



Using Wavelets to Identify Linear Dynamic Models

Youns M. Th. Al.Obeady¹ , Heyam A.A.Hayawi² , Mohamed Ahmed Elkhoul³ 

^{1,2}Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

³ Sadat Academy for Management Science (SAMS), Department of Mathematics and Statistics, Cairo, Egypt.

Article information

Article history:

Received: March 1, 2024

Revised: December 20, 2024

Accepted: December 25, 2024

Available online June 1, 2025

Keywords: Time Series, Dynamic System, Identification, Wavelet

Correspondence:

Heyam A.A.Hayawi

he.hayawi@uomosul.edu.iq

Abstract

The process of forecasting water purification in the city of Mosul was investigated by analyzing the input and output variables, which were derived from several tests conducted on raw water prior to the filtration process. This water undergoes treatment through multiple filtration stages. In order to assess the safety of the water for human consumption, the filtration process was examined using forecasting techniques with dynamic models. These included the autoregressive model with additional inputs, the moving averages model, autoregression with additional inputs, the output error model, and the Box-Jenkins model. The most effective model, determined from the data, was identified using statistical criteria. Subsequently, a comparison was made to evaluate the predictive accuracy based on forecasting criteria, and the model was applied to water quality data. The study also emphasized the importance of selecting the most appropriate forecasting model to ensure accurate predictions of water quality throughout the filtration process. By applying different dynamic models, it was possible to compare the performance of each model in terms of its ability to predict water safety for human consumption. The results demonstrated that certain models, such as the autoregressive model with additional inputs, showed greater precision in forecasting water quality outcomes. This highlights the crucial role of advanced statistical methods in improving water treatment processes. Moreover, the research emphasized the need for continuous monitoring and adaptation of forecasting models to account for seasonal variations and potential changes in water sources.

DOI: [10.33899/ijoss.2025.187731](https://doi.org/10.33899/ijoss.2025.187731) , ©Authors, 2025, College of Computer Science and Mathematical, University of Mosul.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Building any dynamic system involves creating a mathematical model from a set of input and output data for the dynamic system under study. This is achieved through a mathematical representation of the correlation between a group of variables. The process of constructing a dynamic system involves describing the output based on preceding outputs and inputs, as well as other exogenous variables, which can be defined through obfuscation and model diagnosis in various ways. The diagnostic approach relies on the nature of the system and the specific goals of the diagnosis. In many cases, researchers have utilized stochastic linear motor systems diagnosis methods, incorporating modern techniques such as fuzzy logic and neural networks, in addition to traditional methods, to enhance the diagnostic process. These

advanced techniques have enabled more accurate modeling and prediction of system behavior, allowing for improved system performance, better troubleshooting, and optimized control strategies. Studies in this area have made significant contributions by demonstrating how these techniques can provide more detailed insights into system dynamics and improve decision-making processes in dynamic system management. The application of dynamic system modeling is crucial in various fields, including engineering, environmental science, and economics, as it provides a structured approach to understanding complex systems. In dynamic systems, the relationship between inputs, outputs, and intermediate states is often nonlinear, which makes the modeling process more challenging but also more insightful. One of the key advantages of dynamic models is their ability to predict future states based on historical data, which is essential for making informed decisions in real-time. For instance, in water purification systems, dynamic models can help forecast the impact of various filtration stages, allowing for better control over the quality of treated water. Similarly, in industrial applications, dynamic system models can predict machine performance, enabling maintenance schedules to be optimized and downtime minimized. In addition to the traditional methods used in dynamic system modeling, the integration of modern techniques such as machine learning, deep learning, and artificial intelligence has significantly enhanced the diagnostic process. These technologies allow for the processing of large volumes of data, which is often a limitation of conventional methods. Machine learning models, for example, can be trained to recognize patterns in data that are not immediately apparent, thus providing more accurate predictions and diagnostics. Neural networks, in particular, are adept at handling highly complex relationships between input and output variables, making them valuable tools for dynamic system modeling. Furthermore, fuzzy logic has become increasingly popular in system diagnosis because it can handle the uncertainty and imprecision that are inherent in many real-world systems. In many cases, the variables involved in dynamic systems are not perfectly defined, and fuzzy logic allows for reasoning in situations where data may be vague or incomplete. This makes it especially useful in systems where human judgment or subjective data is involved, such as in decision support systems or expert systems. Despite these advancements, the development and implementation of dynamic models still face challenges, especially in terms of computational complexity and data quality. Accurately modeling a dynamic system requires high-quality, reliable data, which is not always available. Additionally, the computational resources required to process and analyze large datasets can be significant, especially when using advanced techniques like deep learning. Therefore, ongoing research and development are necessary to refine existing methods and explore new approaches to dynamic system modeling. In conclusion, dynamic system modeling plays a vital role in various domains by providing predictive insights that help improve system performance and decision-making. The integration of modern computational techniques has revolutionized the field, making it possible to model more complex systems with greater accuracy. However, challenges such as data quality, computational demands, and model interpretability remain, and addressing these issues will be key to further advancements in the field. As the technology continues to evolve, dynamic system modeling will undoubtedly play an even more central role in optimizing systems and solving complex real-world problems.

2.Dynamic Systems

The construction of the Linear regression model Structure from which most linear models are derived and from which the outputs of the designated linear system are calculated as follows:[1] u_t Through filtering inputs at the time t Y_t

$$y_t = G(q)u_t + H(q)v_t \quad (1)$$

where $G(q)$ and $H(q)$ are linear filters They can be expressed in terms of polynomials system identification and the input-output from:[8,2]

$$Y_t = \frac{B(q)}{F(q)A(q)}u_t + \frac{C(q)}{D(q)A(q)}v_t \quad (2)$$

$$A(q) = 1 + a_1q^{-1} + a_2q^{-2} + \dots + a_{na}q^{-na}$$

$$B(q) = b_0 + b_1q^{-1} + b_2q^{-2} + \dots + b_{nb}q^{-nb}$$

$$C(q) = 1 + c_1q^{-1} + c_2q^{-2} + \dots + c_{nc}q^{-nc}$$

$$D(q) = 1 + d_1q^{-1} + d_2q^{-2} + \dots + d_{nd}q^{-nd}$$

$$F(q) = 1 + f_1q^{-1} + f_2q^{-2} + \dots + f_{nf}q^{-nf}$$

The linear filter ($G(q)$) is known as the input transfer function, which connects the input to the output. In contrast, the linear filter ($H(q)$) is referred to as the noise transfer function, relating the noise ($v(t)$) to the output. The stochastic linear dynamic systems models are divided into two parts.:

3. The Equation Error Model

They include ARX and ARMAX models and feature a common linear filter $1/A(q)$ in both the determinant operations and the stochastic operations models, in other words, all *ARX and ARMAX* models share a polynomial $A(q)$ as a dynamic denominator of the input transform function and the noise transformation function. This is consistent with The noise does not directly affect the model's output but instead enters the model before the filter($1/A(q)$). In other words, interference enters the process later, causing the features of repetition to be shaped by the dynamics of the process. [1,7].

- **"Autoregressive with exogenous input model"(ARX)**

The ARX model is a widely used linear dynamic model, suitable for many real-world applications. Its practicality stems from the ease of calculating its parameters using the least squares method. This can be illustrated by the following equation: [2,11]

$$y_t = \frac{B(q)}{A(q)} u_t + \frac{1}{A(q)} v_t \quad (3)$$

The optimal prediction of this model can be calculated from the subsequent equation:

$$\hat{y}_t(t|t-1) = B(q)u_t + (1 - A(q))y_t \quad (4)$$

- **"Autoregressive Moving Average with Exogenous input model"(ARMAX)**

The ARMAX model is the second simplified linear model following the ARX model. It offers some dominant designs with reduced variance by leveraging information from the interference model, in contrast to the ARX model. It is an adaptable model and because it has an expanded noise model, in old research and publications the ARX model is expressed through the ARMAX model, due to the presence of a polynomial for the numerator and denominator, and in general, the terms adopted by Ljung that the ARMAX model is a time series. This model can be represented by the following equation: [1,8]

$$y_t = \frac{B(q)}{A(q)} u_t + \frac{C(q)}{A(q)} v_t \quad (5)$$

The optimal prediction of this model can be calculated from the subsequent equation:

$$\hat{y}(t|t-1) = \frac{B(q)}{C(q)} u_t + \left(1 - \frac{A(q)}{C(q)}\right) y_t \quad (6)$$

4. The Output Error Models

It includes both the Output Error (OE) model and the Box-Jenkins (BJ) model. Output Error models are distinct because their noise models do not include a dynamic process. In other words, Output Error models are characterized by the independence of the noise model from the specific process model. [8].

5. Wavelet Analysis

The research aims to use the wavelet method to diagnose stochastic linear kinetic systems using the wavelet method and compare the prediction in this method with the traditional method using real data [9] and [3].

6. Criteria to Choose Best Model

There are many statistical and engineering criteria that are often used in diagnosing stochastic linear dynamic models, including:

a) Cost Function

It is sometimes referred to as the Lost Function, where it is the best basic necessity in Selecting the model's rank by examining how this function behaves as it increases in the order of the model, as its value decreases with the increase in the rank of the model and that the decrease in the value of this function stops at a certain point, which means that the increase in the order of the model becomes useless [1] and [2]. This function can be mathematically calculated as follows:

$$V = \frac{1}{2} \sum_{t=1}^N e_t^2 \quad (7)$$

b) Akaike's Final Prediction Error Criterion

One of the important criteria in determining the appropriate rank of the model, which was known by the scientist Akaike in 1969 and represents the measure of the final prediction error and is defined as The variance of the prediction error for a future period is calculated as follows: [2] and [6]

$$FPE = \frac{1 + \frac{m}{N}}{1 - \frac{m}{N}} \quad (8)$$

c) Akaike's Information Criterion

This criterion was likewise known by the scientist Akaike in the year (1973-1974) and is abbreviated as he provided information for selecting the appropriate order of the ARIMA model from among several models so that the appropriate rank corresponds to the lowest value of the AIC criterion and depicts the most suitable order. It is mathematically expressed as follows: [10] and [3].

$$AIC = 2n - 2 \ln L \quad (9)$$

d) Fitting Criteria

It represents a measure to know the accuracy of the model as a percentage, where (Ljung, 2004) relied in determining the appropriate rank of dynamic models on dividing the input and output observations used into two groups after merging them into an object. The first group represents the input-output observations that are used in the estimation process as known as estimation data which is used to obtain a set of estimated models, while the second group represents the other part of the input-output observations that are used to test the legitimacy of the models that were estimated from the first group, and it is known as the Validation Data, where the amount of congruence is calculated as a percentage and this percentage is calculated as follows [5] and [4].

$$Fit = 100 * \left(1 - \sqrt{\frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{\sum_{t=1}^N (y_t - \bar{y})^2}} \right) \quad (10)$$

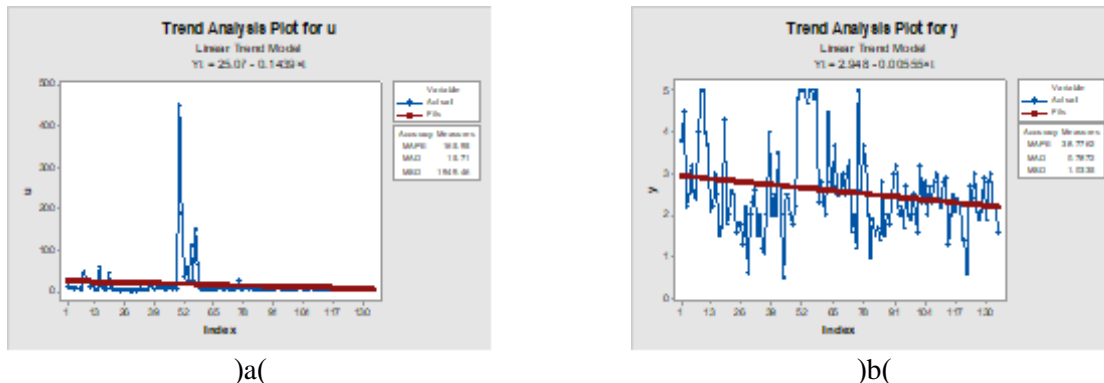
7. Methods for Estimation Stochastic Linear Dynamic Systems

There are many methods that are used in diagnosing stochastic linear dynamic systems, including [3]: least squares method, greatest possibility method, and, prediction error method.

8. Application Side

One of the first steps that can be followed to describe the time series, is use the time-graph of the data. Through the drawing, note that, the nature of the fluctuations. whether the series is stable or velocity or not in terms of mean and variance, as the stability of the series has a great role in the modeling process of time series and dynamic systems as well as their relationship to forecasting.

The data used in this paper, which includes the turbidity in the water before and after filtering, as the turbidity data before filtering was represented by the inputs, which are symbolized by u and the turbidity after filtering, symbolized by y , It is clear from the diagram that the stability of the two series is as follows:



Figures (1): analysis of the general trend of the time series of inputs and outputs of turbidity before and after water purification,

(a): represents the input (turbidity before filtering), (b): represents the output (turbidity after filtering)

It is clear from Figure (1) above that both series, the input for turbidity before filtering, and the output for turbidity after filtering are unstable in the mean and variance, also the instability was confirmed by dividing the series into parts and then finding the variance for each parts, We noticed that the variance values for all parts were unstable. The logarithmic transformation was performed for the purpose of fixing the variance, and also the first difference of the two series was taken to remove the effect of the general trends in the two series.

The diagnostic process in stochastic linear dynamic systems is obtained by reconciling many kinetic models with different and multiple parameters, according to the model used, which is done by dividing the data represented by the inputs, which represents the turbidity series of the water before filtering, and the output, which is the turbidity series of the water after filtering into two parts: The first part is used to estimate the parameters of the model. It is known as the Estimation object, Z_e , While the second part is used to test the fit of the model which is known as the Validation object, Z_v , Through the use of Matlab 2020 system, the delay time for stochastic linear dynamic systems was estimated through one of the estimation methods for the delay time, as follows view these for ways:

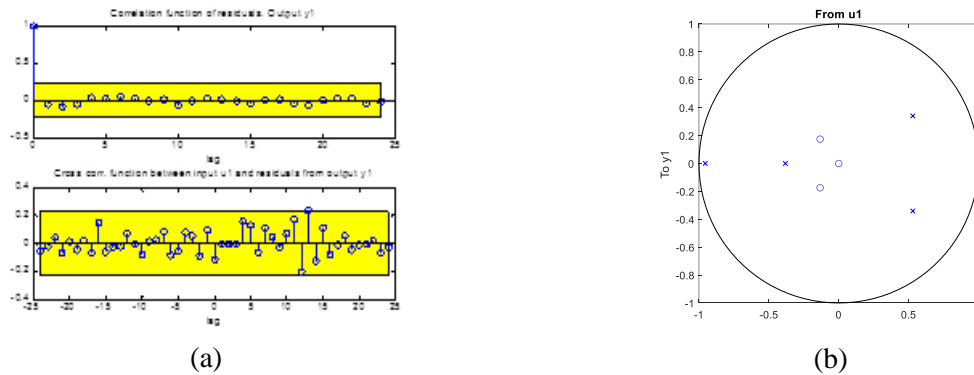
The delay time was estimated using the method proposed by Ljung by using the ARX model by fixing the parameter values of the model $na = nb = 2$. Take compensation for the delay time $nk = 1: 10$ the choice of the delay time for the model corresponding to the minimum value of the statistical criteria and was the value of $nk = 1$.

After that, adjustments are made for the equation error models, namely the autoregressive model with additional inputs ARX, the autoregressive model and moving averages with additional inputs ARMAX, and the best model is chosen by applying the statistical criteria as in the following table (1):

Table (1): represents the models of the final stochastic linear dynamic systems

Models	Aic	Fitt	FPe	Resid	Cross	Mse
ARX(4,3,1)	-1.724	100%	0.178	Random	Uncorrelated	0.150
ARMAX(4,3,3,1)	-1.857	100%	0.156	Random	Uncorrelated	0.133

Figure (2) shows the stability of the best model obtained



Figures (2): (a) shows the unit circuit of the best model using real data
(b): Plotting the cross-correlation and residuals of the best model using real data

We noticed in the figure that the best model obtained is completely unstable, as shown in Figure A, a drawing of the unit circle, while in the Figure B it has completely random horns.

When comparing the results obtained from diagnosing equation error models with its two parts, the autoregressive model with additional inputs ARX and the autoregressive model and moving averages with additional inputs ARMAX, we note that the ARMAX model corresponds to the lowest values of the statistical criteria that were obtained, This model is considered the best among the error equation models.

For the purpose of comparing the dynamic models with the wavelet. for the purpose, the data that represents the turbidity of the water before and after filtering will be processed, meaning that X represents the turbidity before filtering and Y represents the turbidity after filtering, which is about (135) views, through the use of the small wave Haar and finding the discontinuous wavelet transformation (DWT).) by relying on the ready-made software MATLAB (SOFTWARE) through the following:

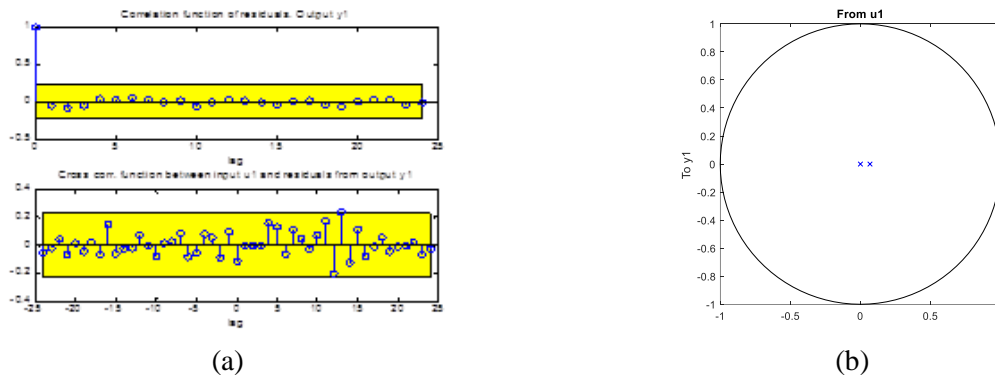
First, the original data is entered into the ready-made MATLAB program and the following instructions are used:

(Start → Toolboxes → More → Wavelet → Wavelet Toolbox Main Menu(Wave menu As the wavelet HAR was used on the data of the study, and by defining the five-levels wavelet HAR, the equation error models were diagnosed in the same method as before on the wavelet HAR data, and the results were shown in the following table:

Table (2): The models of the final stochastic linear dynamic systems using the wavelet har

Models	Aic	Fitt	FPe	Resid	Cross	Mse
ARX(1,2,1)	-2.836	100%	0.058	Random	Uncorrelated	0.054
ARMAX(1,2,2,1)	-2.983	100%	0.055	random	Uncorrelated	0.051

The best model obtained using the wavelet HAR can be seen from the following figure (3):



Figures (3): (a) The unit circle of the best model using the wavelet HAR
(b): Draw the cross-correlation and residuals of the best model using the wavelet har

We note from the figures that the best model obtained is completely stable and drawing of the unit circle through the occurrence of all poles inside the unit circle zeros were located outside the unit circle ,In figure B, the model has completely random residuals. Thus the best model has number of parameters is less than in the real data.

After getting the best model $ARMAX(na = 1, nb = 2, nc = 2, nk = 1)$ according to the statistical and engineering standards for real data and waveform data, and as shown in table (1) and (2), the predictive values for the best real data and waveform data are found by applying the predictive equation of this model, which was previously referred to in the theoretical side. The results were shown in table (3) as follows:

Table (3): The predictive values of the best kinematic model using real data and wavelet data

Series	Real values	Predictive value of real data	Predictive value of wavelet data
130	0.387766	0.254	0.098
131	0.068993	0.401	0.147
132	-0.31015	0.249	0.181
133	0.03523	0.189	0.202
134	-0.31845	0.392	0.216

Figures (3): A shows the unit circuit of the best model using the wavelet HAR
B: Draw the cross-correlation and residuals of the best model using the wavelet har

From table (3), we notice the superiority of the wavelet method in giving better prediction values than the classical method for predicting stochastic linear dynamic systems, and this was confirmed by the criteria for testing prediction accuracy, as the standards gave lower values when using the wavelet than in the real data.

9. Conclusions and Recommendations

The research reached some conclusions, including:

The turbidity data series for drinking water before and after the filtration process was unstable in the mean and variance and was converted into a stable time series by taking the square root transformation and the first difference.

The best model obtained using wavelet data is $ARMAX(na = 1, nb = 2, nc = 2, nk = 1)$ while the best model obtained for real data is $ARMAX(na = 4, nb = 3, nc = 3, nk = 1)$

We notice that the wavelet model has fewer parameters than it is from the real data, as the fewer parameters the model is, the better.

Adoption of the wavelet method to predict the linear dynamic systems, as it gave the wavelet prediction values that are closer to the real values than it is in the use of the usual method of forecasting.

When using the prediction accuracy criteria, these criteria gave lower values when using the wavelet method than the real data.

We recommend using the wavelet method to predict time series and other stochastic linear dynamical systems, as well as using other wavelets such as Dobby's wavelet to give better prediction values.

Acknowledgement

None

Funding

None

Conflict of Interest

The authors do not have any conflict of interest

References

1. Chiras, N.,(2002):"Linear and Nonlinear Modeling of Gas Turbine Engines",Ph.D. Thesis University of Glamorgan Lim-ssol Gyprus.
2. Hayawi, H.A.A., Alsharabi, N.S.I.,(2022):" Analysis of multivariate time series for some linear models by using multi-dimensional wavelet shrinkage" ,Periodicals of Engineering and Natural Sciences., Vol.10 ,No.4, pp. 120–129. DOI:<http://dx.doi.org/10.21533/pen.v10i4.3121>
3. Hayawi, H.A.A.,(2022):"Using Wavelet in identification state space models", International Journal of Nonlinear Analysis and Application, Vol.13,Pp. 2573-2578. <http://dx.doi.org/10.22075/ijnaa.2022.5958> .
4. Hayawi,H.A.A.,(2020):"Employ the Principle Components in the Detection of feedback",journal of physics: conference series., vol.1591. <http://pen.ius.edu.ba/index.php/pen/article/view/3121>. [doi:10.1088/1742-6596/1591/1/012100](https://doi.org/10.1088/1742-6596/1591/1/012100)
5. Heyam A. Hayawi , Ibrahim, Najlaa Saad,(2022),"Comparison of prediction using Matching Pattern and state space models",IRAQI JOURNAL OF STATISTICAL SCIENCES 19 (1), 30-37. Doi: [10.33899/ijoss.2022.0174329](https://doi.org/10.33899/ijoss.2022.0174329)
6. Heyam A. Hayawi , Ibrahim, Najlaa Saad ,Mohammed, Lamyaa Jasim,(2021),"Using the fuzzy technique to identification stochastic linear dynamic systems",Journal of Statistics and Managment Systems 24 (4), 801-808.<https://doi.org/10.1080/09720510.2020.1859808>
7. 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. <https://future-in-tech.net/19.1/R-Al-Hashimi.pdf>
8. 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. <https://future-in-tech.net/19.1/R-MuzahemAl-Hashimi.pdf>
9. Fahad S. Subhy, Heyam A. Hayawi,(2021),"Comparison Prediction of Transfer Function Models and State Space Models Using Fuzzy Method",Iraqi Journal Of Statistical Sciences 18 (2), 73-81. DOI: [10.33899/ijoss.2021.0169968](https://doi.org/10.33899/ijoss.2021.0169968)
10. 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. DOI:[10.19139/soic-2310-5070-2122](https://doi.org/10.19139/soic-2310-5070-2122)
11. 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>

12. Shahla H. Ali, Heyam A. Hayawi, Nazeera Sedeek, Taha H. Ali,(2023), "Predicting the Consumer price index and inflation average for the Kurdistan Region of Iraq using a dynamic model of neural networks with time series ",The 7th International Conference of Union if Arab Statistician-Cairo.
13. 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. DOI: [10.18517/ijaseit.11.1.14094](https://doi.org/10.18517/ijaseit.11.1.14094)
14. Kahane, J.P, P.G. Le Marie, "Fourier Series and Wavelets" Gordon and Research puplishers,1995.
15. Ljung, L., & Söderström, T.,(1983): "Theory and Practice of Recursive Identification", IT Press Cambridge Massachusetts London England.
16. Ljung, L., (2002): "System Identification Toolbox for use with MATLAB", 5.0 Math works Inc.
17. Ljung, L., (2004): "System Identification Toolbox for use with MATLAB", 6.0 Math works Inc.
18. Mallat, S. (1989), "A theory for multi resolution signal decomposition: the wavelet representation," IEEE Pattern Anal. and Machine Intell., vol. 11, no. 7, pp. 674-693.
19. Nelles, O.,(2001): "Nonlinear System Identification from Classical Approach to Neural Network and Fuzzy Models", Springer Verlag Belin Heidelberg Germany.
20. Wang,J.,Pai,P& Lin, Y.,Grey ,(2013): 'models in seasonal time series forecasting', Journal of Statistics and Management Systems, P.469-476.
21. Wei, W.W.S., 2006, "Time Series Analysis: Univariate and Multivariate Methods", Addison-Wesley Publishing Company, U.S.A.

توظيف الموجات لتشخيص النماذج الحركية الخطية

يونس محمد ذنون¹ ، هيام عبد المجيد حياوي²، محمد احمد الخولي³

^{1,2}، قسم الإحصاء والمعلوماتية، كلية علوم الحاسوب والرياضيات، جامعة الموصل ،موصل ، العراق

³ أكاديمية السادات للعلوم الإدارية (SAMS)، قسم الرياضيات والإحصاء، القاهرة، مصر.

الخلاصة: تم دراسة عملية التنبؤ لتنظيف المياه في مدينة الموصل باستخدام متغيرات المدخلات والمخرجات التي تمثلها بعض الاختبارات التي يتم إجراؤها على المياه الخام قبل عملية الترشيح ، ليتم معالجتها من خلال مراحل الترشيح المتعددة .لتحديد سلامة المياه للاستهلاك البشري ، تمت دراسة عملية الترشيح من خلال التنبؤ باستخدام النماذج الديناميكية ومن ضمنها نموذج الانحدار الذاتي مع مدخلات إضافية ونموذج المتوسطات المتحركة والانحدار الذاتي مع مدخلات إضافية فضلا عن نموذج خطأ المخرجات ونموذج بوكس جنكينز وتم تشخيص افضل نموذج تم الحصول عليه من البيانات باستخدام المعايير الاحصائية، ومن ثم إجراء مقارنة للتنبؤ من خلال معايير التنبؤ وبالتطبيق على بيانات المياه.

الكلمات المفتاحية: السلسلة الزمنية ، النظام الديناميكي ، التشخيص ، الموجات