STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–16.
Published online in International Academic Press (www.IAPress.org)

Using Bayesian AR-ESN for climatic time series forecasting

Shahla Tahseen Hasan, Osamah Basheer Shukur*

Department of Statistics and Informations, College of Computer science and Mathematics, University of Mosul, Mosul, Iraq

Abstract Bayesian ARIMA models will offer a solid approach for analyzing time series data, providing more flexibility than traditional recursive models. They also effectively combine previous knowledge with current data to handle uncertainty. A particular kind of Bayesian ARIMA model with comparable considerations is called a Bayesian AR model. While Bayesian models employ prior information to estimate a wide range of possible parameter values, older methods frequently use maximum likelihood estimation to obtain single values for parameters. In order to effectively handle uncertainty, they also develop a posterior distribution. The applicability of Bayesian techniques to AR(p) models is examined in this work. It demonstrates their capacity to manage noisy, non-stationary, or incomplete data while allowing for thorough probabilistic inference, which improves uncertainty comprehension and validates probabilistic forecasts. The Bayesian AR model states that present values are linearly dependent on past values, which are further amplified by white noise. We use previous distributions to evaluate the variance and establish the model parameters. Consequently, these values are adjusted in response to observations, resulting in more complex and adaptable dimensional distributions. The Bayesian ARIMA model aids in forecasting and drawing conclusions when time series are more complicated and need variance considerations. Bayesian AR(p) models display the temporal correlations between data points regardless of how stationary they are. These models are commonly estimated using Markov Monte Carlo (MCMC) techniques like Metropolis-Hastings and Gibbs sampling. These models perform well when handling asymmetry, incomplete data, and structural changes. Even when used in a Bayesian manner, traditional models struggle to capture uncertain time series or intricate nonlinear patterns. These contemporary issues can be resolved with the appropriate use of an Echo State Network (ESN). An effective recursive neural network for forecasting evolving time series is the ESN. To identify the most effective inputs for the ESN, the hybrid Bayesian AR-ESN methodology utilizes the optimal configuration of the Bayesian AR model. The capacity of this approach to accurately simulate nonlinear interactions is recognized. A Bayesian AR model and an ESN model were integrated in this hybrid Bayesian AR-ESN methodology study. The results show that combining Bayesian AR and ESN significantly increases forecasting accuracy, particularly when forecasting error metrics are used. When compared to conventional techniques, the Bayesian model significantly increases predictive accuracy.

Keywords Autoregressive (AR) Model, Bayesian Autoregressive (AR) Model, Echo State Network (ESN), Bayesian AR-ESN Model, Wind Speed Forecasting

DOI: 10.19139/soic-2310-5070-2902

1. Introduction

A chronologically ordered set of observations where each value depends on the previous one is called a time series. This pattern produces a sort of movement that can be used to forecast potential future events. Weather-related variables and maximum wind speed are examples of chaotic time series. Among the variables that influence these characteristics are temperature and air pressure. Combining these two creates a nonlinear pattern that is challenging to depict using traditional linear techniques like the autoregressive (AR) model. [1, 2] . Forecasting is challenging due to the variability and instability of maximum wind speed data over time. Linear models may inadequately represent the complexity of structures and the rapidity of changes. Various models, including AR

*Correspondence to: Osamah Basheer Shukur (Email: drosamahannon@uomosul.edu.iq). Department of Statistics and Informations, College of Computer science and Mathematics, University of Mosul, Mosul, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

and ARIMA, have been proposed for forecasting time-dependent variables. However, they encounter issues as they solely rely on point estimates and are unable to address uncertainty or nonlinear interactions. Consequently, Bayesian models are particularly intriguing, as they represent a significant advancement in time series analysis. They provide a comprehensive probabilistic representation of parameters and predictors, utilizing past knowledge to address data deficiencies and rectify discrepancies [3, 4]. Barnett, Kohn, and Sheather (1996) present a complete Bayesian treatment of autoregressive model estimation incorporating choice of autoregressive order, enforcement of stationarity, treatment of outliers and allowance for missing values and multiplicative seasonality. [5] proposes a Bayesian procedure to the ARIMA model to forecast the peak season demand which uses a dummy variable to account for the past years intermittent demand. To capture uncertainty in the Bayesian ARIMA model, the non-informative prior distributions are assumed for each parameter. Oyelam and Ogundeji in 2024 employed Bayesian statistical methods to model and forecast inflationary dynamics in Nigeria over two decades (2003 to 2024) by Adopting a Bayesian Seasonal ARIMA. The analysis incorporates prior knowledge and provides robust uncertainty quantification in parameter estimation and forecasting [6]. In 2025, Singh et, al. attempt to analyze infant mortality rate data using one of the well-known time series models (ARIMA model) and focused on Bayesian estimation of the parameters of the model considered [7]. Bayesian models are robust; nonetheless, they are incapable of completely encapsulating intricate, nonlinear events. Researchers have been employing artificial intelligence techniques, particularly the Echo State Network (ESN), a variant of recurrent neural networks designed for analyzing time-dependent series [8, 9] . This work employs a hybrid Bayesian AR-ESN strategy that integrates a Bayesian AR model with an ESN. Initially, the time series is analyzed to determine the optimal rank utilizing autocorrelation functions (ACF) and partial autocorrelation functions (PACF). This assists in determining the structure of the AR model. The ESN subsequently receives this structure as input. employing this model to predict the maximum wind speeds in Mosul on a daily basis. To ensure uniformity within each seasonal group, the data were categorized into two seasons (hot and cold) from (2012/07/16 to 2023/01/17) [10, 11]. Metrics like the Mean Squared Error (MSE) and the Mean Absolute Percentage Error (MAPE) showed that the hybrid model produced noticeably more accurate forecasts when comparing the results of the conventional AR model with the ESN-based AR model. The enhancement is attributable to the ESN's capacity to demonstrate nonlinear correlations and dynamic interactions among temporal variables that linear models fail to encapsulate [12, 13]. By employing a state space model to ascertain the fundamental structure of the Echo State Network, including the formulation of internal matrices and the randomization of feedback weights. The geographical model effectively analyzes climate data. Researchers have utilized it to illustrate historical temperature data and seasonal variations. For instance, they employed discrete time series to forecast the minimum and maximum temperatures in the Makkah region or utilized deep ESNs within additive decomposition models [14]. Other research has employed Echo State Networks (ESNs) with wavelet transforms to analyze both seasonal and non-seasonal time series components, while in our study, we developed a novel method to integrate ESNs with an AR model. Initially, a Bayesian technique was used to compute the AR model. With random errors removed, the variables that were obtained from it were used as ESN inputs. Thus, we used two different approaches. The AR coefficients, their magnitudes, and their signals are used as inputs to the ESN in the first technique. The second approach ignores the values and signals and only uses the variables [15]. This hybrid approach provides a versatile and adaptive means to forecast chaotic time series, such as wind velocity. The results indicate that it significantly outperforms existing models in terms of accuracy and the management of structural complexity [16].

2. Methodology

2.1. Autoregressive (AR) Model

The autoregressive process can be used to express the current value of the time series using the linear regression function of the previous p values of the time series. The AR(p) autoregressive model of order p^{th} can generally be written as follows.

$$\phi(B)Z_t = a_t \tag{1}$$

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)Z_t=a_t \tag{2}$$

or

$$Z_t=\phi_1 Z_{t-1}+\phi_2 Z_{t-2}+\cdots+\phi_p Z_{t-p}+a_t \tag{3}$$

where $\phi-k$ is the kth autoregressive coefficient or parameter that reflects the impact of altering the kth autoregressive variable Z_{t-k} on Z_t the autoregressive model,$k=1,2,3,\ldots,p,\ a_t$, is the independent and identically distributed random error that signifies a white noise process with a mean of zero and constant variance σ_a^2 , or can be expressed as $a_t \sim i.i.d.N(0,\sigma_a^2)$, and . If the time series data requires any order of differencing to achieve mean stationarity $W_t=(1-B)^d Z_t$, W_t is utilised in place of Z_t in equations (1), (2), and (3). The primary properties of AR(p) are enumerated below.

i. The variance of AR(p) is written as follows.

$$\gamma_0=\sigma_Z^2=\sigma_a^2/\left(1-\phi_1\rho_1-\phi_2\rho_2-\cdots-\phi_p\rho_p\right). \tag{4}$$

ii. The autocorrelation of AR(p) is as follows.

$$\rho_k=\phi_1\rho_{k-1}-\phi_2\rho_{k-2}-\cdots-\phi_p\rho_{k-p} \tag{5}$$

where $k=1,2,3,\ldots$.

iii. iii. The partial autocorrelation of AR(p) can be just evaluated as significant or insignificant such as follows.

$$\phi_{kk}\neq 0; k\leq p \tag{6}$$

$$\phi_{kk}=0; k>p. \tag{7}$$

The present value is contingent upon the two preceding values. Each ϕ coefficient indicates the extent of the impact of the respective historical value on the present value. A large coefficient value indicates that past values significantly influence current values, rendering the model appropriate for forecasting in time series that are heavily reliant on historical data.

2.2. Bayesian ARIMA and Bayesian AR models

The traditional ARIMA(p,d,q) framework is enhanced by Bayesian ARIMA models, which integrate prior distributions on parameters and estimate posterior distributions via simulation methods. The conventional ARIMA equation is

$$\phi(B)(1-B)^d Z_t=\theta(B)a_t, a_t\sim N\left(0,\sigma_a^2\right) \tag{8}$$

where
- $\phi(B)=1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p$ denotes the autoregressive (AR) polynomial,
- $\theta(B)=1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$ signifies the moving average (MA) polynomial,
- B is the backshift operator $(BZ_t=Z_{t-1}$.
- d is the order of differencing.
- a_t is the white noise error term, assumed to be normally distributed: $a_t\sim N\left(0,\sigma_a^2\right)$.

Bayesian approaches utilize previous knowledge regarding parameters $(e.g.,\phi_i,\theta_j,\sigma^2)$ and observed data to generate posterior distributions. This diverges from conventional methods that depend on maximum likelihood estimation. This facilitates analysts in elucidating uncertainty and doing model averaging when the model order (p, d, q) is indeterminate. Bayesian ARIMA models are particularly beneficial for: • Bayesian ARIMA models are particularly useful for analyzing time series that display data gaps or are intermittently accessible, such as situations where seasonal data is not available. • Incorporating priors to integrate expert viewpoints, • It is crucial to assess the uncertainty of forecasts and parameters, particularly in unstable systems like inventory management or financial prediction. Bayesian ARIMA facilitates the integration of structural knowledge or external constraints

(such as anticipated totals over time) in high-dimensional or limited forecasting scenarios through the application of conditional simulation techniques. Bayesian implementations typically use sampling techniques such as Gibbs sampling or Metropolis-Hastings, both of which are forms of Markov Chain Monte Carlo (MCMC). These techniques facilitate the acquisition of posterior samples for the latent variables of the ARIMA model [17, 18, 19]. ARIMA models based on point estimates by Maximum Likelihood Estimation and do not inherently quantify parameter uncertainty, while Bayesian ARIMA improve the traditional methodology via prior distributions for its parameters and using simulation methods MCMC to obtain the posterior distributions in order to handle uncertainty.

2.2.1. Prior Distribution In Bayesian statistics, the prior distribution reflects the analyst's beliefs regarding the model parameters prior to observing any data. Bayesian ARIMA models use prior distributions for the moving average coefficients $\phi = (phi_1, phi_2, dots, phi_p$, autoregressive coefficients $\theta = (\theta_1, \theta_2, \ldots, \theta_q$, and innovation variance σ^2 . Employing distinct Gaussian priors for the autoregressive and moving average coefficients, alongside an inverse-Gamma prior for the variance, is a prevalent and straightforward approach.
$\phi_i \sim N(\mu_\phi, \sigma_\phi),\ \theta_j \sim N(\mu_\theta, \sigma_\theta),\ \sigma^2 \sim inverse - Gamma(\alpha, \beta)$

These selections are favored because of their conjugacy qualities, which facilitate posterior updates in Gibbs sampling. The inverse-Gamma prior σ^2 ensures that the complete conditional posterior of the variance remains within the same distributional family, facilitating calculations during each sampling step. In hierarchical Bayesian ARIMA models, the variances σ^2_ϕ and σ^2_θ may be assigned hyperpriors, often derived from the inverse-Gamma distribution. This enables the model to determine the appropriate level of regularization based on the data.

$$\sigma^2_\phi, \sigma^2_\theta \sim inverse - Gamma(\alpha, \beta)$$

Such formulations are most effective in scenarios with limited data and a significant danger of overfitting. In real life, people often use weak or unhelpful priors (like a small epsilon) to reduce the effect of priors when they don't know much about the topic. In probabilistic modeling of uncertainty, prior distributions are occasionally assigned to the parameters p, d, and q in the ARIMA model. It is commonly presumed that there exists a consistent prior over a specific interval.

$$p, d, q \sim Discrete\ Uniform(0, p_{max})$$

The diverse priors facilitate the application of Bayesian model averaging across various model configurations, significantly benefiting automated forecasting systems. In situations where predictions need to follow certain rules (like staying within a set range or adding up to a specific number), helpful priors or adjustments can be added to the simulation process to make sure the results are useful ([18, 19] . The priors were chosen to be non-informative and that the results are therefore not expected to be sensitive to minor changes in their specification.

2.2.2. Likelihood Function Within a Bayesian ARIMA paradigm, the likelihood function indicates the probability of observing the data $Z = (Z_1, Z_2, \ldots, Z_n)$ contingent upon the model parameters. To obtain a stationary series using an ARIMA (p, d, q) model, one must difference the observed time series d times $Z_t = (1 - B)^d \gamma_t$. The altered series can be represented as an ARMA(p, q) process:

$$\phi(B) Z_t = \theta(B) a_t, a_t \square N\left(0, \sigma^2\right) \tag{9}$$

Assuming the innovations are conditionally normal $\theta = (\phi, \theta, \sigma^2)$ the probability of the model parameters is articulated as:

$$L\left(\theta_0, \theta, \phi, \sigma^2 \mid Z\right) \propto \left(\frac{1}{\sigma^2}\right)^{-\frac{T-d-p}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=p+1}^{T-d} \left[Z_t - \theta_0 - \sum_{j=1}^{q} \theta_j Z_{t-j} - \sum_{i=1}^{p} \phi_i a_{t-i}\right]^2\right\} \tag{10}$$

This formulation regards the ARIMA model as a conditional regression issue, wherein Z_t it is forecasted based on its historical values and prior residuals. The ARMA process is recursive and reliant on lagged values a_t ; hence,

Kalman filtering, state-space representations, or conditional likelihood approximations are necessary for precise likelihood assessment. The likelihood is crucial for determining the posterior distribution, as it indicates the extent to which the model parameters account for the observed data. Bayes' theorem integrates the prior distribution with the likelihood of deriving the posterior in Bayesian ARIMA models. In practical applications, beginning values (e.g.,$Z-1,\ldots,Z_q$) and residuals a_t are either regarded as constant or eliminated using data augmentation techniques. In Gibbs sampling, a_t the model parameters can be iteratively modified along with them [17, 18] .

2.2.3. Posterior Distribution The posterior distribution is essential in Bayesian ARIMA inference. The model parameters have been adjusted based on the observed data. Bayes' theorem posits that

$$p\left(\theta_0,\theta,\phi,\sigma^2 \mid Z\right) \propto L\left(\theta_0,\theta,\phi,\sigma^2 \mid Z\right) \cdot p\left(\theta_0,\theta,\phi,\sigma^2\right) \tag{11}$$

where:
• $L(\bullet)$ signifies the likelihood of the data.
• $p(\bullet)$ denotes the joint prior distribution of the parameters.
Due to the nonlinear and recursive nature of ARMA processes, it is often challenging to analytically modify the resulting posterior. We simulate and estimate the posterior distribution utilizing sampling techniques such as Metropolis-Hastings or Gibbs sampling. In contrast to other modeling methodologies, Bayesian modeling offers the complete distribution of each parameter rather than merely a point estimate. This enables the generation of posterior prediction intervals and credible intervals that illustrate the uncertainty of the parameters. Utilizing the data and the priors as a reference, the posterior distribution ϕ_i will illustrate the complete spectrum of potential AR coefficients. The posterior may incorporate model index variables if the model's structure is ambiguous (for instance, if p, d, and q are indeterminate). Consequently, the posterior probability for each ARIMA configuration may be computed. Bayesian model averaging, or BMA, is a technique employed in time series ensemble models and automated forecasting systems. It is founded on this ([18, 19, 20]. The final Bayesian ARIMA equation keeps the basic structure of the regular ARIMA model but adds uncertainty about the parameters through probabilistic inference.

$$\prod_{i=1}^{k}\left(1-\lambda_i L\right) y_t = \prod_{j=1}^{q}\left(1-\delta_j L\right) a_t \tag{12}$$

where
•λ_i the roots of the autoregressive polynomial.
• δ_j while the roots of the moving average polynomial. For the model to exhibit both stationarity and invertibility, the roots must be outside the unit circle. Bayesian estimations do not utilize fixed numbers. It determines the posterior distributions of parameters ϕ and θ , providing dependable ranges and simulations for future predictions. The Bayesian ARIMA approach generates multiple predictions rather than a single one. Samples $\{\theta^{(s)}\}_{s=1}^{s}$ from the posterior distribution are utilized to forecast subsequent observations:

$$p\left(y_{T+h} \mid y_{1:T}\right) = \int p\left(y_{T+h} \mid \theta\right) p\left(\theta \mid y_{1:T}\right) d\theta \approx \frac{1}{S}\sum_{s=1}^{S} p\left(y_{T+h} \mid \theta^{(s)}\right) \tag{13}$$

This mindset is highly beneficial for decision-making in ambiguous situations, such as financial planning, supply chain forecasting, or evaluating climate risk [18, 5] . The autoregressive (AR) model is a fundamental approach for analyzing time series data. Univariate time series data are effective for model development and forecasting. The classical AR(p) model posits that the current value of a time series is a linear function of its preceding p values, augmented by random noise. Conventional frequentist estimation methods, like maximum likelihood estimation and ordinary least squares, fail to consider uncertainty in the model or existing knowledge regarding the parameters. The Bayesian framework is advantageous as it provides a complete posterior distribution and incorporates prior knowledge on the model's parameters. Bayesian AR(p) models can be advantageous when uncertainty exists over the appropriate lag order, managing outliers, and imputing missing data [21].

It is necessary to establish prior distributions for the parameters $\phi=(\phi_1,\ldots,\phi_p)$ and σ_a^2 to conduct a Bayesian analysis. A prevalent option is:

- The standard prior for ϕ ,
$\phi \sim N(\mu_\phi, \sigma_\phi)$
- For σ_a^2 , employ an inverse-Gamma prior:
$\sigma_a^2.G.(a, b)$

In the absence of substantial prior knowledge, one may employ flat or non-informative priors [22] . The likelihood function for the AR(p) process given a sample $(Z_1, Z_2, \ldots, Z_r)$ is [21] :

$$L\left(\phi, \sigma^2 \mid Z_{1:\mathrm{T}}\right) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(Z_t - \sum_{i=1}^{p} \phi_i Z_{t-i}\right)^2}{2\sigma^2}\right) \tag{14}$$

Bayes' Theorem provides the joint posterior distribution of ϕ and σ_a^2 :

$$p\left(\phi, \sigma_a^2 \mid Z_{1:T}\right) \propto L\left(\phi, \sigma_a^2 \mid Z_{1:T}\right) \cdot p(\phi) \cdot p\left(\sigma_a^2\right) \tag{15}$$

Using conjugate priors, specifically a normal distribution for and an inverse-gamma distribution for σ_a^2 , leads to posterior distributions that belong to established families. In certain instances, they can be computed analytically, while in others, simulation may be employed to approximate the results [21, 22].

2.3. Markov Chain Monte Carlo (MCMC)

MCMC techniques are essential for Bayesian ARIMA modeling. They facilitate sampling from intricate posterior distributions when analytical integration is unfeasible. In Bayesian ARIMA, MCMC creates a Markov chain that eventually represents the posterior distribution of the model parameters based on the data.$p(\theta_0, \theta, \phi, \sigma_a^2|Z)$ The primary objective is to generate a collection of samples $\{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(S)}\}$ that closely resemble the actual posterior about their empirical distribution. Once the chain has converged during the "burn-in" phase, these samples are utilized to estimate quantities such as:
• "Posterior" denotes the back or rear position $E[\phi_i, Z]$.
• Credible intervals.
• Projected distributions.

Bayesian ARIMA employs two primary MCMC algorithms:
• The Metropolis-Hastings algorithm proposes a new value from a proposal distribution based on a specified probability and subsequently determines whether to accept or reject it.
• The Gibbs sampling method repeatedly selects samples from the entire conditional distributions of each parameter. MCMC is a crucial element of the ARIMA architecture due to the recursive error components of the likelihood, which render the posterior decomposition into simple forms unfeasible. MCMC allows for the direct incorporation of parameter constraints, such as stationarity and invertibility, into the sampling process by eliminating non-compliant draws. MCMC facilitates the comparison of Bayesian models by computing posterior model probabilities or Bayes factors. MCMC-based Bayesian ARIMA is an effective and adaptable forecasting instrument that accounts for both model order and parameter uncertainty. In practical applications like limited policy simulations or sporadic demand, MCMC enables the modeling of complete uncertainty distributions and the generation of forecasts that are probabilistically coherent [18, 19, 20].

2.3.1. Gibbs Sampling Bayesian ARIMA models utilize Gibbs sampling, an effective variant of Markov Chain Monte Carlo (MCMC), and particular kinds of prior knowledge. This method involves keeping the other values fixed while iteratively choosing each parameter based on its entire conditional distribution. It entails repeatedly selecting each parameter from its whole conditional distribution, under the assumption that the other values remain constant. The subsequent approach is employed to modify the Gibbs sampler for a Bayesian ARIMA model:

$$\phi^{(t+1)} \sim p\left(\phi \mid \theta^{(t)}, \sigma^{2(t)}, Z\right) \tag{16}$$

$$\theta^{(t+1)} \sim p\left(\theta \mid \phi^{(t+1)}, \sigma^{2(t)}, Z\right) \tag{17}$$

$$\sigma^{2(t+1)} \sim \mathrm{I}nverse - Gamma\left(\alpha', \beta'\right) \tag{18}$$

Direct sampling of the entire conditional distribution is feasible when a closed-form solution exists, as demonstrated with conjugate priors. Metropolis-Hastings steps are includedin Gibbs sampling to address challenging conditional distributions. This is occasionally referred to as Metropolis-within-Gibbs [18]. Chib and Greenberg (1994) demonstrate the application of Gibbs sampling in models with ARMA error mechanisms. The authors demonstrate the feasibility of employing numerical approaches to estimate or compute complete conditionals, especially in models incorporating moving average components. They provide detailed guidelines for estimating probabilities. This is the conventional method for conducting Bayesian time series modeling currently. Gibbs sampling offers several advantages:
• User-friendly
• Exhibits favorable convergence characteristics when appropriately configured
• Compatible with latent variable models and hierarchical priors. Initialization, chain autocorrelation, and convergence diagnostics must be conducted with meticulous attention. When parameters exhibit high similarity, reparameterization or block sampling techniques are occasionally employed to enhance the mixing of ARIMA models. Gibbs sampling facilitates comprehensive Bayesian analysis, encompassing thorough result summaries, predictions, and assessments of probability models [17, 18].

2.4. State Space (SS) model

Employing the AR model as a linear statistical framework to analyze nonlinear time series data will increase stochastic uncertainty, hence diminishing predicting accuracy and addressing the data's nonlinearity. The SS approach will be employed to address random uncertainty, enhancing forecasting accuracy owing to its effective performance in time series analysis. The SS technique can be articulated as a statistical approach for estimating and forecasting unmeasured state space equations. The SS approach integrates observations and current predicting values by employing weights that mitigate biases and mistakes. The state equation (SE) and the observation equation (OE) constitute a set of equations known as the linear model of the state space, which can be expressed as follows:

$$Z_t = AZ_{t-1} + Bu_{t-1} + e_{1,t} \tag{19}$$

$$Y_t = CZ_t + e_{2,t} \tag{20}$$

Z_t is the m-dimensional state vector, u_t is a specific input vector, Y_t is an output observation vector, $e_{1,t}$ and $e_{2,t}$ are independent white noise vectors; A, B, and C are constant matrices. The SE equation, denoted as equation (19), and the OE equation, denoted as equation (20), can be expressed in state space as follows:

$$Z_t = AZ_{t-1} + Bu_{t-1} + C'a_t \tag{21}$$

$$Y_t = CZ_t \tag{22}$$

$r = max(g, j)$
g denotes the count of lagged series of the variable Z_t, j signifies the count of lagged series of the residuals Z_t , is the state vector with dimension r, u_t is the lagged series vector of the residuals, and a_t is the rotated vector of the current residuals.
Equations (21) and (22) are complex in application; therefore, they can be restructured for simplicity as follows [23, 24].

$$Z_t = AZ_{t-1} + C'a_t \tag{23}$$

$$\hat{y}_t = CZ_t \tag{24}$$

$$a_t = Y_t - \hat{Y}_t \tag{25}$$

Z_t the m-dimensional state vector, A represents the state transition matrix of dimensions $(m \times m)$, while C signifies the display transition matrix of dimensions $(m \times 1)$. $\hat{y}_t$ the displayed rotation vectors illustrate the SS output series, where m indicates the number of lagged series terms Z_t . All lagged series of the residuals a_t are included on the right side of the AR data model equation after simplification, Z_t retaining only the left side.

$$Z_t = \begin{bmatrix} Z_{1,t} & Z_{2,t} & \dots & Z_{m,t} \end{bmatrix}' \tag{26}$$

$$A = \begin{bmatrix} P_1 & P_2 & P_3 & \cdots & P_m \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{m \times m} \tag{27}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}_{\mathrm{b} \times m} \tag{28}$$

$(P_1, P_2, P_3, \dots, P_m)$ After simplification, just the left half of the AR data model equation remains, and all parameter values of the lagged series and all time-lagged residual series are integrated on the right side. The input variables and the ESN architecture used in this investigation will be defined by the matrix A, the row vector C, and the variables in the AR data model equation.

2.5. Echo state network (ESN)

ESN is a highly efficient computational technique for time series forecasting because it is made to handle time-dependent data. The design is more accurate than conventional neural networks, has an easy-to-understand learning process, and has substantial processing power to solve nonlinear problems. Since its inception as a deep learning neural network, ESN has emerged as a preferred option for researchers in time series forecasting [23, 24]. The ESN technique uses a large number of randomly initialized recurrent neurons, called reservoirs, that act as a hidden layer. Training a recurrent neural network, such as the Echo State Network (ESN), is made easier by the randomly generated recurrent reservoir. Since (p) represents the dimension of the input variables and (m) represents the number of hidden nodes in the Echo State Network (ESN), training the output weights is a key idea in ESN. One of the benefits of ESN over traditional neural networks, however, is that the input and reservoir weights are allocated at random and are not taught. ESN is widely used in many fields, such as time series forecasting. [25, 26, 27, 28]. ESN is a neural network developed through the linear regression method, allowing it to address the limitations of conventional neural network models. It demonstrates superior performance and an accelerated learning rate [29]. The following two equations articulate the modeling process of the Echo State Network (ESN):

$$Z_t = f\left(W_z Z_{t-1} + W_{in} X_t\right) \tag{29}$$

$$\hat{y}_t = W_{out} Z_t \tag{30}$$

Equations (29) and (30) are derived from the state space model tailored for time series analysis, based on the SE equation (23) and the OE equation (24), incorporating a nonlinear function such as the tangent function.
In this context, t denotes the number of time steps, X_t while the rotated input vector is represented with dimensions $(1 \times t)$ and corresponds to the series of residues in the SS model as outlined in equation (23). Additionally a_t, Z_t the rotated matrix of internal states has dimensions $(m \times t)$ and is situated in the hidden layer, and finally, $\hat{y}_t$ the output vector has dimensions (t1 x) for the neural network. The output weights may be determined through the general inverse method utilizing the subsequent formula:

$$W_{\text{out}} = \hat{y}_t' Z_t^+ = \hat{y}_t' \left(Z_t' Z_t\right)^{-1} Z_t' \tag{31}$$

$$W_z = A, W_{\text{in}} = C', W_{\text{out}} = C$$

$$z = W_z Z_{t-1} + W_{\text{in}} X_t$$

$f(.)$ this research employs the nonlinear tangent function, represented by the following formula:

$$f(z) = \tanh(z) = \frac{1 - e^{-2z}}{1 + e^{-2z}} \tag{32}$$

If only the Z_t 's lagged series which represent the variables of AR model regardless of the parameters and signs are used to determine the inputs structure of the Artificial neural network (ANN) generally or ESN as in this study, this method can also be called hybrid AR-ESN model [30, 31, 32] or it can just be ESN [33, 34, 35]. Using an AR model instead of ARIMA model, regardless of several time series conditions, resulted in a simple input structure for the ESN according to [31, 36, 37]. In the first stage, the AR model will be constructed. An AR(p) model was used to determine the inputs structure of the ESN. The inputs structure of the ESN depended on the number of autoregressive parameters. ACF and PACF are used directly for the original time series data and for the first difference series to provide an idea about the autoregressive order. AR(p) model is a function of linear components that can be written as follows.

2.6. Time Stratification (TS)

An efficient technique for analyzing time series data is time stratification (TS). It makes it easier to arrange data according to seasonal variations, which are important factors that affect time series dynamics and forecasting results. By classifying data according to particular seasonal time strata, this approach aims to provide a deeper understanding of how seasons or cyclical temporal patterns affect the variable being studied. When time series exhibit consistent seasonal trends under similar correlation and influence conditions, time series analysis can be effectively applied to a diversity of time series. By classifying data into more uniform categories, this method reduces random volatility and boosts the accuracy of statistical and predictive modeling. By separating recurring seasonal patterns from general behavior, this technique improves forecast model performance and produces more reliable results than those obtained from aggregate, unclassified data. [38, 39].

2.7. The measurement of forecasting errors

The average of the absolute discrepancies between expected and actual values is known as the mean absolute error, or MAE. ([11]. The subsequent formula is employed to ascertain it:

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |y_t - \hat{y}_t| \tag{33}$$

- y_t : the actual value.
- $\hat{y}_t$: the anticipated value.
- n: the quantity of samples (or observations).

Root Mean Square Error (RMSE) quantifies the accuracy of a data forecast. This is the notation for RMSE [11] and [40]:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} a_t^2}{n}} \tag{34}$$

$t = 1, 2, ..., n;\ a_t = y_t - \hat{y}_t$, where n represents the number of observations, y_t denotes the true value at time $\hat{y}_t$ signifies the predicted value of the observation at time t, and a_t indicates the error series at time$t$.

3. Results and discussion

3.1. Data Description and Study Framework

This study establishes a dual forecasting framework that integrates nonlinear modeling with the Bayesian AR(p)-ESN hybrid model, employing the Echo State Network (ESN) inside a Bayesian analytical context, alongside probabilistic linear modeling with the Bayesian AR(p)-ESN model. The daily maximum wind speed in Mosul is a complex climatic time series that this combination aims to predict with greater accuracy. The agricultural weather station located in Nineveh Governorate, at coordinates 43.16°E and 36.33°N, supplied the data. Given the significant temporal fluctuations and evident seasonal variations in these data, analytical models capable of capturing both linear and nonlinear aspects of the time series are required. A time-stratified (TS) methodology was employed to establish seasonal homogeneity in the data, dividing the series into two discrete seasons: the cold season (November to March) and the hot season (May to September), covering the years 2012 to 2023. To ensure the robustness of the probabilistic models used, sufficient training and testing observations were employed. Consequently, the seasonality of the data and associated models will be presumed, with each seasonal cycle comprising five months (s = 5). Initially, non-informative prior distributions were employed to develop Bayesian AR(p) models, and Markov Chain Monte Carlo (MCMC) techniques were utilized to estimate the dimensional distribution of the coefficients and error variance. Modeling was conducted over several lags (2, 3, 4, and 5) to determine the optimal series structure for each season. These models provide a comprehensive probabilistic framework that facilitates uncertainty analysis and the establishment of confidence intervals for future projections, rather than solely depending on traditional point estimates. In the subsequent phase, the resonant state network was constructed utilizing the outputs of the Bayesian AR(p) model. The input matrix Z is created by combining the lagged time variables and residuals from the Bayesian model after the initial random error has been removed. This matrix is entered into the ESN using a stochastic internal structure (matrix A) and an output vector (C), which is intended to improve the model's forecasts. The hybrid Bayesian AR(p)-ESN model combines Bayesian inference, which accurately captures the characteristics of time series, with a neural network architecture that can capture complex interactions between variables and nonlinear patterns. Two implementations of this model were carried out: one that uses only the variables and not the AR coefficients, and another that includes the AR coefficients estimated using the Bayesian model, along with their values and signs. This combination of flexible neural networks and a strict Bayesian approach is more predictive and ideal for unpredictable climatic data because of the rapid changes in the environment. Preliminary comparative data indicate that the hybrid model exhibits superior prediction accuracy and resilience, especially when evaluated using metrics such as MAPE and MSE. General Study Framework: The subsequent steps provide the overarching framework of the study:

1. We divide the climate data into two seasons, a warm season and a cold season, using the Time Stratified (TS) method to achieve seasonal homogeneity in the time series attributes.
2. Dividing the season-specific data into two subsets: training data for model development and testing data for evaluating predictive model efficacy. The training period data is analyzed using a Bayesian AR(p) model over several lags (2, 3, 4, 5). The parameters of each model are estimated via MCMC techniques and dimensional inference.
3. We utilize the outputs of the Bayesian AR(p) model to build the input matrix Z, which includes the residuals and lagged temporal variables, while eliminating the initial random error. The network relies on an output vector (C) and an internal stochastic state matrix (A). The methodology is referred to as the hybrid Bayesian AR(p)-ESN model.
4. The study's general framework can be articulated as depicted in Figure (1). Figures (2) and (3) illustrate the maximum wind speed variable for both the hot and cold seasons during the training and testing periods.

3.2. Bayesian Autoregressive Model (BAR)

The initial step in this framework is to ascertain the configuration of the Bayesian AR(p) model. The partial autocorrelation function (PACF) of the original series is employed to determine the optimal rank of the model, indicating the number of prior time periods that exert a statistically significant influence. Currently, we do not

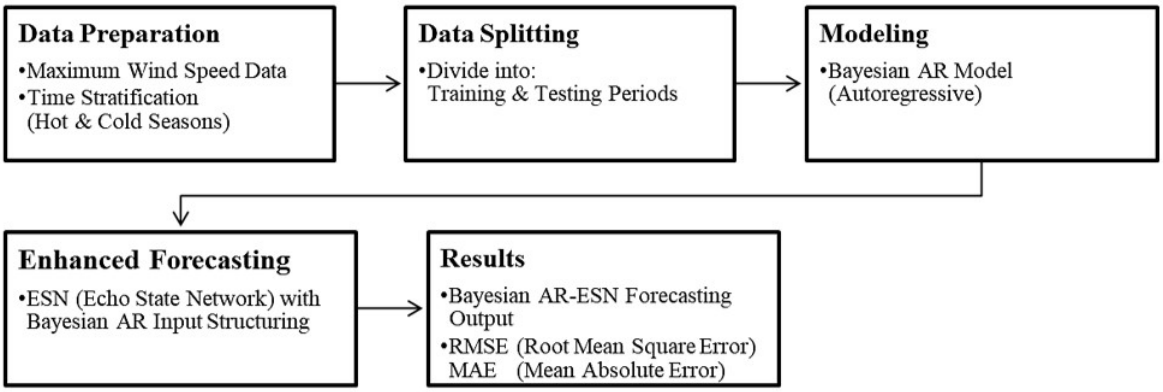

Figure 1. General framework for the research

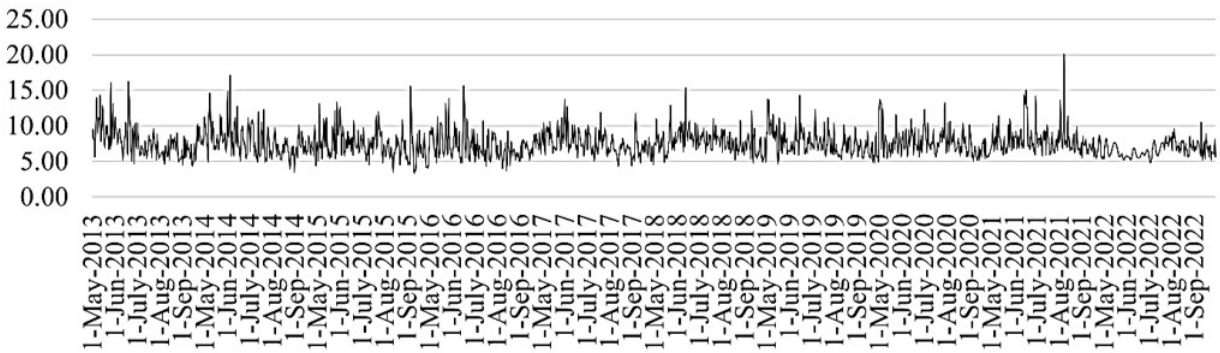

Figure 2. time series plot of the maximum wind speed of hot season

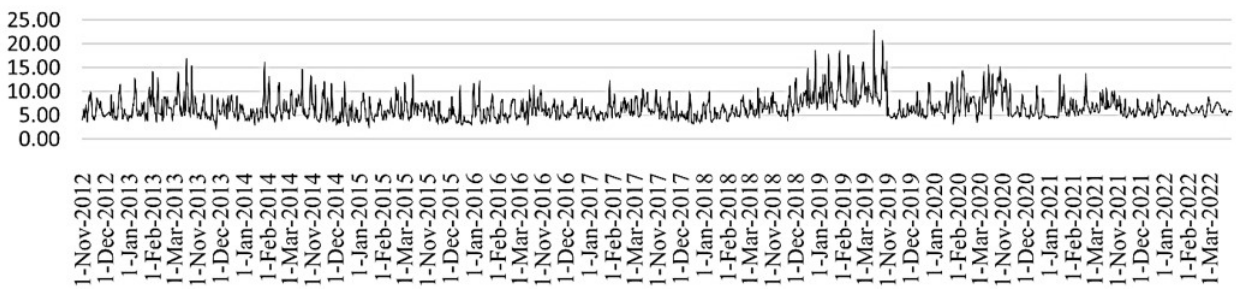

Figure 3. time series plot of the maximum wind speed of cold season

discuss stationarity criteria, as the primary objective of employing the Bayesian AR(p) model is to ascertain the characteristics of the input layer of the ESN, rather than to generate predictions. Figure (4) below illustrates the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) for the maximum wind speed data for the summer and winter seasons.

Figure (4) demonstrates that the ranks (p=2 till p=5) are the optimal Bayesian AR(p) models. These models can be applied to the data, despite Figures 2 and 3, which do not display any evident non-stationarity. These models can be expressed numerically using Equation (3), as demonstrated below.

$$Z_t = 0.537912 Z_{t-1} + 0.01947 Z_{t-2} + a_t \tag{35}$$

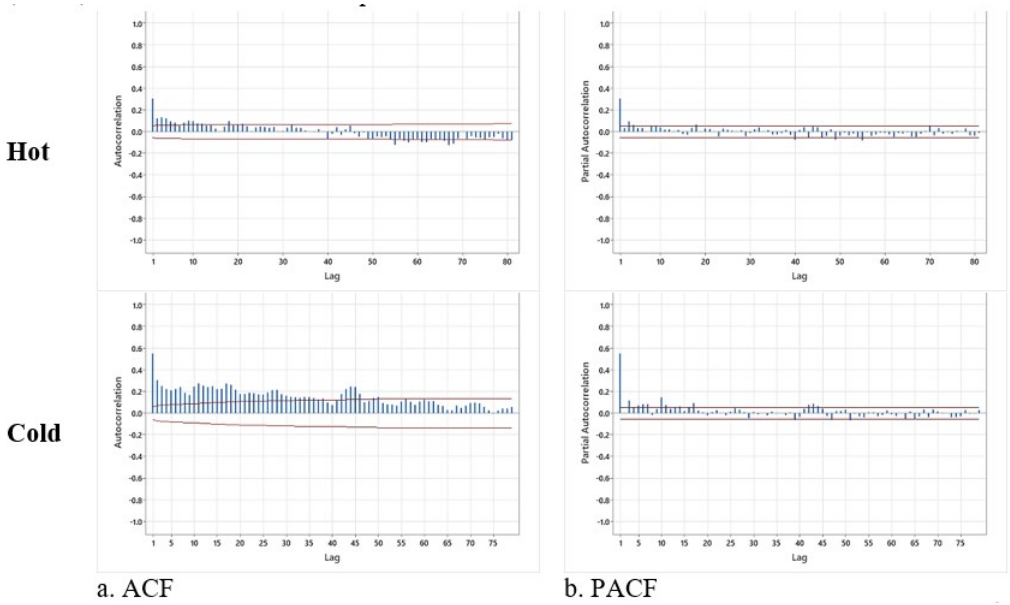

Figure 4. ACF and PACF of the maximum wind speed data for the hot and cold seasons.

$$Z_t = 0.535701Z_{t-1} - 0.03666Z_{t-2} + 0.104532Z_{t-3} + a_t \tag{36}$$

$$Z_t = 0.529889Z_{t-1} - 0.03438Z_{t-2} + 0.075616Z_{t-3} + 0.052781Z_{t-4} + a_t \tag{37}$$

$$Z_t = 0.525834Z_{t-1} - 0.042Z_{t-2} + 0.079266Z_{t-3} + 0.007439Z_{t-4} + 0.0852Z_{t-5} + a_t \tag{38}$$

$$Z_t = 0.300404Z_{t-1} + 0.021043Z_{t-2} + a_t \tag{39}$$

$$Z_t = 0.297208Z_{t-1} - 0.01525Z_{t-2} + 0.114033Z_{t-3} + a_t \tag{40}$$

$$Z_t = 0.290876Z_{t-1} - 0.01522Z_{t-2} + 0.099838Z_{t-3} + 0.046963Z_{t-4} + a_t \tag{41}$$

$$Z_t = 0.290581Z_{t-1} - 0.01879Z_{t-2} + 0.099144Z_{t-3} + 0.038189Z_{t-4} + 0.029118Z_{t-5} + a_t \tag{42}$$

3.3. State Space (SS) model Based on Bayesian AR(p)

We employed Bayesian AR(p) models to determine the maximum wind speed in both the hot and cold seasons. Subsequently, we employed these models to construct a State Space (SS) framework that distinctly illustrates the temporal variations of the time series. The Bayesian equations employed to estimate the SS model originate from Bayesian AR(p). The model is transformed into a format (23) and (24) that encompasses:

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} &= \begin{bmatrix} 0.537912 & 0.01947 \\ 1 & 0 \end{bmatrix}_{2\times 2} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} a_t \\ Y_t &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}'_{1\times 2} \times \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} \end{aligned} \tag{43}$$

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \end{bmatrix} &= \begin{bmatrix} 0.535701 & -0.03666 & 0.104532 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}_{3\times 3} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \\ Z_{3,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} a_t \\ Y_t &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}'_{1\times 3} \times \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \end{bmatrix} \end{aligned} \tag{44}$$

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \end{bmatrix} &= \begin{bmatrix} 0.529889 & -0.03438 & 0.075616 & 0.052781 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}_{4\times 4} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \\ Z_{3,t-1} \\ Z_{4,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} a_t \\ Y_t &= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}'_{1\times 4} \times \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \end{bmatrix} \end{aligned} \tag{45}$$

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \\ Z_{5,t} \end{bmatrix} &= \begin{bmatrix} 0.525834 & -0.042 & 0.079266 & 0.007439 & 0.0852 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}_{5\times 5} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \\ Z_{3,t-1} \\ Z_{4,t-1} \\ Z_{5,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ Y_t &= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}'_{1\times 5} \times \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \\ Z_{5,t} \end{bmatrix} \end{aligned} \tag{46}$$

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} &= \begin{bmatrix} 0.300404 & 0.021043 \\ 1 & 0 \end{bmatrix}_{2\times 2} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} a_t \\ Y_t &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}'_{1\times 2} \times \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} \end{aligned} \tag{47}$$

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \end{bmatrix} &= \begin{bmatrix} 0.297208 & -0.01525 & 0.114033 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}_{3\times 3} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \\ Z_{3,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} a_t \\ Y_t &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}'_{1\times 3} \times \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \end{bmatrix} \end{aligned} \tag{48}$$

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \end{bmatrix} &= \begin{bmatrix} 0.290876 & -0.01522 & 0.099838 & 0.046963 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}_{4\times 4} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \\ Z_{3,t-1} \\ Z_{4,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} a_t \\ Y_t &= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}'_{1\times 4} \times \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \end{bmatrix} \end{aligned} \tag{49}$$

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \\ Z_{5,t} \end{bmatrix} = \begin{bmatrix} 0.290581 & -0.01879 & 0.099144 & 0.038189 & 0.029118 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}_{5\times 5} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \\ Z_{3,t-1} \\ Z_{4,t-1} \\ Z_{5,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} a_t$$
$$Y_t = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}'_{1\times 5} \times \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \\ Z_{5,t} \end{bmatrix} \tag{50}$$

The state equation: illustrates the temporal evolution of the system's state. The observation equation illustrates the relationship between real observations and the underlying conditions. This paradigm provides a dynamic representation that facilitates the management of the series' underlying temporal structure. Indicators such as the Mean Absolute Error (MAE) and Root mean square error (RMSE) presented in Table (1) and (2) can be utilized to evaluate the efficacy of a model. This image illustrates a phase in the construction of the hybrid ESN model, subsequently employed to determine the network inputs.

Table 1. MAE values of different BAR models

	Lags 2		Lags 3		Lags 4		Lags 5	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Cold	1.0018	1.6009	1.0046	1.5997	1.0075	1.5937	0.994	1.5915
Hot	1.1235	1.4015	1.1087	1.3927	1.1066	1.3932	1.1021	1.3928

Table 2. RMSE values of different BAR models

	Lags 2		Lags 3		Lags 4		Lags 5	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Cold	1.4116	2.1891	1.4134	2.179	1.4175	2.1747	1.4012	2.1682
Hot	1.6026	1.8498	1.5908	1.8355	1.5916	1.8338	1.5816	1.833

3.4. Hybrid Bayesian AR-ESN method

Method is a hybrid model that integrates the internal architecture of the Echo State Network (ESN) with the fundamental equations of the Bayesian AR(p) model. A hybrid model, termed the Hybrid Bayesian AR-ESN, was developed utilizing the internal architecture of the ESN, which is founded on the fundamental equations of the Bayesian AR(p) model. The lagged time variables and residuals from the Bayesian model were utilized to determine the network's input matrix Z. We applied the entire ESN methodology with all the details in Table (3) to the identical dataset for both seasons and utilized the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics to assess the outcomes, as illustrated in Table (4) and (5).

The ESN was initialized with random matrices U, V, and W. The Burgers' equation, gradient-based training was applied to the readout instead of the standard pseudo-inverse method.

Tables (1) and (2), (4), (5) indicate that the hybrid Bayesian AR-ESN model outperforms the model utilizing solely Bayesian AR. The hybrid model exhibited reduced the Mean Absolute Error (MAE) and Root mean square error (RMSE) values, enhancing its predictive accuracy. Utilizing the TS technique to segment maximum wind speed data by season and employing a hybrid model based on Bayesian analysis and the ESN network significantly

Table 3. Hyperparameters of the ESN.

Hyperparameter	**Symbol/Value**	**Description**
Reservoir size	5	Number of neurons in the reservoir.
Spectral radius	0.9	The largest absolute eigenvalue of the reservoir weight matrix U.
Input scaling	1	Scaling factor applied to the input weights V.
Readout regularization	0	Ridge regularization parameter applied to the output weight W for training, while Gradient descent (GD) used for the regularization.
Sequence length (for training)	5	Number of time steps per training subsequence.
Training epochs	1000	Number of GD iterations on the readout.
Learning rate	1	Learning rate used for training optimizer W.

Table 4. MAE values of different Bayesian AR-ESN models

	Lags 2		Lags 3		Lags 4		Lags 5	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Cold	0.9877	1.537	0.9639	1.5458	0.9943	1.5365	0.99774	1.52564
Hot	1.1338	1.363	1.0828	1.3742	1.1	1.3849	1.0723	1.361

Table 5. RMSE values of different Bayesian AR-ESN models

	Lags 2		Lags 3		Lags 4		Lags 5	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Cold	1.3986	2.091	1.3951	2.092	1.4379	2.0745	1.43221	2.06323
Hot	1.6151	1.7869	1.5698	1.8058	1.5828	1.8219	1.5705	1.7902

improved the forecasting results. This indicates that the proposed model effectively articulates complex temporal variations and generates predictions more accurately.

4. Conclusions

Compared to standard Bayesian models as Bayesian AR(p), the hybrid Bayesian AR-ESN model is a more accurate forecasting tool, particularly when examining complex and nonlinear climatic data as maximum wind speed series. This conclusion is supported by the data shown in the Results and Discussion section. The nonlinear adaptive features of the Echo State Network (ESN) are combined with Bayesian analysis, which offers an accurate probabilistic representation of temporal structure, the forecast accuracy is significantly improved, according to the comparison of the two models. Forecasting error measurement parameters like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were significantly reduced when the hybrid model was used. According to the study, Bayesian AR(p) estimates model parameters and represents linear temporal effects effectively in a probabilistic context; it falls short in capturing the complex correlations and nonlinear dynamics present in the time series. Therefore, it must be combined with ESN as a nonlinear layer that enhances the linear representation and increases the predictive power of the model. The suggested approach predicts univariate climatic time series quite well, particularly those with abrupt changes and seasonal volatility.This increases the method's suitability for use in upcoming real-world meteorological and renewable energy applications.

REFERENCES

1. G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, John Wiley and Sons, 2015
2. H. Kantz, and T. Schreiber, *Nonlinear time series analysis*, Cambridge University Press, 2003.
3. A. Gelman, *Bayesian Data Analysis, 3rd: Boca Raton*, Texts in Statistical Science, 2013.
4. C. P. Robert, *The Bayesian choice: from decision-theoretic foundations to computational implementation*, Springer, 2007.
5. M. A. Rahman, and B. R. Sarker, *Intermittent demand forecast and inventory reduction using Bayesian ARIMA approach*, in *Proceedings of the 2010 International Conference on Industrial Engineering and Operations Management*, 2010.
6. A. Oyelam, and R. Ogundeji, *Bayesian Seasonal autoregressive integrated moving average: Modelling two decades of inflation dynamics in Nigeria*, Science World Journal, vol.20, no.1, pp.56-63, 2024.
7. A. Singh, T. Tripathi, R. Ranjan, and A. K. Tiwari, *Time series forecasting of infant mortality rate in India using Bayesian ARIMA models*, BMC Public Health, vol.25, no.1, pp.2855, 2025.
8. H. Jaeger, *The "echo state" approach to analysing and training recurrent neural networks-with an erratum note*, German National Research Center for Information Technology, GMD Technical Report 148(34), pp.13, Bonn, Germany, 2001.
9. M. Lukoševičius, and H. Jaeger, *Reservoir computing approaches to recurrent neural network training*, Computer Science Review, vol.3, no.3, pp.127–149, 2009.
10. P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, 2002.
11. R. J. Hyndman, and G. Athanasopoulos, *Forecasting: principles and practice*, OTexts, 2018.
12. S. F. Crone, M. Hibon and K. Nikolopoulos, *The elements of statistical learning: data mining, inference, and prediction*, International journal of forecasting, vol. 27 no. 3, pp. 635-660 2011.
13. G. Zhang, B. E. Patuwo, and M. Y. Hu, *Forecasting with artificial neural networks: The state of the art*, International Journal of Forecasting, vol.14, no.1, pp.35–62, 1998.
14. J. Durbin and S. J. Koopman, *Time series analysis by state space methods*, Oxford university press, 2012.
15. S. Løkse, F. M. Bianchi, and R. Jenssen, *Training echo state networks with regularization through dimensionality reduction*, Cognitive Computation, vol.9, no.3, pp.364–378, 2017.
16. J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, *Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach*, Physical Review Letters, vol.120, no.2, 024102, 2018.
17. M. Agarwal, P. K. Tripathi and S. Pareek, *Forecasting infant mortality rate of India using ARIMA model: A comparison of Bayesian and classical approaches*, Stat Appl, vol. 19 no. 2,101–114, 2021.
18. S. Chib and E. Greenberg, 64(1-2), 183-206. *Bayes inference in regression models with ARMA (p, q) errors*, Journal of Econometrics, vol.64, no. 1-2, pp. 183–206, 1994.
19. R. S. Ehlers and S. P. Brooks, *Bayesian analysis of order uncertainty in ARIMA models*, University of Parana, Tech. Rep, 2004.
20. W. Uturbey, *Identification of ARMA models by Bayesian methods applied to streamflow data*, in *Proceedings of the 2006 International Conference on Probabilistic Methods Applied to Power Systems*, 2006.
21. G. Barnett, R. Kohn and S Sheather, *Bayesian estimation of an autoregressive model using Markov chain Monte Carlo*, Journal of Econometrics, vol. 74, no. 2, pp. 237–254, 1996.
22. J.-M. Marin, and C. P. Robert, *Bayesian essentials with R*, Springer, vol.48, 2014.
23. P. G. Gould, A. B. Koehler, J. K. Ord, R. D. Snyder, R. J. Hyndman, and F. Vahid-Araghi, *Forecasting time series with multiple seasonal patterns*, European Journal of Operational Research, vol.191, no.1, pp.207–222, 2008.
24. H. Madsen, *Time Series Analysis*, CRC Press, 2007.
25. M. Han, and M. Xu, *Laplacian echo state network for multivariate time series prediction*, IEEE Transactions on Neural Networks and Learning Systems, vol.29, no.1, pp.238–244, 2017.
26. Q. Ma, E. Chen, Z. Lin, J. Yan, Z. Yu, and W. W. Ng, *Convolutional multitimescale echo state network*, IEEE Transactions on Cybernetics, vol.51, no.3, pp.1613–1625, 2019.
27. J. Qiao, F. Li, H. Han, and W. Li, *Growing echo-state network with multiple subreservoirs*, IEEE Transactions on Neural Networks and Learning Systems, vol.28, no.2, pp.391–404, 2016.
28. C. Yang, J. Qiao, L. Wang, and X. Zhu, *Dynamical regularized echo state network for time series prediction*, Neural Computing and Applications, vol.31, pp.6781–6794, 2019.
29. H. Duan and X. Wang, *Echo state networks with orthogonal pigeon-inspired optimization for image restoration*, IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 11, pp. 2413-2425 2015.
30. M. Khashei, and M. Bijari, *A novel hybridization of artificial neural networks and ARIMA models for time series forecasting*, Applied Soft Computing, vol.11, pp.2664–2675, 2011.
31. H. Liu, H.-q. Tian, and Y.-f. Li, *Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction*, Applied Energy, vol.98, pp.415–424, 2012.
32. O. B. Shukur, and M. H. Lee, *Daily Wind Speed Forecasting Through Hybrid AR-ANN and AR-KF Models*, Jurnal Teknologi, vol.72, no.5, pp.89–95, 2015.
33. M. Khashei, and M. Bijari, *An artificial neural network (p,d,q) model for time series forecasting*, Expert Systems with Applications, vol.37, no.1, pp.479–489, 2010.
34. O. B. Shukur, N. S. Fadhil, M. H. Lee, and M. H. Ahmad, *Electricity Load Forecasting using Hybrid of Multiplicative Double Seasonal Exponential Smoothing Model with Artificial Neural Network*, Jurnal Teknologi, vol.69, no.2, pp.65–70, 2014.
35. G. P. Zhang, *Time series forecasting using a hybrid ARIMA and neural network model*, Neurocomputing, vol.50, pp.159–175, 2003.
36. K. Chen and J. Yu, *Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach*, Appl. Energy, vol. 113, pp. 690-705, 2014.
37. S. Tatinati, and K. C. Veluvolu, *A hybrid approach for short-term forecasting of wind speed*, Scientific World Journal, vol.2013, doi:10.1155/8086, 2013.

38. B. J. Malig, D. L. Pearson, Y. B. Chang, R. Broadwin, R. Basu, R. S. Green, and B. Ostro, *A time-stratified case-crossover study of ambient ozone exposure and emergency department visits for specific respiratory diagnoses in California (2005–2008)*, Environmental Health Perspectives, vol.124, no.6, pp.745–753, 2016.
39. A. Tobias, B. Armstrong, and A. Gasparrini, *Analysis of time-stratified case-crossover studies in environmental epidemiology using Stata*, in *United Kingdom Stata Users' Group Meetings*, 2014.
40. R. J. Hyndman, and A. B. Koehler, *Another look at measures of forecast accuracy*, International Journal of Forecasting, vol.22, no.4, pp.679–688, 2006.