed, A. (2025). Comparing Methods for Estimating Gamma Distribution Parameters with Outliers Observation. Journal of Economics and Administrative Sciences, 31(145), 163-174 https://doi.org/10.33095/cc5b 9h49

[3] Ali, Taha Hussein & Qais Mustafa. (2016)"Reducing the orders of mixed model (ARMA) before and after the wavelet de-noising with application." Journal of Humanity Sciences 20.6: 433-442 https://web.archive.org/web/20200604023557/http://zancojournals.su.edu.krd/index.php/AHS/article/download/1230/702

[4] Ali, Taha Hussein, Mohammed Abdul Majeed Badal, & Safaa S. (2010)"Estimations of AR(p) Model using Wave Shrink" Iraqi Journal of Statistical Sciences, 17: 97-114.

[5] Ali, Taha Hussein, Nasradeen Haj Salih Albarwari, and Diyar Lazgeen Ramadhan. (2023) "Using the hybrid proposed method for Quantile Regression and Multivariate Wavelet in estimating the linear model parameters." Iraqi Journal of Statistical Sciences 20.1: 9-24.

[6] Bhar, L. & Gupta, V., (2001) "A Useful Statistic for Studying Outliers in Experimental Design", The Indian Journal of Statistic, Vol.63, Series B pt.3, pp338-350.

[7] Box, G., Jenkins, G., Reinsel, G. and Ljung G., (2016)," Time Series Analysis Forecasting and control", John Wiley & Sons, Inc. Hoboken, New Jersey.

[8] Cryer, J. D. and Chan, K. S. (2008)." Time Series Analysis with Application in R".2nd ed., Springer Science Business Media.

[9] Duaa Faiz Abdullah Faiz Abdullah; Jwana Rostom Qadir; Diyar Lazgeen Ramadhan; Taha Hussein Ali. (2024) "CUSUM Control Chart for Symlets Wavelet to Monitor Production Process Quality.". IRAQI JOURNAL OF STATISTICAL SCIENCES, 21, 2, , 54-63. doi:10.33899/iqjoss.2024.185240

[10] Elias, I. I., & Ali, T. (2025), Optimal Level and Order of the Coiflets Wavelet in the VAR Time Series De-noise Analysis. Frontiers in Applied Mathematics and Statistics, 11, 1526540. https://doi.org/10.3389/fams.2025.1526540

[11] Friedman, J., (1999)," Multivariate Adaptive Regression Spline", The Annals of Statistics, Vol.19, No.1, pp1-141.

[12] Hample, F., Ronchetti, E., Roussee, U. and Stahel, W. (1986)," Robust Statistics the Approach Based on Influence", John Wiley & Sons, New York.

[13] Haydier, Esraa Awni, Nasradeen Haj Salih Albarwari, and Ali, Taha Hussein. (2023) "The Comparison Between VAR and ARIMAX Time Series Models in Forecasting." Iraqi Journal of Statistical Sciences 20.2: 249-262.

[14] Heyam Hayawi, Muzahem Al-Hashimi, Mohammed Alawjar.(2025)," Machine learning methods for modelling and predicting dust storms in Iraq", STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING, Vol.13, pp 1063–1075.

[15] DOI:10.19139/soic-2310-5070-2122 http://www.iapress.org/index.php/soic/article/view/2122/1163

[16] Heyam A. Hayawi, Najlaa Saad Ibrahim, Omer Salem Ibrahim, (2021)," Forecasting The Fuzzy Hybrid Arima-Garch Model Of Stock Prices In Iraqistock Exchange", Int. J. Agri cult. Stat. Sci. 17 (1), 2229-2238.

[17] Heyam A. Hayawi, Shakar Maghdid Azeez, Sawen Othman Babakr and Taha Hussein Ali ,(2025)," ARX Time Series Model Analysis with Wavelets Shrinkage (Simulation Study)", Pak. J. Statist. Vol. 41(2), 103-116 https://www.pakjs.com/wp-content/uploads/2025/04/PJS-41201.pdf

[18] Kadir, D.H., Yahya, R.O., Khudhur, A.M. and Ali, T.H., (2024). Data De-Noise for Multivariate T2 and S-Charts Using Multivariate Wavelets. Zanco Journal of Human Sciences, 28(6), pp.343-360. https://doi.org/10.21271/zjhs.28.6.19

[19] Keller, G. & Brian W., (2000)," Statistic for Management and Economics", Edition, Duxbury, Thomson Learning U.S.A.
Liu, L. M., (2006)." Time Series Analysis and Forecasting", 2nd ed., Scientific Computing Associates Crop., Illinois, USA.

[20] Muzahem M. Al-Hashimi, Heyam A. Hayawi, (2023)," Nonlinear Model for Precipitation Forecasting in Northern Iraq using Machine Learning Algorithms", International Journal of Mathematics and Computer Science 19 (1),171-179.

[21] Muzahem M. Al-Hashimi, Heyam A. Hayawi, Mowafaq Al-Kassab, (2023) "A Comparative Study of Traditional Methods and Hybridization for Predicting Non-Stationary Sunspot Time Series", International Journal of Mathematics and Computer Science 19 (1), 195–203.

[22] Najlaa Saad Ibrahim, Heyam A. Hayawi, (2021)," Employment the State Space and Kalman Filter Using ARMA models", International Journal on Advanced Science Engineering Information Technology, Vol.11,(1),145-149.

[23] Omer, A. W., Sedeeq, B. S., & Ali, T. H. (2024). A proposed hybrid method for Multivariate Linear Regression Model and Multivariate Wavelets (Simulation study). Polytechnic Journal of Humanities and Social Sciences, 5(1), 112-124. https://doi.org/10.25156/ptjhss.v5n1y2024.pp112-124

[24] Sakar Ali Jalal; Dlshad Mahmood Saleh; Bekhal Samad Sedeeq; Taha Hussein Ali. (2024) "Construction of the Daubechies Wavelet Chart for Quality Control of the Single Value". IRAQI JOURNAL OF STATISTICAL SCIENCES,21,1,160-169 Doi:10.33899 iqjoss.2024.183257