Al-Nahrain Journal of Science
ANJS, Vol.28 (4), December, 2025, pp. 259-269

ARIMA-NN Model for Drugs Sales Forecasting in the United States

Ghadeer Jasim Mohammed Mahdi[1,*], Zahraa Ibrahim Al-Share[2]

[1] Department of Mathematics, College of Science, University of Baghdad, Baghdad, Iraq.
[2] Directorate General of Education Karkh 2, Ministry of Education, Baghdad, Iraq,

Article's Information

Received: 28.02.2025
Accepted: 10.08.2025
Published: 15.12.2025

Keywords:
ARMA,
ARIMA,
ANNs,
Forecasting,
Time Series,
Multi-layer Perceptron.

Abstract/ Overview

This study proposes a new version of the Autoregressive Integrated Moving Average (ARIMA) model using Artificial Neural Networks (ANNs) denoted by ARIMA-NN. The new model incorporates a multi-layer perceptron with matrix multiplication within a feed-forward network. The logistic, hyperbolic tangent (tanh), and sigmoid activation functions are used for weight updates in ARIMA-NN. A new forecasting algorithm is proposed, and one-step and multiple-steps forecasting procedures are rigorously analyzed. The proposed model was evaluated against existing forecasting model using performance metrics such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to assess its effectiveness. The U.S. Census Bureau (www.census.gov) provides a dataset of monthly drug sales spanning ten years (2014-2024), which is utilized in the study. The ARIMA-NN model is applied to generate forecasts for drug sales in the U.S. for the next four years to demonstrate the models' utility and efficacy. All the computations and visualizations are performed using various R packages in version 4.3.2.

http://doi.org/10.22401/ANJS.28.4.19
*Corresponding author: ghadeer.j@sc.uobaghdad.edu.iq

1. Introduction

Techniques applied to the predictor parameters are utilized in the ARMA and ARIMA models. The models use data from the variables to predict the trend, and they require an understanding of the stochastic and probabilistic features of the variable. The Box-Jenkins (1976) approach is the term used to refer to these models [1]. Before fitting the model, two assumptions should be made. The model is stationary, to begin with. It is considered stationary if a series has a finite variance, a declining autocorrelation over time, and mean reversion. The second prerequisite is the invertibility assumption. Invertibility is necessary for the series to be represented by a finite order [2]. The autoregressive (AR) and moving average (MA) models, of which the ARMA model is a part, each explain a series' performance from different objectives. The dependence between an observation and the residual error from the moving average model is combined by MA. Returning to the stationary topic, the integration component distinguishes ARMA from ARIMA [3]. The three terms that make up the ARIMA model are AR, MA, and integration (I). To maintain stationarity, raw data points are differentiated. As stated differently, a data point will be substituted for the discrepancies between two values. Since most variables are non-stationary, they must undergo a transformation process known as differencing or integration to transfer the time series to stationary series [4]. The ARIMA model is typically expressed as ARIMA(p, d, q)(P, D, Q), where:

- p represents the number of lag observations in the autoregressive (AR) component.
- d is the number of non-seasonal Lag-1 differences required to achieve stationarity.
- q denotes the size of the moving average (MA) window.
- P is the number of seasonal Lag-h differences, used for capturing seasonal autoregressive effects.
- D is the order of seasonal differencing (also based on Lag-h).
- Q is the order of the seasonal moving average.

This detailed structure enables the model to handle both regular and seasonal components of the time series effectively. The parameters are usually determined by analyzing the Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and performing visual inspection for trend and seasonality [5]. The user must supply the parameters p, d, q, P, D, and Q in the SARIMA model. Selecting the settings is a difficult decision that calls for knowledge and experience. These parameters are typically determined through visual inspection for trends and seasonality, and analyzing the autocorrelation and partial autocorrelation..

2. Methodology

The time series is a random variable set $\{y_t: t \in N\}$, where N is the set of natural numbers [6]. This set is referred to as a time series [5]. Time series models typically have a stationary critical assumption, which can take either a strong or weak form [7]. For some time shift, say h, and observation y_i, the statistical distribution of $\{y_t: t \in N\}$ remains intact after shifting time scale, then:

$$P(Y_{t_1} \le y_1, Y_{t_2} \le y_2, \dots, Y_{t_k} \le y_k) = F(y_{t1}, y_{t2}, \dots, y_{tk}) = F(y_{h+t_1}, y_{h+t_2}, \dots, y_{h+t_k}) = P(Y_{h+t_1} \le y_1, Y_{h+t_2} \le y_2, \dots, Y_{h+t_k} \le y_k)$$

The variable does not affect the marginal distribution of Y_t. The absolute location associated with t_1 and t_2 has no bearing on the two-dimensional distributions of $\{Y_{t_1}, Y_{t_2}\}$. Consequently, the covariance $Cov(Y_t, Y_{t-k})$ is essentially a function of k, and the mean function E(X) is constant. Similarly, adding a continuous-time does not affect a higher-order moment [8]. A random time series is independent with continuous variance and constant mean. If the random variables in a discrete process $\{W_t\}$ come from a series of mutually identical variables, the process is called purely random. This procedure implies that:

For any k, $\gamma(k) = cov(W_t, W_{t+k}) = 0$, indicating that the system's mean and variance are constants [9]. The process $\{Y_t\}$ is said to be a random walk process if $Y_t = Y_{t-1} + W_t$. With mean of μ and variance of σ_w^2, W_t is a completely random process with $t = 0$ and $Y_1 = W_0$ at the beginning of the process, so we have:

$Y_1 = Y_0 + W_1$, at $t = 1$

$Y_2 = Y_1 + W_2 = Y_0 + W_1 + W_2$, at $t = 2$

$Y_3 = Y_2 + W_3 = Y_1 + W_2 + W_3 = Y_0 + W_1 + W_2 + W_3$, at $t = 3$

In general, $Y_t = Y_0 + \sum_{j=1}^{n} W_j$. From the above procedure, the first-order moment and the variance are equal to

$$E[Y_t] = Y_0 + \sum_{j=1}^{n} E[W_j] = Y_0 + t\mu_W = t\mu_W \quad \text{and} \quad Var(Y_t) = t\sigma_W^2$$

Disparities that create a time series data set stationary by eliminating trends are one type of filtering. It is typically necessary to separate seasonal data from the initial order for the data to achieve average stationarity [10]. Assuming that the time series $Y_t = \{Y_1, Y_2, \dots, Y_n\}$ is non-stationary, the first-order difference is $\Delta Y_t = Y_t - Y_{t-1}$ and the second-order difference is $\Delta^2 Y_{t+2} = \Delta Y_{t+2} - \Delta Y_{t+1}$. Assume that the process $\{Z_t\}$ is completely random, with a mean and variance of 0 and σ_W^2, respectively. Moving Average of order q, or MA (q), is what it is known as if $Y_t = b_0 W_t - b_1 W_{t-1} - \dots - b_q W_{t-q}$, where $\{b_1, b_2, \dots, b_q\}$ are constants, and b_0 is 1 because $W_t, t \in N$.

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + W_t$$

$$Y_t = \delta + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + W_t$$

If $E[Y_t] = E[Y_{t-1}] = \dots = E[Y_p] = \mu$, then $E[Y_t] = \mu = \delta + \theta_1\mu + \theta_2\mu + \dots + \theta_p\mu + 0$, hence

$$\mu = \frac{\delta}{1 - \theta_1 - \theta_2 - \dots - \theta_p}$$

The parameter μ is constant when $\theta_1 - \theta_2 - \dots - \theta_p < 1$ is taken into account. When $p = 1$, then the first-order autoregressive AR(1) can be express as follows [9]:

$$Y_t = \theta Y_{t-1} + W_t \quad \dots (1)$$

An alternative name for Equation 1 is the Markov process. As an endless MA process, AR(1) can be represented by the backshift operator

$$bY_t = Y_{t-1}, \text{ or } (1 - \theta B)Y_t = W_t$$

$$Y_t = W_t/(1 - \theta B) = (1 + \theta B + \theta^2 B^2 + \cdots)W_t = W_t + \theta W_{t-1} + \theta^2 W_{t-2} + \cdots$$

The variance of Y_t is the series $\sigma^2 W(1 + \theta^2 + \theta^4 + \cdots)$ converging under the constraint $|\theta| < 1$, and the mean of Y_t is zero . The autoregressive moving average (ARMA) process is produced when the model combines both AR and MA . Equation 2 represents the process ARMA, which is of order (p, q) when it contains p autoregressive terms and q moving average terms; respectively. Equation 1 can be written as follows:

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + W_t - b_1 W_{t-1} - b_2 W_{t-2} - \dots - b_q W_{t-q}$$

$$\Theta_p(H) Y_t = B_q(H) W_t \quad \dots (2)$$

where $\Theta_p(H)$ is a polynomial of order p, and $B_q(H)$ is polynomial of order q. A gradually declining autocorrelation graph combined with a steep fall in the partial autocorrelation graph, or calls for q equals 0 and p equals the biggest partial

Al-Nahrain Journal of Science
ANJS, Vol.28 (4), December, 2025, pp. 259-269

autocorrelation, are some examples of rules of thumb for selecting the parameters based on visual inspection [11]. Most of the time, p eliminates, and q is less than or equal to 3, and either p or q is 0. To keep the model as basic as feasible, a few universal ideas are to use differencing to the seasonality and trend by selecting low values for p and q, setting P and Q to 0 or 1, and differencing no more than once.

3. Multivariate Time Series

By nature, the real-time series is nonlinear, yet linear models often afford precise estimates for conclusions. A k-dimensional series z_t is said to be a linear series if and only if

$$Y_t = \nu + \sum_{j=0}^{\infty} \theta_j y_{t-j} \quad \ldots (3)$$

where ν is vector, and $\theta_0 = I_k$ the $k \times k$ identity matrix, θ_j (j >0) is a constant matrix of order $k \times k$, and $\{y_t\}$ is a sequence of random vectors which are independent and identically distributed with 0 mean and $\sum_y$ is covariance matrix. $\sum_y$ is involved to be positive-definite. The condition that $\theta_0 = I_k$ is satisfied because $\sum_y$ is allowed to be a positive-definite matrix. Another way to represent a linear time series requires θ_0 to be a lower triangular matrix with diagonal elements 1 and $\sum_y$ is a diagonal matrix [12]. Using the Cholesky decomposition of $\sum_y$, it can be achieved. Specifically, it the covariance matrix is decomposed as follows:

$$\Sigma_y = MGM^T,$$

where M and G are matrices such that, M is a $k \times k$ lower triangular with 1 being its diagonal elements, and G is a diagonal. Suppose that $b_t = M^{-1} y_t$, then $y_t = M\, b_t$,

$$\mathrm{Cov}(b_t) = \mathrm{Cov}(M^{-1}y_t) = M^{-1} \sum_y (M^{-1})^T$$
$$= M^{-1}(MGM^T)(M^T)^{-1} = G$$

With the sequence $\{b_t\}$, Equation 3 can be written as

$$w_t = \nu + \sum_{j=0}^{\infty} (\theta_j M) b_{t-j} = \nu + \sum_{j=0}^{\infty} \theta_i^* \, b_{t-j} \quad \ldots (4)$$

such that $\theta_0^* = M$ (lower triangular). $\theta_j^* = \theta_j M$, for $(j > 0)$, and the covariance matrix of b_t is a diagonal. A purely stationary stochastic process w_t, according to Wold decomposition, can be written as a linear combination of a sequence of consecutively uncorrelated process e_t. This is not identical to Equation 3, but it is close because $\{e_t\}$ do not necessarily have the identical distribution. An example of w_t , that satisfies the Wold decomposition, but not a linear time series, is the multivariate autoregressive conditional heteroscedastic process. The Wold decomposition, though, illustrates that the conditional mean of w_t, can be expressed as a linear combination of w_{t-j} for $(j > 0)$, when w_t is purely stapptionary stochastic. This explains starting with linear or nonlinear time series model because the conditional mean of w_t plays a significant role in forecasting [13]. From Equation 3, w_{t-1} is a function of $\{y_{t-1}, y_{t-2}, \ldots\}$, so the time index $t-1$, the only unfamiliar quantity of w_t, is y_t, hence, y_t is the innovation of the model w_t, at time t. The linear series w_t to be stationary, in Equation 3, the coefficient in matrices should satisfy the following $\sum_{j=0}^{\infty} ||\theta_j|| < \infty$, such that $||F||$ represent a matrix representation; for instance, the Frobenius norm $||F|| = \sqrt{tr(FF^t)}$, and Based on the properties of a convergence series, this infers that $||\theta_j|| \to 0$ as $j \to \infty$. As a result, for a stationary linear series w_t, in Equation 4, $\theta_j \to 0$ as $j \to \infty$. Moreover, it follows that:

$$E(w_t) = \nu, \qquad \text{and} \qquad \mathrm{Cov}(w_t) = \sum_{j=0}^{\infty} \theta_j \Sigma_y \theta_j^T \quad \ldots (5)$$

4. Measurement Criteria

Although autocorrelation and partial autocorrelation can be used to determine the order of p, d, and q, other specification criteria such as Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be used [14] AIC is used to select the model that minimizes the following:

$$AIC = -2\log(\text{Maximal Likelihood}) + 2K$$

where $K = p + q + 1$ if the model contains a constant, and $K = p + q$ otherwise.

BIC is used to select the model that minimizes the following:

$$BIC = -2\log(\text{Maximal Likelihood}) + K * \log(n)$$

To get a better time series model, the objective is to minimize AIC and BIC. After investigating some models, the one with the lower value can be selected as the best model. In addition, the accuracy of forecasts sales models can be measured as follows:

$$F_{Af_k} = \frac{\sum_{i=1}^{k} (A_{f_i} - F_{f_i})}{\sum_{i=1}^{n} F_{f_i}} * 100\%$$

F_{Af_k} is the forecast accuracy of the model for the item (f) in period (t). It refers to the further actions which should be carried out in the future to check up on the real behavior of the sales in the next period for which the forecast was performed. A_{f_i} and F_{f_i} denotes the actual and forecasting sales for the item (f) in period (t); respectively.

5. Proposed Model

The multi-layer perceptron (MLP) is the most commonly used in Artificial Neural Networks (ANN), and it is also called Feed-forward network. MLP computes a response (output) by propagating the explanatory variable, x, in the organized process elements, which are named nodes [14]. The j^{th} node in the hidden layer can be defined as follows:

$$h_j = G_j(\alpha_{0j} + \sum_{i\to j} w_{ij}x_i) \quad \dots (6)$$

where $G_j(.)$ is a transformation function such as Logistic, tangent, sigmod, and hyperpoplic functions.
Using Equation 6, the output node can be written as follows:

$$s = \Gamma\,(a_{0s} + \sum_{j\to s} w_{is}h_j) \quad \dots (7)$$

where $\Gamma(.)$ is a linear or an indicator function. From Equation 6 and Equation 7, it follows that:

$$s = \Gamma\left[\alpha_{0s} + \sum_{j\to s} w_{js}G_j\left(\alpha_{0j} + \sum_{i\to j} w_{ij}x_i\right)\right] \quad \dots (8)$$

The processes of ANN can be used in the time series application. Suppose $\{(x_i, y_i)\}_{i=1}^{N}$ denote to a time series data set where x_i is a vector of inputs, and y_i the output [15]. For a given x_i in Equation 8, the output of MLP denoted by $m(x_i;\theta)$ and can be defined as follows:

$$s = \Gamma\left[\alpha_{0s} + \sum_{i\to s} \alpha_{is}x_i + \sum_{j\to s} w_{js}G_j\left(\alpha_{0j} + \sum_{i\to j} w_{ij}x_i\right)\right] \quad \dots (9)$$

where θ is the neural network weight vector, and it is adjustable parameter of the neural network system that can be obtained from a technique called training [15]. Now, using the ordinary least squared (OLS) function,

$$L_N(\theta) = \sum_{i=1}^{N}\{y_i - s(x_i;\theta)\}^2 \quad \dots (10)$$

Hence, the weights for ANN time series can be found using $\hat{\theta} = \operatorname{argmin}_{\theta\in\Theta}\{L_N(\theta)\}$
By assuming that Θ is compact network weight space in R^r. Some minimizing algorithm can be used to estimate θ and the most popular is the Back-propagation algorithm [16]. For a single output, AR-NN(k;p) can be denote to the autoregressive neural network of order p with k experiments.

$$\begin{aligned} Y_t &= h(X_{t-1};\theta) + \epsilon_t \\ &= \phi_o + \phi^T X_{t-1} \\ &\quad + \sum_{j=1}^{k} \epsilon_j G\left(w_j^T X_{t-1} - C_j\right) + \epsilon_t \quad \dots (11) \end{aligned}$$

Where $X_{t-1} = \left(Y_{t-1}, \dots, Y_{t-p}\right)^T, \phi = \left(\phi_1, \dots, \phi_p\right)^T, \epsilon = (\epsilon_1, \dots, \epsilon_k)^T, C = (c_1, \dots, c_k)$, and $w_j = \left(w_{1j}, \dots, w_{pj}\right)^T$.

AR_NN can be generalized to the autoregressive moving average network which denoted by ARMA_NN(k; p, q) and defined as follows:

$$\begin{aligned} Y_t &= h(X_{t-1}, e_{t-1};\theta) + \epsilon_t \\ &= \phi_0 + \phi^T X_{t-1} + \Gamma^T e_{t-1} \\ &\quad + \sum_{j=1}^{k} \epsilon_j G\left(w_j^T X_{t-1} - \nu_j^T e_{t-1} - c_j\right) \end{aligned}$$

where $e_{t-1} = \left(e_{t-1}, \dots, e_{t-q}\right)^T$ is a $q \times 1$ input vector, and ν_j is a linear polynomial function that used to filter residuals.

ARIMA-NN Model Algorithm

Step 1: Plot the time series data to identify the unusual observations and understand the patterns.
Step 2: Stabilize the time series data using Box-Cox transformation (if needed).
Step 3: Difference the data time series to get stationary data (if needed).
Step 4: Determine a possible candidate model using ACF or PACF criterion.
Step 5: Use ARMA-NN to determine the model parameters.
Step 6: Determine the best model by using the AIC and BIC criteria.
Step 7: Plot the ACF to test the model residuals.
Step 8: If the residuals look like white noise, then calculate forecasts; otherwise (go to step 4)

6. Exploratory Data Analysis

The dataset used in this article is obtained from the United States Census Bureau (www.census.gov). The examined dataset runs from January 2014 to December 2023, a roughly decade-long timeframe. To help visualize trends and volatility [18]. Figure 1.a shows drug sale data in the original, and Figure 1.b shows the logarithmic scales. The research demonstrates a broad increasing trend in most variables, indicating that drug sales have consistently increased over time. However, a substantial drop in data level occurs at the end of 2023, indicating a likely shift or abnormality in the dataset [19]. Figure 2 depicts the first difference in the dataset, which is used to test stationarity. The transformation to first differences successfully stabilizes the mean, indicating that the data series is stationary following this adjustment. This means that no additional differencing is needed for time series modeling [20]. The difference values range from 0.20 to -0.15, indicating relatively moderate swings around the mean, reinforcing the dataset's stability post-transformation [21]. These discoveries are critical for future time series analysis and forecasting since stationary data simplifies the modelling process and improves prediction accuracy.

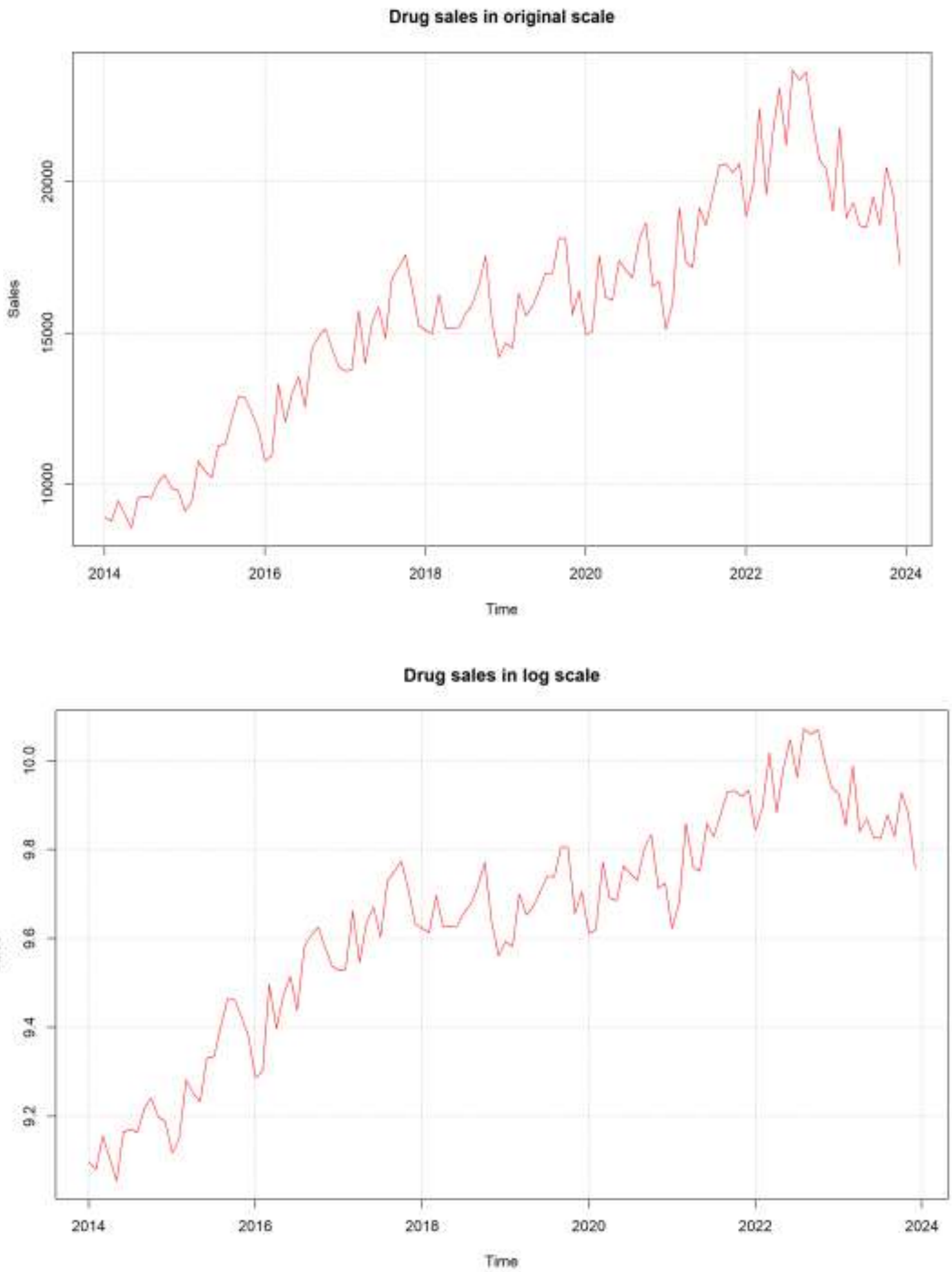

Figure 1. Drug sales in the US between 2014 and 2024 in both original and logarithmic scales.

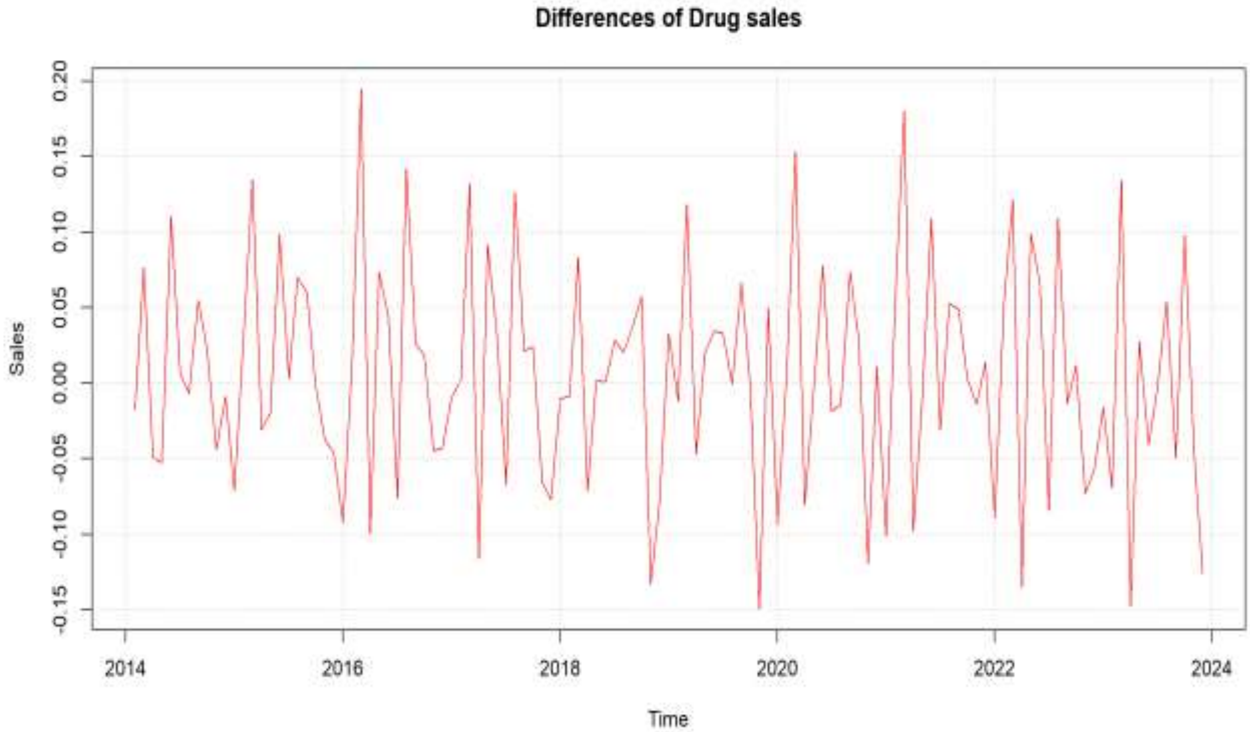

Figure 2. The first difference in drug sales in US between 2014 and 2024.

Figure 3 displays the diagnostic plot for the medication sales dataset, which shows apparent seasonal patterns and a significant rising trend over time. The seasonal fluctuations reflect periodic differences in sales; however, the increasing trend represents a general increase in sales, which is critical for long-term forecasting [22]. The insights gathered from this plot are useful in establishing the best model for estimating drug sales over the following four years. Given the seasonal and trend components observed, models such as the Seasonal Autoregressive Integrated Moving Average (SARIMA) or state space models for exponential smoothing may be useful for making accurate predictions [23]. These models can account for both seasonal effects and long-term patterns, resulting in accurate estimates for future drug sales. Further investigation of residuals after applying these models is required to validate their accuracy and ensure robust projections. As a result, a rigorous approach to model selection and validation will improve forecasting precision and facilitate strategic decision-making [24] . Although some yearly fluctuations may appear uncorrelated, this is likely due to external shocks or non-seasonal factors. The ARIMA-NN model accounts for such behavior through its nonlinear component.

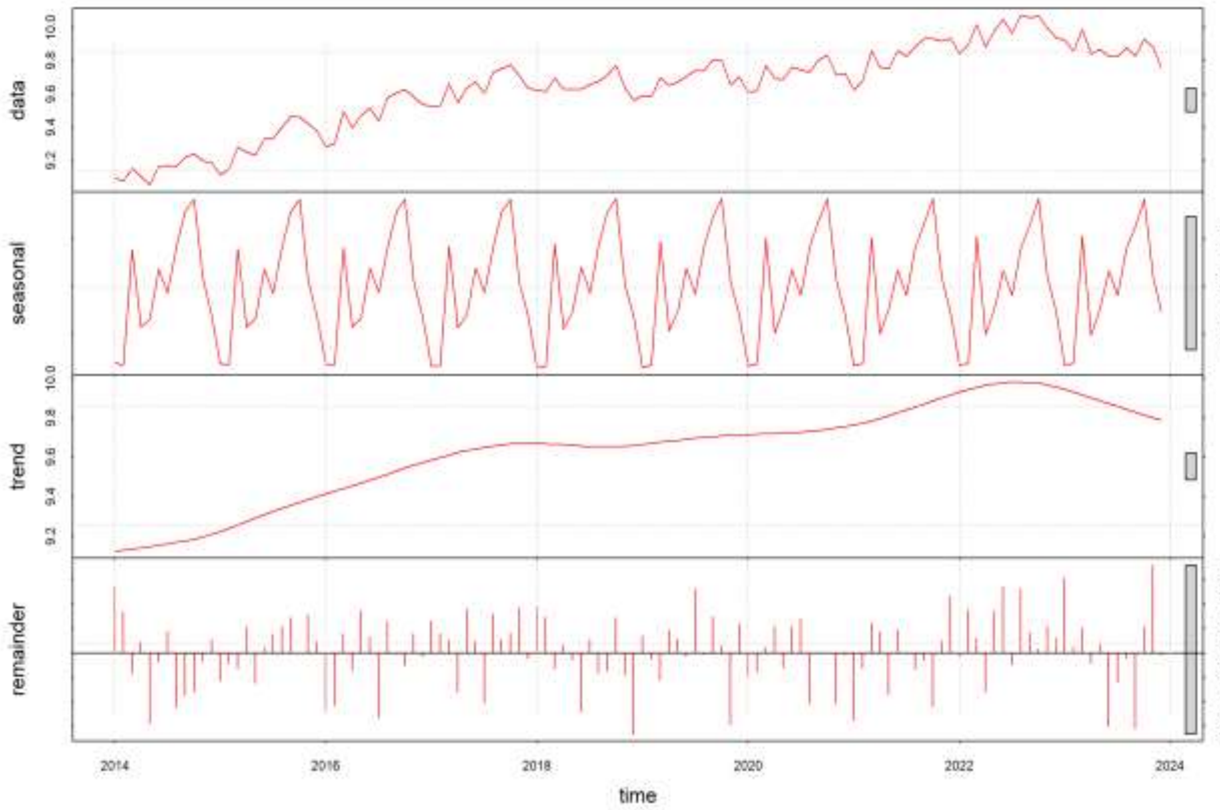

Figure 3. Comprehensive drug sales data set decomposition: seasonal, trend, and remainder

7. Result and Discussion

Table 1 compares ARIMA and ARIMA-NN models using the AIC and the BIC for different model orders and seasonal parameters. These metrics assess model fit and complexity; lower numbers indicate better performance. Model 1 (Order: c(0, 0, 1); Seasonal: c(0, 1, 1)) has the greatest AIC (175.13 for ARIMA, 183.94 for ARIMA-NN) and BIC values (193.94 for ARIMA, 143.02 for ARIMA-NN), indicating that it fits the data less effectively. A notable improvement can be seen in Model 2 (Order: c(0, 1, 1); Seasonal: c(1, 1, 1)), especially in the ARIMA-NN model, which has a significantly lower AIC of 154.42 and BIC of 144.11 than ARIMA's AIC of 177.26 and BIC of 158.95. This implies that the ARIMA-NN model better captures the structure of the data. Model 3: Seasonal: c(1, 1, 1); Order: c(1, 1, 1) demonstrates competitive results with AIC values of 160.28 for ARIMA and 142.86 for ARIMA-NN, and BIC values of 145.06 for ARIMA-NN and 142.24 for ARIMA. It does not outperform Model 4, despite its good performance. The best model is determined to be Model 4 (Order: c(0, 1, 1); Seasonal: c(1, 1, 2)), which has the lowest AIC of 130.73 and BIC of 118.14 for ARIMA-NN as opposed to 150.14 and 136.19 for ARIMA. This suggests that the model fits better and has less complexity, which makes it the most efficient model. Model 5 (Order: c(2, 1, 1); Seasonal: c(2, 1, 1)) has good performance for ARIMA-NN, with AIC and BIC values of 154.93 and 154.04, respectively, but falls short of Model 4's results. According to AIC, BIC, and the accuracy%, Model 4 effectively forecasts and analyses. It offers the greatest fit with the lowest AIC and BIC values and higher accuracy, especially when combined with the ARIMA-NN technique.

Table 1. Comparing ARIMA and ARIMA-NN models using AIC and BIC criteria for different order and seasonal values.

Model No.	Order	Seasonal	AIC		BIC		Accuracy %	Quality
			ARIMA	ARIMA-NN	ARIMA	ARIMA-NN		
Model 1	c(0, 0, 1)	c(0,1,1)	175.13	183.94	193.94	143.02	52.2%	Poor
Model 2	c(0, 1, 1)	c(1, 1, 1)	177.26	154.42	158.95	144.11	81.3%	Good
Model 3	c(1, 1, 1)	c(1, 1, 1)	160.28	142.86	142.24	145.06	78.6%	Good
Model 4	c(0, 1, 1)	c(1, 1, 2)	150.14	130.73	136.19	118.14	93.9%	The best
Model 5	c(2, 1, 1)	c(2, 1, 1)	174.26	154.93	164.04	154.04	84.7%	Good

The standardized residuals for the time series model in question are shown in Figure 4. A positive indicator of model adequacy is the plot showing that the residuals are spread within a normal range. The Normal Q-Q plot and the residuals' Autocorrelation Function (ACF) further validate the model. The model's validity is supported by the ACF plot, which indicates that residuals do not show significant autocorrelation, and the Q-Q plot, which indicates that residuals approximate a normal distribution [24]. The Ljung-Box test statistic's p-values, which determine if autocorrelation is still present in the residuals, are also given. This test's non-significant p-values attest to the model's good fit and ability to accurately represent the dynamics of the time series. All of these diagnostic tests, confirm that the forecasting model is accurate and dependable.

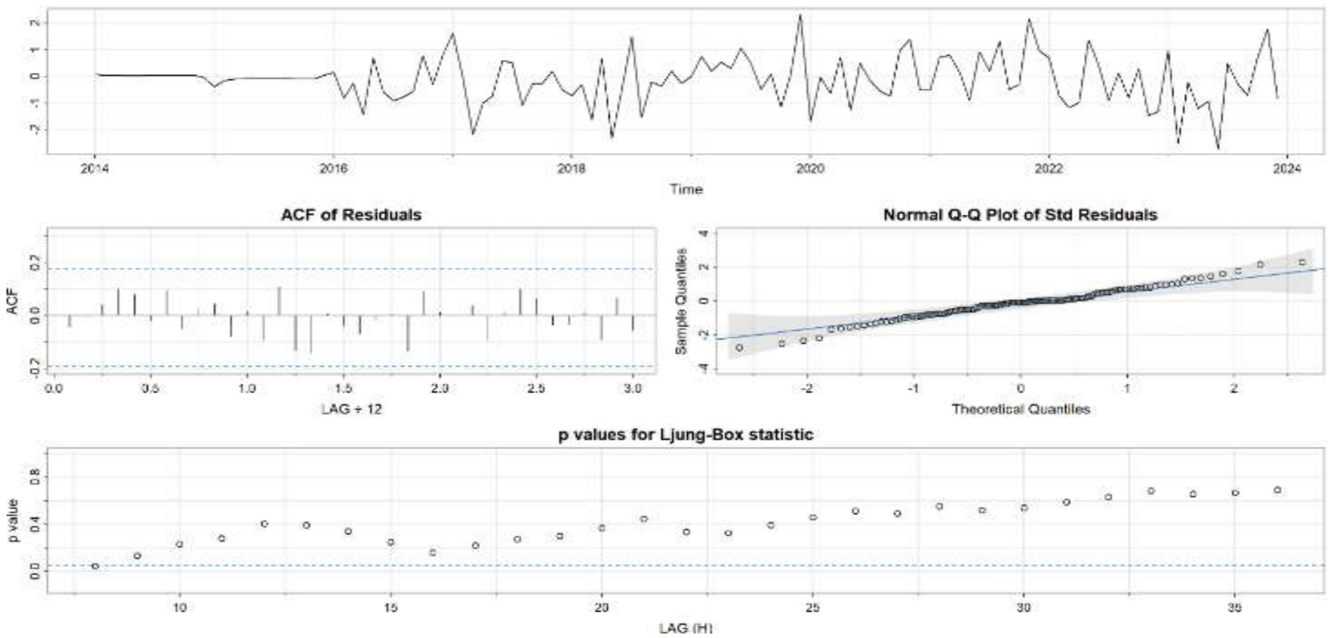

Figure 4. Auto-correlation function for the Proposed time series model in the US from 2014 to 2024.

The ACF plot and Ljung-Box test proved highly effective in assessing the autocorrelation structure of the model residuals. Their results demonstrated that the residuals approximate white noise, exhibiting no significant autocorrelation across lags. This characteristic is a critical indicator of model adequacy, as it confirms that the model has successfully captured the underlying patterns in the data without leaving systematic structures unexplained. Consequently, these diagnostic tools reinforce the reliability and robustness of the ARIMA-NN model in forecasting applications. The red line in Figure 5 represents the observed drug sales statistics from January 2014 to December 2023. The smooth level trend, which offers a deseasonalized time series perspective, is concurrent. The influence of the COVID-19 pandemic on medicine sales can be explained by the smooth level curve, which shows a distinct upward shift beginning in early 2020 [25]. During this time,

Al-Nahrain Journal of Science
ANJS, Vol.28 (4), December, 2025, pp. 259-269

medicine sales increased due to the pandemic's major effects on consumer behavior and disruptions to healthcare systems. However, an apparent decrease in the smooth level is shown as the plot moves closer to the end of 2022. As the pandemic's immediate effects subsided and healthcare systems started to normalize, this decline might indicate that drug sales patterns were returning to normal. Studies demonstrating changes in drug usage as a result of outside events and their eventual reversion as circumstances stabilize are consistent with these tendencies.

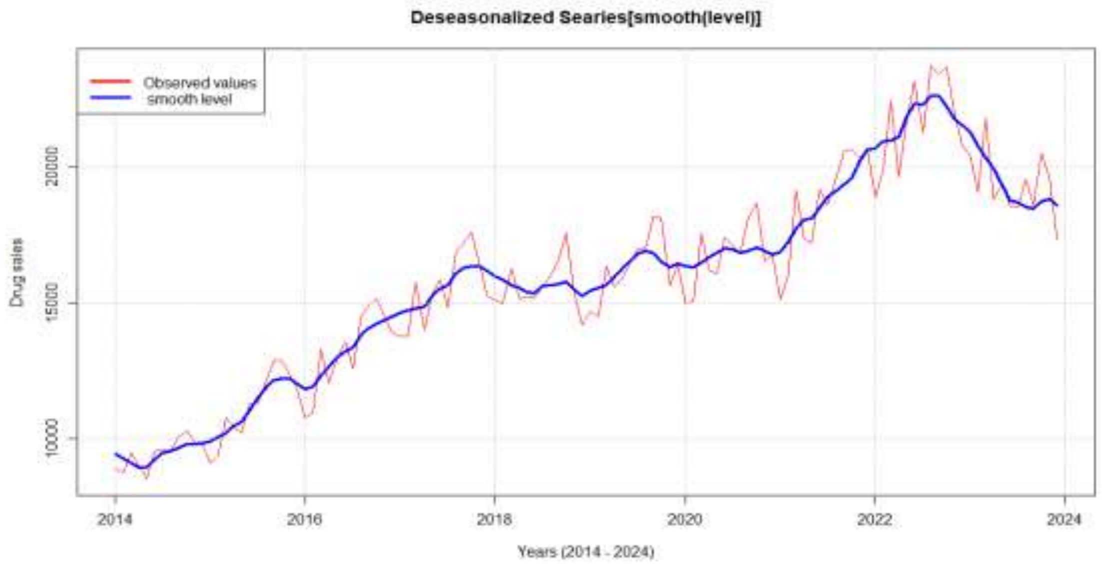

Figure 5. Deseasonalized series [smooth(level)] for the Drug sales model from 2014 to 2024.

In Figure 6, the actual values for drug sales are represented with a red line, the ARIMA fitted model is displayed with a green line, and the ARIMA-NN is illustrated with a blue line. As shown, the fitted ARIMA-NN is closer to the actual values, giving us a better prediction for the future.

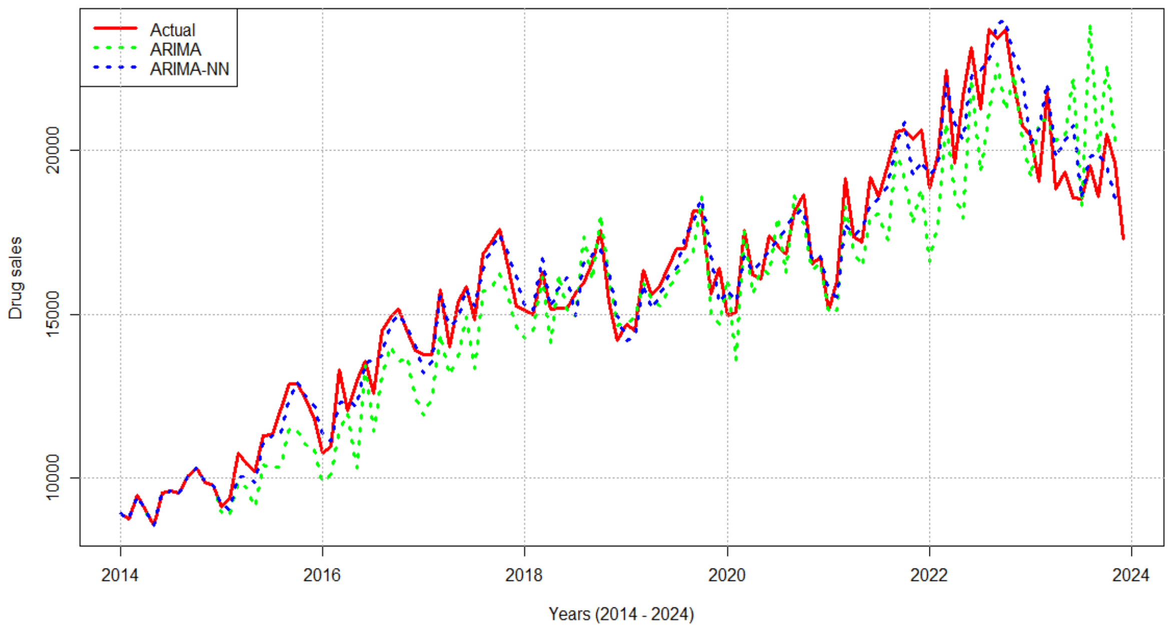

Figure 6. Fitted models and actual series for the Drug sales model from 2014 to 2024.

The chosen model [ARIMA-NN(0,1,1)(1,1,2)] is used to forecast drug sales for the next four years, from 2024 to 2027. Figure 7 illustrates these expected numbers, demonstrating the projected trends in millions of dollars during this time duration. To measure the projections' dependability, three confidence ranges were calculated: 80%, 90%, and 95%. These intervals define the range within which genuine future values are likely to fall at the chosen level of confidence. The map displays the upper and lower boundaries for each confidence interval, providing a visual picture of forecast uncertainty. This precise forecasting strategy improves knowledge and planning by accounting for variability and providing a comprehensive view of likely future events. The forecasting and confidence intervals were created with R version 4.3.2, resulting in accurate and up-to-date projections.

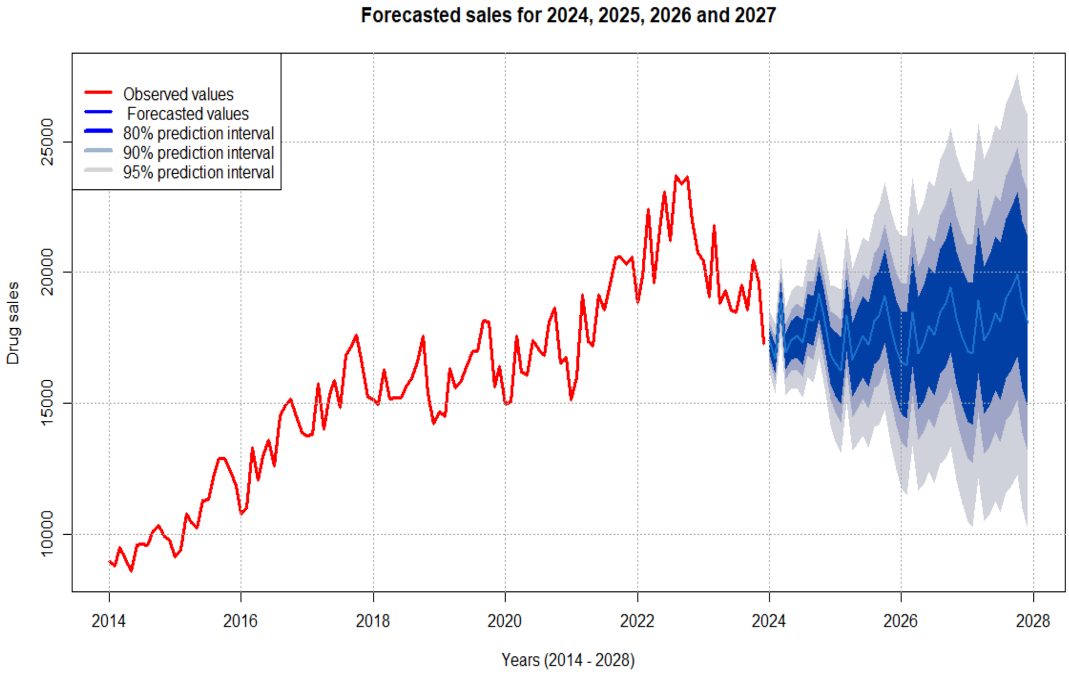

Figure 7. Forecasting for the Drug sales in billions of dollars for 2024, 2025, 2026 and 2027.

8. Conclusions

This article presents a modified version of the ARIMA model. The novel model, ARIMA-NN, is applied to the drug sales time series data set. It is compared with some ARIMA models after removing the stationarity and seasonality from the data set. The AIC and BIC criteria suggest that ARIMA-ANN with order c(0,0,1) and seasonal c(1,1,2) is the best. The residuals were assessed using ACF and a normal Q-Q plot of standardized data where the p values were more than 0.05, indicating that ARIMA-NN fit the data well and was the best. The four-year forecast sales are estimated with three confidence interval ranges, which are 80%, 90%, and 95%.

Acknowledgments: I would like to express my sincere gratitude to all those who supported me throughout this project. I am deeply thankful to [Zahraa Al-Sharea] for providing scientific guidance and invaluable insights that greatly contributed to the success of this work.

Funding: No funding has been received.

Conflicts of Interest: The authors declare no conflict of interest. The sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] Alsheheri, G.; "Comparative Analysis of ARIMA and NNAR Models for Time Series Forecasting". J. Appl. Math. Phys., 13 (1): 267–280, 2025.

[2] Kaur, J.; Parmar, K. S.; Singh, S.; "Autoregressive models in environmental forecasting time series: a theoretical and application review". Environ. Sci. Pollut. Res.,

30 (8): 19617–19641, 2023.
[3] Verma, P.; Reddy, S. V; Ragha, L.; Datta, D.; "Comparison of time-series forecasting models". in 2021 International Conference on Intelligent Technologies (CONIT) 1–7, IEEE:, 2021.
[4] Zhang, G. P.; "Time series forecasting using a hybrid ARIMA and neural network model". Neurocomputing, 50: 159–175, 2003.
[5] Dubey, A. K.; Kumar, A.; García-Díaz, V.; Sharma, A. K.; Kanhaiya, K.; "Study and analysis of SARIMA and LSTM in forecasting time series data". Sustain. Energy Technol. Assessments, 47: 101474, 2021.
[6] Shmueli, G.; Polak, J.; 'Practical Time Series Forecasting with r: A Hands-on Guide'. 2nd ed Axelrod schnall publishers:, Rockville, MD, USA, 2024.
[7] Diggle, P.; Giorgi, E.; Time Series: A Biostatistical Introduction. 1st ed, Oxford University Press:, Oxford, UK, 2025.
[8] Asparouhov, T.; Muthén, B.; "Comparison of models for the analysis of intensive longitudinal data". Struct. Equ. Model. A Multidiscip. J., 27 (2): 275–297, 2020.
[9] Singh, V. P.; Singh, R.; Paul, P. K.; Bisht, D. S.; Gaur, S.; "'Hydrological Processes Modelling and Data Analysis'". Water Sci. Technol. Libr., 127: 105–120, 2024.
[10] Brockwell, P. J.; Davis, R. A.; 'Time Series: Theory and Methods'. Springer science & business media:, New York, USA, 1991.
[11] Alsuwaylimi, A. A.; "Comparison of ARIMA, ANN and Hybrid ARIMA-ANN models for time series forecasting". Inf. Sci. Lett., 12 (2): 1003–1016, 2023.
[12] Wagner, B.; Cleland, K.; "Using autoregressive integrated moving average models for time series analysis of observational data". bmj, 383: 2023.
[13] Satrio, C. B. A.; Darmawan, W.; Nadia, B. U.; Hanafiah, N.; "Time series analysis and forecasting of coronavirus disease inp Indonesia using ARIMA model and PROPHET". Procedia Comput. Sci., 179: 524–532, 2021.
[14] Beeram, S.R.; Kuchibhotla, S.; "Time Series Analysis on Univariate and Multivariate Variables: A Comprehensive Survey". In: Proceedings of the 11th International Conference on Communication Systems and Networks (COMSNETS INDIA 2019), Bengaluru, India, 7–11 January 2019; IEEE, Eds.; IEEE: New York, NY, USA, 2020.
[15] Tsay, R. S.; "Multivariate Time Series Analysis: With R and Financial Applications". John Wiley & Sons:, 2013.
[16] Hajirahimi, Z.; Khashei, M.; "A novel parallel hybrid model based on series hybrid models of ARIMA and ANN models". Neural Process. Lett., 54 (3): 2319–2337, 2022.
[17] Jafarian-Namin, S.; Shishebori, D.; Goli, A.; "Analyzing and predicting the monthly temperature of tehran using ARIMA model, artificial neural network, and its improved variant". J. Appl. Res. Ind. Eng., 11 (1): 76–92, 2024.
[18] Elsaraiti, M.; Merabet, A.; "A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed". Energies, 14 (20): 6782, 2021.
[19] He, R.; Zhang, L.; Chew, A. W. Z.; "Modeling and predicting rainfall time series using seasonal-trend decomposition and machine learning". Knowledge-Based Syst., 251: 109125, 2022.
[20] Cheng, C.; Sa-Ngasoongsong, A.; Beyca, O.; Le, T.; Yang, H.; Kong, Z.; Bukkapatnam, S. T. S.; "Time series forecasting for nonlinear and non-stationary processes: a review and comparative study". Iie Trans., 47 (10): 1053–1071, 2015.
[21] Ahmed, S.; Nielsen, I. E.; Tripathi, A.; Siddiqui, S.; Ramachandran, R. P.; Rasool, G.; "Transformers in time-series analysis: A tutorial". Circuits, Syst. Signal Process., 42 (12): 7433–7466, 2023.
[22] Sun, F.; Meng, X.; Zhang, Y.; Wang, Y.; Jiang, H.; Liu, P.; "Agricultural product price forecasting methods: A review". Agriculture, 13 (9): 1671, 2023.
[23] Abbas, M.K.; Mahdi, G.J.; Mseer, H.A.H.; "A Multivariate Bayesian Model Using Gibbs Sampler with Real Data Application". In: Proceedings of the AIP Conference, Baghdad, Iraq, 2024; AIP Publishing: Melville, NY, USA, 2024.
[24] Abbas, I. T.; Ghayyib, M. N.; "Using Sensitivity Analysis in Linear Programming with Practical Physical Applications". Iraqi J. Sci., 907–922, 2024.
[25] Mseer, H.A.H.; Mahdi, G.J.M.; "Comparison Among Variable Selection Models and Its Application to Health Dataset". In: Proceedings of the AIP Conference, Baghdad, Iraq, 2023; AIP Publishing: Melville, NY, USA, 2023.