Hybrid Bayesian ARIAM-ANN for Population Forecasting

Ali Mohammed Ali Chichan[1] and Qutaiba Nabeel Nayef Al-Qazaz[2]
[1]*Department of Statistics Techniques, Institute of Management Rusafa, Middle Technical University, 10053 Baghdad, Iraq*
[2]*Department of Statistics, College of Administration and Economics, University of Baghdad, 10071 Baghdad, Iraq*
alimohammed520@mtu.edu.iq, dr.qutaiba@coadec.uobaghdad.edu.iq

Keywords: Autoregressive Integrated Moving Average, Bayesian Method, Markov Chain Monte Carlo Algorithm, Artificial Neural Network, Feed-Forward Back Propagation, Forecasting.

Abstract: This paper compares two forecasting methodologies: The Bayesian ARIMA model and a hybrid Bayesian ARIMA- ANN framework. This study uses historical population data for Iraq (1970–2024) to build predictive models for the period 2025–2034. In the hybrid model, the Bayesian method is employed to optimally estimate the ARIMA parameters while calculating forecast uncertainty intervals. The outputs of the Bayesian model are subsequently utilized as inputs for an artificial neural network (ANN). This integration allows the neural network to capture nonlinear patterns and complex relationships between the Bayesian outputs and actual population trends. The results, as indicated by the Mean Absolute Percentage Error (MAPE) criterion, demonstrated a substantial superiority of the hybrid model, which achieved the lowest criterion value of 0.31, in contrast to the Bayesian ARIMA model's value of 49.31. This improvement is attributed to the model's ability to combine the precision of Bayesian estimation with the flexibility of neural networks in modelling complex relationships. The study confirms that integrating Bayesian methods with artificial intelligence techniques significantly enhances the accuracy of long-term population forecasts, offering a reliable tool for strategic development planning.

1 INTRODUCTION

Precise population forecasts are crucial for strategic planning, particularly in unstable environments such as Iraq. Traditional hybrid models (ARIMA-ANN) can capture linear and nonlinear associations; however, they fail to account for the significant uncertainty stemming from socio-political swings in Iraqi data and lack a method for incorporating past knowledge into the forecasts.

This research introduces a sophisticated hybrid model that integrates Bayesian ARIMA with artificial neural networks. Bayesian integration facilitates the quantification of uncertainty and the integration of prior information, yielding more stable and dependable predictions in the intricate context of Iraqi data.

A multitude of studies has examined this subject. For instance, [1] employed the Box-Jenkins methodology to develop an ARIMA model for forecasting Pakistan's population over a twenty-year horizon. Multiple ARIMA models were evaluated, and the ideal model was selected based on error metrics. The study determined that the ARIMA (1,2,0) model is the most effective and straightforward model, as it produced low AIC values and a random distribution of residuals. [2] developed a hybrid model that integrates the linear statistical approach ARIMA with artificial neural networks (ANNs) for nonlinear modeling. The hybrid model initially employed ARIMA models to forecast wind speed, subsequently utilizing an ANN to examine the resultant errors and model the residual nonlinear patterns of those errors, so enhancing the overall prediction accuracy. The employed ARIMA models were of the AR(p) variety without differencing (d=0), and the model parameters were estimated using the Box-Jenkins methodology for ARIMA models. The artificial neural networks (ANNs) employed a basic two-layer architecture featuring two inputs and one output following experimentation with various combinations. The findings indicated the dominance of the hybrid model, exhibiting the lowest error levels across all statistical metrics. [3] utilized the Box-Jenkins ARIMA model to forecast the annual output of 34 agricultural crops using annual data. The model technique involved identifying the order (p, d, q) by the analysis of autocorrelation and partial autocorrelation functions, as well as assessing stationarity with the Augmented Dickey-Fuller test. The ideal model was selected based on metrics including the modified coefficient of determination (R^2), Akaike's information criterion (AIC), and the mean absolute percentage error (MAPE). The ideal models differed across various crops. The optimal model for tea was ARIMA (2,1,2), but the optimal model for cardamom was ARIMA (0,1,2). [4] developed a novel hybrid model integrating discrete wavelet analysis (DWT), linear ARIMA models, and nonlinear artificial neural networks (ANNs). The model utilized Discrete Wavelet Transform (DWT) to analyze the time series, decomposing it into two primary components: the detailed (high-frequency) component, which signifies linear patterns, and the approximate (low-frequency) component, which signifies nonlinear patterns. Subsequently, each component was modeled utilizing the proper technique. ARIMA was employed to model the detailed component, whilst ANN was utilized to model the approximation component, using residual errors from the ARIMA model. [5] Researchers in Zhejiang Province, China, utilized annual data from 1978 to 2016 to construct the model. The research methodology employed time series analysis, commencing with an assessment of data stationarity via series plots and autocorrelation tests. The analysis determined that the original series was non-stationary and necessitated a first difference (d=1) to attain stationarity. The best model was determined with the AIC and SBC criterion. The findings demonstrated that the ARIMA (1,1,0) model exhibited optimal performance across all data sets. The model's accuracy was validated by juxtaposing its 2017 forecasts with actual data, which fell within the 95% confidence interval, thereby confirming the model’s reliability. [6] employed the Autoregressive Moving Average (ARMA) model within a Bayesian framework to estimate parameters, assuming a normal-gamma prior distribution for the likelihood function and utilizing a quadratic loss function. The estimation method involved deriving the posterior distributions of the coefficients by iterative integration and applying Bayes' theorem. The marginal posterior distribution was obtained, and the optimal Bayesian estimator for the model parameters was established. The study's principal finding is the feasibility of effectively implementing the Bayesian method in ARMA modeling with a normal-gamma prior distribution, highlighting that the adopted mathematical framework offers flexibility in incorporating prior knowledge with time series data to enhance estimation and prediction accuracy. [7] examined two primary models: Bayesian structural time series (BSTS) and autoregressive integrated moving average (ARIMA) models, to forecast the incidence of COVID-19 infections, fatalities, and vaccinations across five countries. The results indicated that the BSTS model significantly outperformed in prediction accuracy based on RMSE, MAE, and MAPE for the majority of the data. The study's primary findings demonstrate that Bayesian models exhibit greater flexibility and accuracy in the analysis of time series components compared to traditional models. [8] applied the ARIMA framework for estimating the demographic future of Bangladesh. The ARIMA methodology was employed in accordance with the Box-Jenkins procedure to examine 34 discrete time series. Model parameters were estimated using (MLE), and the optimal model for each series was determined based on the minimum corrected Akaike Information Criterion (AIC) value. The model's efficacy was validated by the Box-Ljung test to confirm that the residuals exhibit characteristics of white noise. The chosen models exhibited significant variation across different age and gender demographics. The findings demonstrated that the integrated forecasting method was superior regarding precision. The research validated the efficacy of the integrated ARIMA model with a comprehensive methodology in delivering an accurate and dynamic representation of future demographics.

This study seeks to create a Bayesian Hybrid ARIMA-ANN model for population estimation in Iraq, emphasizing the mitigation of uncertainty in estimates via incremental probability updates. The study will utilize structured historical data, and the model's efficacy will be assessed by the MAPE metric. This model is anticipated to serve as a valuable instrument for decision-makers in Iraq, facilitating optimal future planning through enhanced accuracy and reliability in population estimates.

2 METHODS AND MATERIALS

2.1 ARIMA Model

Time series models are frequently employed in forecasting. The Autoregressive Integrated Moving Average (ARIMA) model is among the most significant models, owing to its superior forecasting efficacy [9]. This framework was introduced by Box and Jenkins in 1970 [10]. The Box-Jenkins methodology involved integrating autoregressive (AR) and moving average (MA) components, together with an differencing (I) component to achieve stationarity in the series. Consequently, the contemporary ARIMA model, denoted as ARIMA(p, d, q), was developed. This methodology enabled the analysis of non-stationary time series and facilitated precise predictions based exclusively on the series' internal dynamics.

The ARIMA model's time series comprises three primary components, as outlined below [11]:

1) Autoregressive (AR): Illustrates the correlation between the present value of the series and its historical values. This is represented by the subsequent equation:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t. \quad (1)$$

Where: y_t: is the current value of the time series at time t.

- c: is a constant (intercept);
- $\phi_1, \phi_2, \dots, \phi_p$: are model parameters that express the effect of past values.;
- ϵ_t: is the random error at time t.

2) Integration (I): Denotes the procedure of applying differences to transform a time series into a stationary series. If the series is non-stationary, differences are computed until it attains stationarity. The differential process is articulated as follows:

$$\Delta y_t = y_t + y_{t-1}. \quad (2)$$

Where Δy_t represents the difference between the current value and the previous value.

3) Moving Average (MA): Illustrates the correlation between the present value and historical errors. This is articulated as follows:

$$y_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t\,. \quad (3)$$

Where:

- c: Mean of the series
- $\theta_1, \theta_2 \dots, \theta_q$: model coefficients that express the effect of past errors.
- ϵ_t: random error at time t.

The standard representation of the ARIMA model, incorporating these three elements, is:

$$\mathrm{y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t}. \quad (4)$$

Equation (4) may be reformulated as follows:

$$\phi(B)(1-B)^d y_t = \theta(B)\epsilon_t\,. \quad (5)$$

Where:

$$\phi(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right), \quad (6)$$

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right), \quad (7)$$

$$(1-B)^d y_t = \Delta^d y_t\,. \quad (8)$$

The formulation of an ARIMA model necessitates the identification of three key parameters [11]:

- p: The order of Autoregressive, indicating the number of lagged dependent variable values.
- d: The degree of differencing required to render the time series stationary.
- q: The order of the Moving average component, specifying the number of lagged forecast errors.

Figure 1 presents a schematic diagram of the ARIMA modeling procedure.

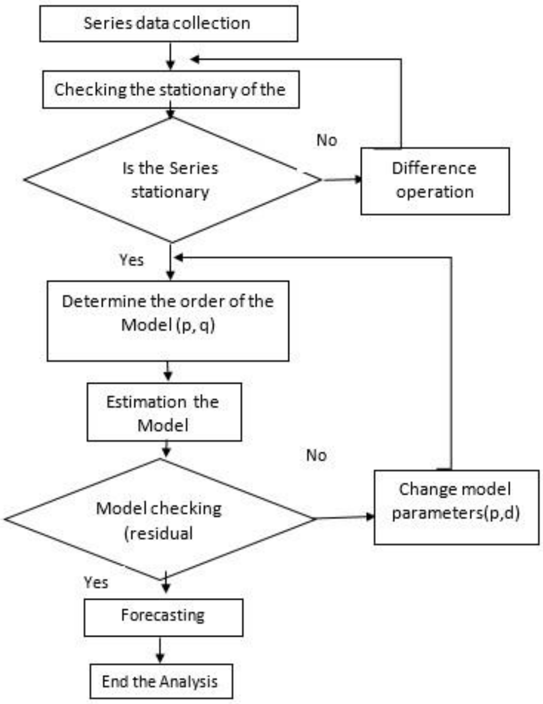


Figure 1: Flow chart for ARIMA model.

2.2 Estimating ARIMA Model Parameters

This section estimates the parameters of the ARIMA model to forecast the population size in Iraq. The model parameters will be determined by the Bayesian method, followed by the use of Feed Forward Back Propagation neural networks to enhance the prediction process. The Bayesian approach will be integrated with neural networks, resulting in a hybrid model.

2.2.1 Bayesian Method

The Bayesian method is a sophisticated statistical technique for estimating parameters of ARIMA models. It integrates existing knowledge of the parameters with the maximum likelihood function of the ARIMA model. The method utilizes Bayes' theorem to deduce the posterior probability distribution of the parameters instead of providing a single point estimate. This allows for the direct computation of uncertainty intervals and the integration of external information into the estimation process [12]. Stochastic simulation techniques, such as Markov Chain Monte Carlo (MCMC), are employed to sample from the posterior distribution [13] and [14]. This approach provides significant adaptability in modeling and uncertainty evaluation, particularly for small samples or complex models. The ARIMA model can be expressed in the following form [1]:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \,. \tag{9}$$

The first step to application Bayes' theorem is to determine the maximum likelihood function for the ARIMA model as follows:

$$\text{L}(\text{y}/\phi_i, \theta_j,\text{p},\text{q},\sigma_\epsilon^2) \approx (2\pi\sigma_\epsilon^2)^{-\frac{(n-p)}{2}} \exp\left(-\frac{1}{2\sigma_\epsilon^2}\textstyle\sum_{i=p+1}^{n} \sigma_\epsilon^2\right). \tag{10}$$

Where:

$$\epsilon_t = y_t - \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_j \epsilon_{t-j}.$$

The second stage involves specifying prior distributions, a fundamental aspect of Bayesian methodology. The prior distributions for the parameters of the ARIMA model are specified as follows:

$$\begin{gathered} \text{p} \sim \text{Discrete Uniform}(0, p_{\max}), \\ \text{q} \sim \text{Discrete Uniform}(0, q_{\max}), \end{gathered} \tag{11}$$

$$\begin{gathered} \phi_i \sim \text{Normal}(0, \sigma_\phi^2) \ \ \text{i=1,...,p}, \\ \theta_j \sim \text{Normal}(0, \sigma_\theta^2) \ \ \text{j=1,...,q}, \end{gathered} \tag{12}$$

$$\begin{gathered} \sigma_\phi^2 \sim \text{inverse} - \text{gamma}\ (\alpha_\phi, \beta_\phi), \\ \sigma_\theta^2 \sim \text{inverse} - \text{gamma}\ (\alpha_\theta, \beta_\theta) \end{gathered} \tag{13}$$

$$\sigma_\epsilon^2 \sim \text{inverse} - \text{gamma}\ (\alpha_\epsilon, \beta_\epsilon). \tag{14}$$

The Bayesian hierarchy is:

$$\text{p}(\phi_i{}^p, \theta_j{}^q,\text{p},\text{q},\sigma_\epsilon^2/y) \propto \text{p}(\text{y}/\phi_i{}^p, \theta_j{}^q,\text{p},\text{q},\sigma_\epsilon^2).\text{p}(\phi_i{}^p/ \text{p},\sigma_\phi^2).\text{p}(\sigma_\theta^2/\alpha_\phi, \beta_\phi).p(\theta_j{}^q/q, \sigma_\theta^2).\text{p}(\sigma_\theta^2/ \alpha_\theta, \beta_\theta).\text{p}(\sigma_\epsilon^2/\alpha_\epsilon, \beta_\epsilon).p(p/\,p_{\max}).p(q/\,q_{\max})\,. \tag{15}$$

The third stage involves determining the posterior distribution, which represents the revised probability distribution of the model parameters after integrating the prior distribution with the maximum likelihood function. This is determined mathematically through the application of Bayes' theorem.

The posterior distribution delineated in (15) poses considerable analytical difficulties. The primary aim is to derive marginal posterior distributions for individual parameters, which requires integrating the joint distribution across all other parameters. In the context of ARIMA models, this integration is computationally intensive. Markov Chain Monte Carlo (MCMC) methods offer a viable alternative by facilitating direct simulation and sampling from the posterior distribution [13]. Adequate sampling results in an empirical distribution that approximates the true posterior distribution. The Gibbs Sampler serves as the fundamental algorithm within this framework [14] and [15]. This methodology iteratively updates each parameter according to its full conditional distribution—the probability distribution conditioned on the current values of all other parameters. The Gibbs sampling procedure functions as follows:

$$\begin{gathered} \sigma_\phi^2/\alpha_\phi, \beta_\phi, \emptyset, p, y \ \sim \text{IG}\left(\alpha_\phi + \frac{p}{2}\beta_\phi + \frac{1}{2}\emptyset^T\emptyset\right), \\ \sigma_\theta^2/\alpha_\theta, \beta_\theta, \theta, q, y \ \sim \text{IG}\ (\alpha_\theta + \tfrac{q}{2}\beta_\theta + \tfrac{1}{2}\theta^T\theta), \\ \sigma_\epsilon^2/\alpha_\epsilon, \beta_\epsilon, \emptyset, \theta, p, q, y \ \sim \text{IG}\left(\alpha_\epsilon + \frac{n}{2}\beta_\epsilon + \frac{1}{2}\epsilon^T\epsilon\right). \end{gathered} \tag{16}$$

The model parameters are updated in two stages, the first for the AR coefficients and the second for the MA coefficients, as follows:

$$\text{p}(\phi_i/p, q, \theta_j, \sigma_\epsilon^2, \sigma_\phi^2, y) \propto e^{-\frac{1}{2\sigma_\epsilon^2}\sum_{t=p+1}^{n}\sigma_t^2)}.p(\phi_i/\sigma_\phi^2 \tag{17}$$

$$\text{p}(\theta_i/p, q, \emptyset_j, \sigma_\epsilon^2, \sigma_\theta^2, y) \propto e^{-\frac{1}{2\sigma_\epsilon^2}\sum_{t=p+1}^{n}\sigma_t^2)}.p(\theta_i/\sigma_\theta^2). \tag{18}$$

The Metropolis-Hastings (MH) algorithm [16] and [17] offers a sampling method for situations where conditional distributions are non-standard or do not have closed-form solutions. Due to the non-standard shapes of the conditional distributions for AR and MA coefficients (Equations (17) and (18), Metropolis-Hastings sampling stages were integrated into the Gibbs Sampler. This methodology produces candidate parameter values and determines their acceptance based on computed acceptance probability. Alterations in parameter space dimensionality due to the addition or removal of coefficients necessitate transitions between various dimensional spaces. The Reversible Jump (RJ) algorithm [18] and [19] tackles this issue by facilitating comparisons across different model dimensions while ensuring consistent sampling for models with identical p and q orders. The procedural steps for employing Bayes' theorem in ARIMA parameter estimation are outlined below [20]:

1) Determine initial values for all parameters.
2) Generate samples for the variances using the probability density function from (16).
3) Generate p from the samples for the AR coefficients from (17).
4) Generate q from the samples for the MA coefficients from (18).
5) Update the model ranks (p) by increasing or decreasing them (p→ $p+1$ or p → $p-1$). This means either losing or generating one coefficient for ϕ_i to take the place of ϕ_i.
6) Update the model ranks (q) by increasing or decreasing them(q→ $q+1$ or q → $q-1$). This means either losing or generating one coefficient for θ_j to take the place of θ_j.

Steps 3 and 4 involve sampling from a non-standard probability density function. Therefore, the Metropolis-Hastings (MH) algorithm will be used. This algorithm uses an accept or reject process to update parameter values.

2.2.2 Neural Networks

Artificial neural networks are computational systems modeled after the human brain. A fundamental type is the feedforward backpropagation (FFBP) network, where data flows from input to output in a forward manner, while errors propagate backward to modify the network's weights and enhance its performance.

This paper employs a feed-forward back propagation (FFBP) neural network architecture of three fundamental layers: an input layer, a single hidden layer, and an output layer. The input layer contains R input nodes, often initialized with random weights. The hidden layer comprises M neurons, with the optimal number determined by the empirical formula $M = 2R + 1$. A random weight w is assigned to each input variable Z during the network initialization [21] and [22]. The weighted inputs from all input variables R across M neurons, along with a bias tearm b, are summed to form the total input to the transfer function. The aggregate input (sum) to the transfer function f can be articulated mathematically as:

$$sum=\sum_{i}^{M}\sum_{j}^{R} w_{i,j} z_j + b. \qquad (19)$$

Prevalent activation functions in neural networks comprise tan-sigmoid, log-sigmoid, and linear functions, with the ideal selection contingent upon the characteristics of the data and the output specifications. Activation functions in the hidden layer establish the correlation between input and output, whereas output layer functions modify the final value range. The mathematical expressions for the fundamental functions are [14]:

$$f(sum) = sum\,, \qquad (20)$$

$$f(sum) = \frac{1}{1+e^{-sum}}, \qquad (21)$$

and

$$f(sum) = \frac{2}{1+e^{-2sum}} - 1\,. \qquad (22)$$

Where sum was defined in (19).

The random weight matrix for R of inputs and M of neurons can be written as [23]:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,R} \\ M & M & 0 & M \\ w_{M,1} & w_{M,2} & \cdots & w_{M,R} \end{bmatrix}$$

The input vector can be formulated as follows:

$$Z = [Z_{l1} \quad Z_{l2} \quad \ldots\ Z_{lR}]^T\,. \qquad (23)$$

In this architecture, V=Z denotes the input vector to the hidden layers, while T=Z represents the input to the output layer. The output of the first hidden layer is given by $T = \hat{Z} = f(N)$, where N is the summation function in the first hidden layer, and the final network output is $U = \hat{Z} = f(A)$, where A denotes the summation function at the output layer. The parameter H indicates the number of neurons in the first hidden layer, which also corresponds to the number of inputs to the output layer. The symbol O refers to the number of units in the output layer, typically equal to 1 in most practical applications. The FFBP network architecture is illustrated in Figure 2 [22].

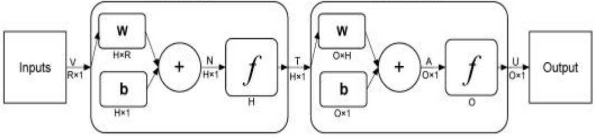


Figure 2: Components of ANN.

Neural network training aims to adjust the random weights and biases at each layer to improve prediction accuracy. through feed-forward back propagation (FFBP) networks require supervised training to minimize modeling and prediction errors. The selection of the optimal training algorithm depends on achieving the lowest possible prediction error, which is influenced by the selection of appropriate weights and biases, as well as an appropriate training function. The Levenberg-Marquardt (LM) and Bayesian regularization (BR) algorithms are among the most efficient training functions for backpropagation algorithms [4] and [24].

The ARIMA model demonstrates inherent statistical constraints in modeling nonlinear time series, as it presupposes linear correlations among observations. This work presents a hybrid model, Bayesian ARIMA-ANN, to overcome this constraint by integrating linear and nonlinear components. This hybrid model integrates the linear stability characteristic of ARIMA with the capacity of neural networks to approximate nonlinear processes.

2.2.3 Hybrid Bayesian ARIMA-ANN

The hybrid Bayesian ARIMA-ANN model is a sophisticated framework for time series analysis, integrating the statistical precision of linear Bayesian modeling with the adaptability of neural networks to capture nonlinear interactions. This integration produces more dependable predictive estimations by concurrently modeling both linear and nonlinear components, while offering a quantitative assessment of uncertainty. The model serves as an optimal solution for intricate forecasting applications within the demographic and economic domains. The hybrid methodology functions in the following manner:

1) Time series analysis is conducted using the Box-Jenkins methodology to determine the optimal ARIMA(p,d,q) model.
2) The ARIMA(p,d,q) model structure identified in Step (1) is employed to construct a Bayesian ARIMA framework.
3) The parameters of the Bayesian ARIMA model are estimated using Bayesian methods to obtain the optimal estimator.
4) The inputs for the Artificial Neural Network (ANN) are structured using the predictive outputs (e.g., forecasts, residuals, or confidence intervals) of the Bayesian ARIMA model.
5) The Artificial Neural Network is trained using a Feedforward Backpropagation (FFBP) algorithm. The hidden layer utilizes a log-sigmoid activation function, the output layer employs a purlin type, and the training procedure is executed using the Levenberg-Marquardt (LM) optimization algorithm.
6) The outputs of the trained network represent the final forecasts generated by the Hybrid Bayesian ARIMA-ANN method.
7) The Mean Absolute Percentage Error (MAPE) is computed to assess and compare the forecasting accuracy of the standalone Bayesian ARIMA method with the Hybrid Bayesian ARIMA-ANN method.

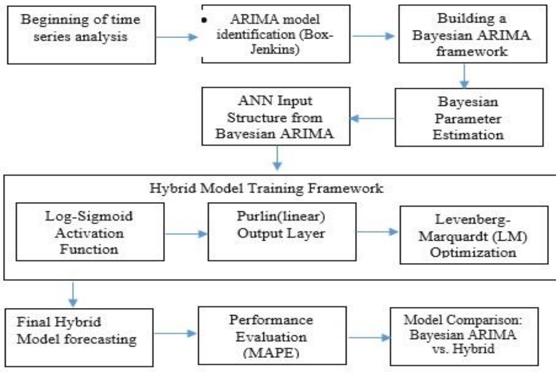


Figure 3: General Architecture of the hybrid Bayesian ARIMA – ANN.

3 EVALUATION METRICS

The predictive efficacy of the Bayesian technique and the hybrid model is evaluated using the Mean Absolute Percentage Error (MAPE) criterion. MAPE assesses the accuracy of time series forecasting by computing the average absolute percentage deviation between actual and projected values. Reduced MAPE values signify enhanced model precision, with figures below 10% generally denoting great accuracy and those above 50% indicating poor performance. This metric is especially appropriate for model comparison because of its intuitive interpretability [25].

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} * 100\% \ . \qquad (24)$$

4 APPLICATIONS

Official population censuses in Iraq, executed by the Ministry of Planning and the Commission of Statistics and GIS, furnish essential data for precise insights into population size, distribution, and diverse demographic, social, and economic characteristics. The research is based on a comprehensive time series of 55 annual measurements spanning from 1970 to

2024. The data, available via official publications and the organization's digital portal (refer to Appendix(A)), are visually encapsulated in Figure 4.

This historical data series underpins the projection of Iraq's population until the year 2034.

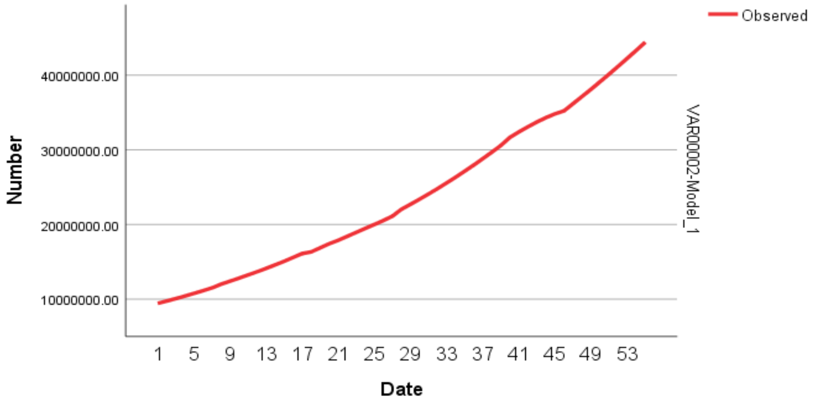
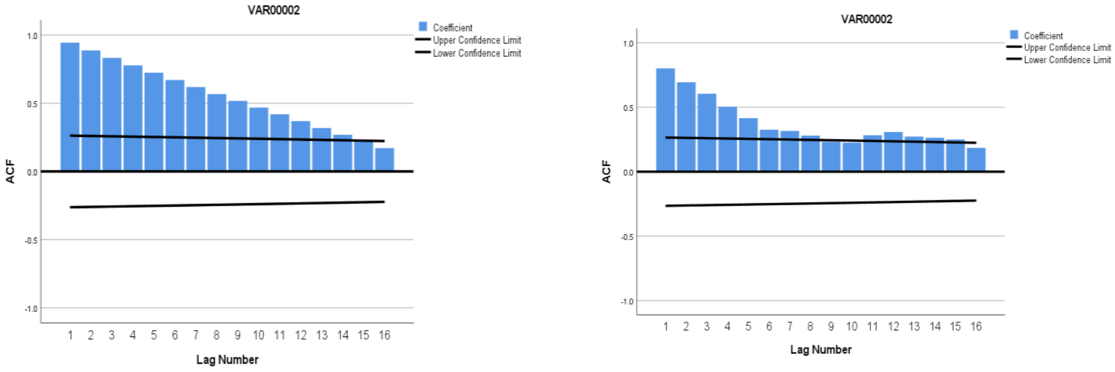
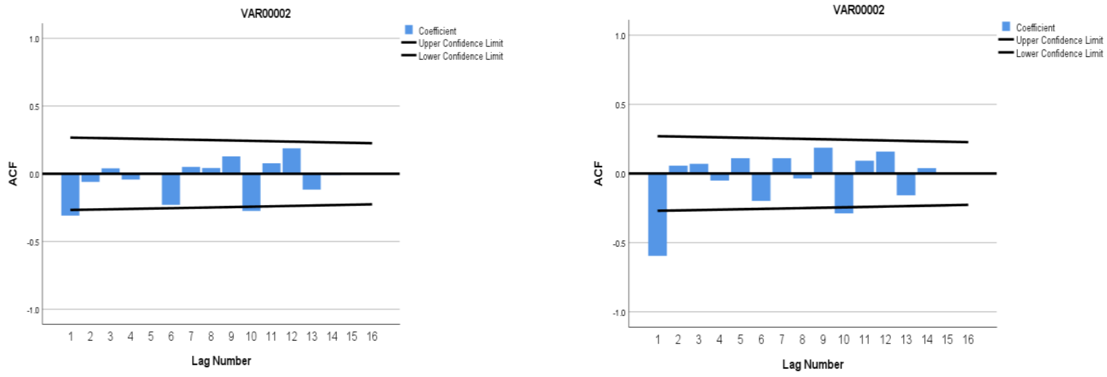


Figure 4: Population census of Iraq from1970 to 2024.



No different

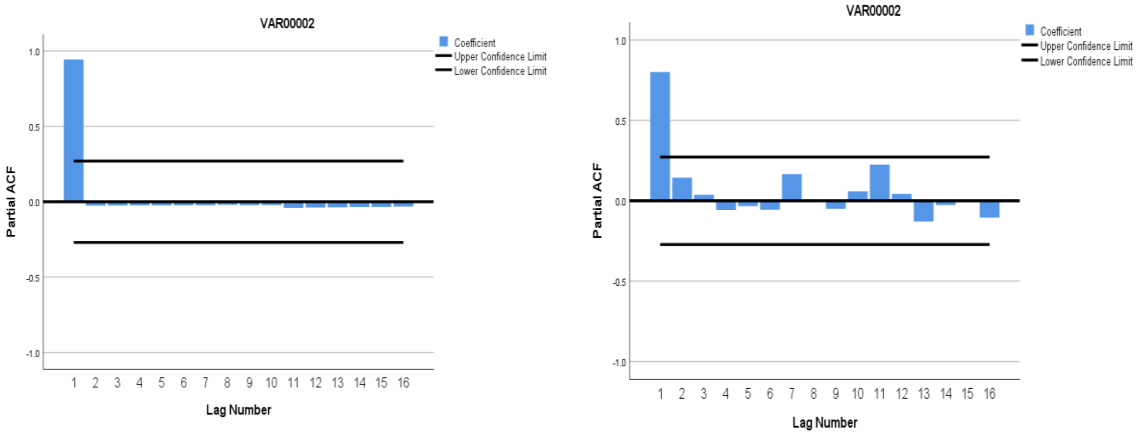
One different



Two different

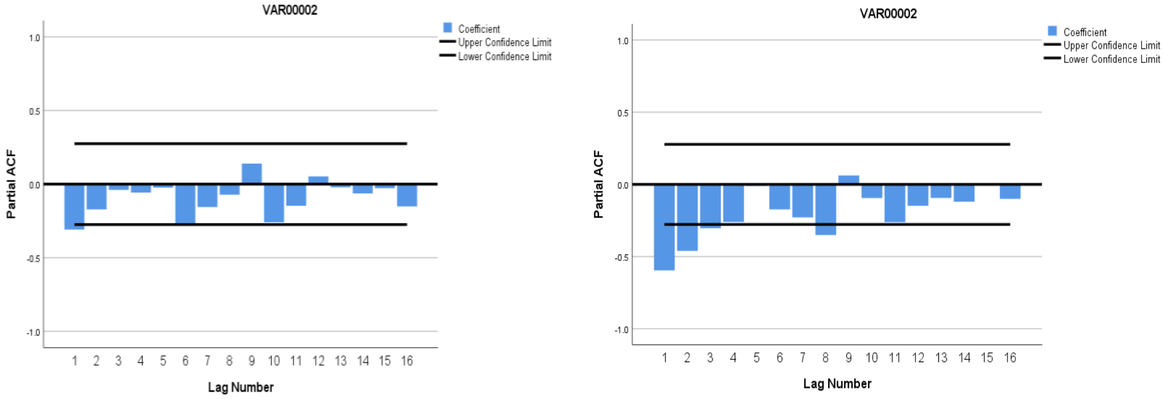
Three different

Figure 5: ACF of the data after taking the differences.



No different

One different



Two different

Three different

Figure 6: PACF of the data after taking the differences.

SPSS was employed to identify the ARIMA model while evaluating the time series of Iraq's population data from 1970 to 2024. The method commences with the assessment of the stationarity of the time series through the analysis of the autocorrelation function (ACF) and the partial autocorrelation function (PACF). Upon analyzing the original data without differentiation, the findings of the autocorrelation function reveal that the values commence at a high level (Lag 1 = 0.944) and diminish gradually as the time lag grows, plainly suggesting that the series is non-stationary and exhibits a general trend. To resolve this issue, first-order differentiation was applied. The autocorrelation and partial autocorrelation values are negligible and fall within the confidence intervals. When the correlations approach the confidence bounds, they are nearly zero, indicating statistical insignificance. This indicator signifies that the series has lost the essential property of autocorrelation in time series analysis. The series can be characterized as independent and identically distributed, sometimes referred to as white noise. Upon applying third-order differentiation, it was noted that the autocorrelation function exhibited one significant lag at lag 1 (ACF = -0.595), whereas the partial autocorrelation function rapidly diminished following the initial lag. This pattern indicates that the most suitable model is a one-parameter moving average model (ARIMA (0,3,1)). According to these findings, third-order differentiation utilizing the ARIMA (0,3,1) model was selected as the ideal model, attaining a coefficient of determination of 0.603, a MAPE of 0.291%, and a Ljung-Box residual statistic of 0.148. This outcome validates the model's efficacy in data representation and its applicability in forecasting once the time series has achieved stationarity. The subsequent Figures 5, 6 depict the autocorrelation function and the partial autocorrelation function.

Analysis revealed that the initial series is non-stationary, and stationarity was not attained after the first difference, although it was nearly obtained following the second difference. Nonetheless, the third difference distinctly revealed the structure of the MA (1) model, corroborated by the SPSS program's findings, which indicated a steady coefficient of determination ($R^2 = 0.603$) and the Ljung-Box test, demonstrating no significant autocorrelation in the residuals (Sig = 0.148). Consequently, the ARIMA (0,3,1) model may be selected as the definitive model to characterize the series.

The ARIMA (0,3,1) model used can be written as an equation as follows:

$$(1-B)^3 y_t = -\theta_1 \epsilon_{t-1} + \epsilon_t \qquad (25)$$

$$y_t = 3y_{t-1} - 3y_{t-2} + y_{t-3} - \theta_1 \epsilon_{t-1} + \epsilon_t \quad (26)$$

Following the establishment of the model representing the population data in Iraq, the researcher proceeded to the parameter estimation phase utilizing the Bayesian method, as outlined in the study's theoretical framework. An algorithm was developed in the R statistical programming language to estimate model parameters using Bayesian inference methods for this purpose. the estimation yielded a parameter value (θ_1 = -0.9663), The estimated model will be used to forecast the population size in Iraq from 2025 to 2034. The predictive outcomes are presented in Table 1.

Upon finalizing the estimation of the ARIMA model parameters via the use of Bayesian methods and acquiring the linear forecasts, the hybrid model construction proceeds using the Bayesian forecasts from the ARIMA model as fundamental inputs for the hybrid model.

An artificial neural network was developed and executed for forecasting using data from 1970 to 2024. The dataset was partitioned as follows: The training period comprised 55 observations from 1970 to 2024, whereas the test period consisted of 10 observations from 2025 to 2034. The network architecture has four input variables and one output variable, utilizing a hybrid algorithm that integrates Bayesian ARIMA with artificial intelligence (artificial neural network) techniques. Figure 7 illustrates the architecture of the fundamental neural network employed in the study.

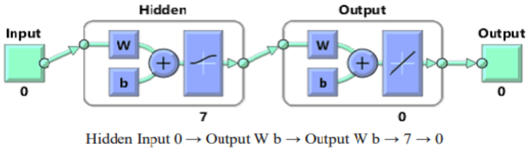


Figure 7: Structure of the ANN.

From Figure 7, shows a feedforward network architecture with single hidden layer contains 7 neurons, the hidden layer utilizes a log-sigmoid activation function, while the output layer employs a purlin type.

The initial theoretical neuron count of 9, derived from the equation M=2R+1, was revised to 7 neurons during experimental validation, since the simplified model demonstrated enhanced predictive accuracy (reduced MAPE) and circumvented overfitting while preserving computational efficiency.

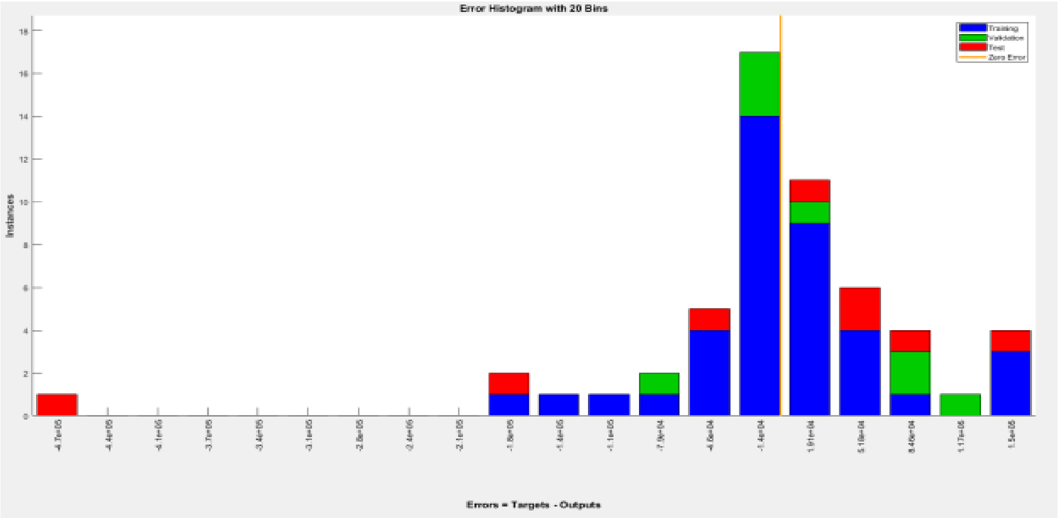


Figure 8: Error distribution (histogram).

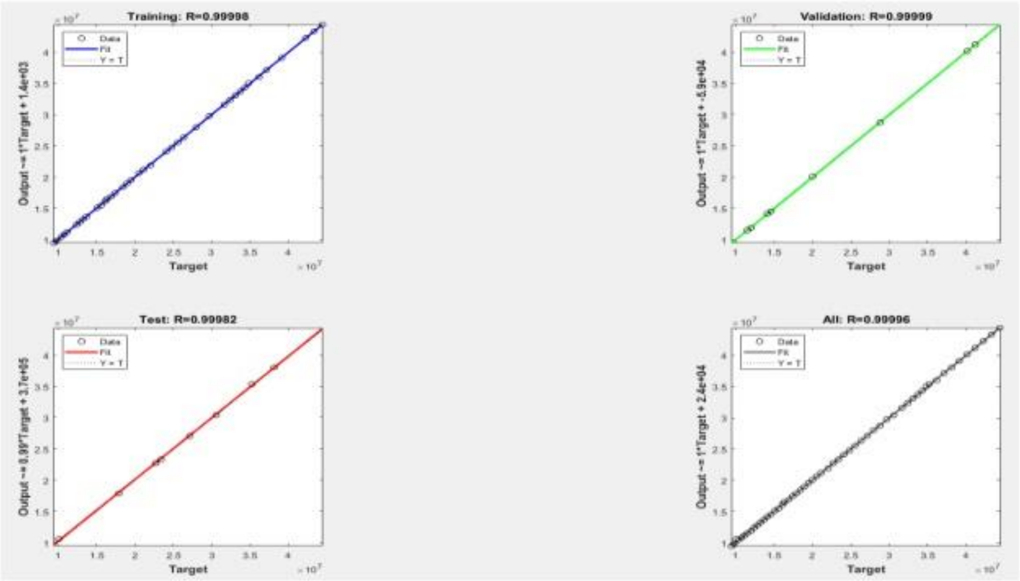


Figure 9: Regression and correlation analysis.

From Figure 8, illustrates that the error distribution consists of 20 bins covering an extensive range of values. The distribution displays Gaussian properties with symmetry centered at zero.

From Figure 9, illustrates the results of the regression and correlation study, revealing remarkably high correlation coefficients: Training (R = 0.99998), Validation (R = 0.99999), and Test (R = 0.99996). The near-unity values (R ≈ 1) signify exceptional model accuracy and reliability, endorsing its use for strategic planning in the 2025-2034 timeframe. Moreover, the study demonstrates uniformly linear correlations throughout all datasets, as indicated by the slope attributes.

After completing the training and validation phases of the hybrid model, population forecasting for Iraq was performed. The ensuing projections, presented in Table 1, are as follows:

Table 1: Forecasting population of Iraq from 2025 to 2034.

Years	Bayesian ARIMA	Hybrid Bayesian ARIMA-ANN
2025	45509530	45487426
2026	46608227	46019308
2027	47710883	46950852
2028	48817500	47741439
2029	49928078	48510663
2030	51042615	49256458
2031	52161113	49976942
2032	53283571	50670444
2033	54409989	51335526
2034	55540367	51970992

After executing forecasts with the Bayesian ARIMA and Hybrid Bayesian ARIMA-ANN methodologies, the models' relative accuracy is evaluated using the Mean Absolute Percentage Error (MAPE) metric.

Table 2: MAPE value.

Model	Bayesian ARIMA	Hybrid Bayesian ARIMA-ANN
MAPE	49.31%	0.31%

The hybrid Bayesian ARIMA-ANN model surpassed the Bayesian ARIMA model in the MAPE evaluation criterion presented in Table 2 by attaining the lowest criterion value. This outcome demonstrates that the integration of ANNs with the Bayesian ARIMA model substantially enhances forecasting precision.

Figure 10: Forecast plot for hybrid Bayesian ARIMA-ANN model.

Figure 10 shows hybrid model forecasts, with blue lines representing historical data and orange lines future estimates. The gray region shows model confidence in forecasts. The model matches the data trend within the margin of error.

5 DISCUSSIONS

The pronounced advantage of the hybrid Bayesian ARIMA-ANN model evidenced by the substantial reduction in MAPE from 49.31% to 0.31% is ascribed to its distinctive capacity to concurrently analyze linear and nonlinear elements of the data. The Bayesian component concentrated on fundamental linear patterns and uncertainty assessment, whereas the artificial neural network effectively captured intricate nonlinear interactions that conventional, models are unable to express. Incorporating linear model residuals as supplementary inputs to the neural network significantly improved the model's capacity to identify concealed and indirect patterns within the data.

6 CONCLUSIONS

The hybrid Bayesian ARIMA-ANN model represents a substantial advancement in the modeling of intricate time series, exhibiting exceptional efficacy in forecasting Iraq's population by attaining the lowest MAPE criterion value relative to conventional models. The strategic significance of this improved forecast accuracy resides in establishing a dependable basis for national planning for from 2025 to 2034 within health, education, and infrastructure sectors. These findings further validate the dependability of population predictions as a vital decision-support instrument, facilitating the government's efficient resource allocation and preparation for future challenges. Thus, the research advocates for the implementation of this hybrid technique as a standard tool for long-term strategic planning, due to its capacity to enhance the precision and dependability of forecasts in demographic and economic sectors.

The research recommends generalizing the model as a standard tool in long-term strategic planning and applying it to population data from different demographic and economic contexts to test its generalizability and robust performance across multiple environments.

REFERENCES

[1] M. Zakria and F. Muhammad, “Forecasting the population of Pakistan using ARIMA models,” Pakistan Journal of Agricultural Sciences, vol. 46, no. 3, pp. 214-223, 2009.

[2] E. Cadenas and W. Rivera, “Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model,” Renewable Energy, vol. 35, no. 12, pp. 2732-2738, Dec. 2010, [Online]. Available: https://doi.org/10.1016/j.renene.2010.04.022.

[3] P. C. Padhan, “Application of ARIMA model for forecasting agricultural productivity in India,” J. Agric. Soc. Sci., vol. 8, no. 2, pp. 50-56, 2012.

[4] I. Khandelwal, R. Adhikari, and G. Verma, “Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition,” in Proc. Int. Conf. Intell. Comput., Commun. Convergence, vol. 48, pp. 173-179, 2015, [Online]. Available: https://doi.org/10.1016/j.procs.2015.04.167.

[5] J. Dai and S. Chen, “The application of ARIMA model in forecasting population data,” in Proc. 2nd Int. Conf. Phys., Math., Stat., vol. 1324, Art. no. 012100, 2019, [Online]. Available: https://doi.org/10.1088/1742-6596/1324/1/012100.

[6] Z. Amry, “Bayesian estimate of parameters for ARMA model forecasting,” Tatra Mt. Math. Publ., vol. 75, pp. 23-32, 2020, [Online]. Available: https://doi.org/10.2478/tmmp-2020-0002.

[7] M. N. Thorakkattle, S. Farhin, and A. A. Khan, “Forecasting the trends of Covid-19 and causal impact of vaccines using Bayesian structural time series and ARIMA,” Ann. Data Sci., vol. 9, no. 5, pp. 1025-1047, 2022, [Online]. Available: https://doi.org/10.1007/s40745-022-00418-4.

[8] K. Z. Hossain, “Predicting the Demographic Future of Bangladesh: Application and Comparison of ARIMA and Combined Population Forecasts,” Romanian Statistical Review, vol. 2, 2024.

[9] H. Song, S. F. Witt, and T. C. Jensen, “Tourism forecasting: Accuracy of alternative econometric models,” Int. J. Forecast., vol. 19, no. 1, pp. 123-141, 2003.

[10] D. Asteriou and S. G. Hall, Applied Econometrics: A Modern Approach Using EViews and Microfit, Rev. ed. New York, NY, USA: Palgrave Macmillan, 2007.

[11] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken, NJ: John Wiley & Sons, 2016.

[12] D. Barber, A. T. Cemgil, and S. Chiappa, Eds., Bayesian Time Series Models. Cambridge, U.K.: Cambridge University Press, 2011.

[13] D. Gamerman and H. F. Lopes, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2006.

[14] G. Lozano Orozco, “Markov Chain Monte Carlo approach to the analysis and forecast of grain prices and volatility monitoring,” M.S. thesis, Dept. Math. and Phys., Inst. Tecnol. y de Estudios Superiores de Occidente, Tlaquepaque, Mexico, 2022.

[15] G. Casella and E. I. George, “Explaining the Gibbs Sampler,” The American Statistician, vol. 46, no. 3, pp. 167-174, Aug. 1992.

[16] S. Chib and E. Greenberg, “Understanding the Metropolis-Hastings Algorithm,” The American Statistician, vol. 49, no. 4, pp. 327-335, Nov. 1995.

[17] C. Fulton, “Bayesian Estimation and Forecasting of Time Series in statsmodels,” in Proc. 21st Python Sci. Conf. (SciPy 2022), pp. 89-96, 2022.

[18] P. J. Green and D. I. Hastie, “Reversible Jump MCMC,” Genetics, vol. 155, no. 3, pp. 1391-1403, Jul. 2009.

[19] D. H. A. Montcho, “Bayesian variable selection using data driven reversible jump: an application to schizophrenia data,” M.S. thesis, Interinstitutional Graduate Program in Statistics, Univ. of São Paulo, São Carlos, Brazil, 2022.

[20] W. G. U. Turbey, “Identification of ARMA Models by Bayesian Methods Applied to Streamflow Data,” in 9th International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, Sweden, Jun. 11-15, 2006.

[21] K. Palit and D. Popovic, Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications. London: Springer London, 2005.

[22] K. Khairudin et al., “Enhancing riverine load prediction of anthropogenic pollutants: Harnessing the potential of feed-forward backpropagation (FFBP) artificial neural network (ANN) models,” Results in Engineering, vol. 22, p. 102072, Apr. 2024, [Online]. Available: https://doi.org/10.1016/j.rineng.2024.102072.

[23] J. L. Ticknor, “A Bayesian regularized artificial neural network for stock market forecasting,” Expert Systems with Applications, vol. 40, no. 14, pp. 5501-5506, 2013.

[24] H. Yonaba, F. Anctil, and V. Fortin, “Comparing sigmoid transfer functions for neural network multistep ahead streamflow forecasting,” Journal of Hydrologic Engineering, vol. 15, no. 4, pp. 275-283, Apr. 2010.

[25] F. Farida, Y. Yuniar, M. F. Mayandah, and N. U. Nurissaidah, and D. Yuliati, “Forecasting population of Madiun Regency using ARIMA method,” CAUCHY: Jurnal Matematika Murni dan Aplikasi, vol. 7, no. 3, pp. 420-431, 2022.

APPENDIX

Table A1: Total population of Iraq from 1970 to 2024.

Year	Population	Year	Population	Year	Population	Year	Population	Year	Population	Year	Population
1970	9440000	1980	13238000	1990	17890000	2000	24085784	2010	32400205	2020	40150174
1971	9750000	1981	13669000	1991	18419000	2001	24813365	2011	33088782	2021	41190658
1972	10072000	1982	14110000	1992	18949000	2002	25564835	2012	33725178	2022	42248883
1973	10413000	1983	14586000	1993	19478000	2003	26340227	2013	34304693	2023	43324018
1974	10765000	1984	15077000	1994	20007000	2004	27139585	2014	34819301	2024	44414794
1975	11124000	1985	15585000	1995	20536000	2005	27962968	2015	35212600		
1976	11505000	1986	16110000	1996	21124000	2006	28810441	2016	36169123		
1977	12000497	1987	16335199	1997	22046244	2007	29682081	2017	37139519		
1978	12405000	1988	16882000	1998	22702211	2008	30577798	2018	38124182		
1979	12821000	1989	17428000	1999	23382068	2009	31664466	2019	39127889		