STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–17.
Published online in International Academic Press (www.IAPress.org)

Using Bayesian Ridge Regression model and ESN for Climatic Time Series Forecasting

Raed Arif Snaan, Osamah Basheer Shukur*

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

Abstract The analysis of climatic time series variables, especially evaporation forecasts influenced by diverse climate components, is essential for mitigating the hazards associated with climate change and its effects on environmental phenomena, subsequently affecting human and plant health. The weather patterns and temporal conditions will be evaluated during a restricted time frame. Numerous climate variables, including temperature and humidity, demonstrate strong connections, resulting in multicollinearity that undermines the efficacy of conventional linear models and induces instability and considerable variability in model parameters. Moreover, climatic data frequently display erratic variations owing to their dependence on several sources, including sensors and satellites. This results in a nonlinear pattern, leading to considerable geographical and temporal fluctuation, hence complicating modelling efforts with conventional methods. Therefore, there is a necessity for statistical models that can address these issues and systematically manage residuals and uncertainties, which adversely affect the precision of time series forecasting. This study utilized the Bayesian Ridge Regression (BRR) model. This Bayesian adaptation of conventional ridge regression considers model parameters as random variables instead of constants, thereby diminishing estimate bias and enhancing stability in the presence of multicollinearity. The model offers a probabilistic representation of outputs, facilitating confidence intervals for forecasting and improving the reliability of results. Moreover, climatic time series forecasting is influenced by their chaotic and nonlinear characteristics. This is the point at which the echo state network (ESN) is relevant. This specific sort of recurrent neural network excels at forecasting nonlinear time series due to its proficiency in managing temporal dynamics and nonlinear modelling. This study integrated and hybridized the BRR model within the ESN architecture to utilize its overfitting mitigation capabilities and structural attributes for enhanced forecast accuracy. Experimental findings indicated that the BRR-ESN hybrid model markedly surpassed traditional models in multivariate time series forecasting, validating its efficacy in addressing the structural and climatic intricacies of evaporation data and the associated climate variables. .

Keywords Bayesian Ridge Regression, Echo State Network, Multicollinearity, Climatic Time Series Forecasting, Hybrid BRR-ESN model

DOI: 10.19139/soic-2310-5070-2637

1. Introduction

Climate studies enhance the comprehension and analysis of alterations in evaporation rates and their effects on diverse climatic aspects and variables. This research will analyze weather patterns and circumstances across the period from 2012 to 2022 to comprehend the characteristics of the climate and its alterations. Climate data were obtained from the agricultural meteorology center at the Mosul Station in Nineveh governorate in northern Iraq., situated at a longitude of 43.16°E and a latitude of 36.33°N. Bayesian Ridge Regression (BRR) and Echo State Networks (ESN) models were utilized to project future climate alterations and examine their effects on the evaporation time series. The utilized database encompassed fundamental climate variables, such as evaporation rate, maximum and minimum temperatures along with their averages, maximum and minimum relative humidity

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

ISSN 2310-5070 (online) ISSN 2311-004X (print)

and their averages, solar radiation, wind direction, average wind speed, and the peak recorded wind speed. Bayesian regression (BRR) is a robust statistical model for forecasting, particularly in scenarios involving several variables and their inter-correlations. It integrates linear regression with Bayesian optimization methodologies to mitigate overfitting .BRR is extensively utilized in the presence of defective or noisy data, rendering it an appropriate option for constructing more stable and precise forecasting models [1] . In the domain of time series forecasting, researchers Kim and King [2] employed Deep ESN as a sophisticated method for addressing extensive, multidimensional, nonlinear time series. Artificial neural networks have demonstrated efficacy in modelling this data type. Researchers Viehweg, Worthmann [3] did a study on the critical relevance of methodically identifying hyperparameter in Echo State Networks (ESN), due to their substantial influence on enhancing the effectiveness of forecasting random time series, including climate time series. In a related context, researchers Peng, Lei [4]devised a methodology for the analysis of time series components by using ESN but with ARIMA model instead of BRR. The BRR model was employed with wavelength variation to examine seasonal time series, while the ESN was utilized to assess non-seasonal components, enhancing forecast accuracy. While the fundamental concept aligns with the approach of this research, the execution and specifics diverge from prior studies. This research seeks to quantify the influence of several climatic parameters on evaporation, identify the existence of multicollinearity among explanatory variables, and evaluate its effect on the precision of regression models. To mitigate the issue of multicollinearity, which results in the interdependence of explanatory variables and subsequently biased estimates due to the intricacies of climatic data and their high correlations, the Bayesian Ridge Regression model was employed as a robust statistical method to resolve this issue, enhance the stability of regression coefficients, and augment the efficacy of forecasted modelling. An Echo State Network was utilized to improve model performance and get precise evaporation forecasts based on the attributes of the analyzed meteorological data. The integration of the BRR and ESN methodologies into a hybrid model (BRR-ESN) constitutes a cohesive framework aimed at enhancing the precision of estimation and forecasting, hence facilitating a more profound comprehension of the variables influencing evaporation in the context of climate change.

2. Methods and Materials

2.1. Multiple linear Regression (MLR)

Linear models are extensively employed throughout diverse scientific disciplines. Multiple linear regression is one of the most often employed linear models for analyzing data in medical, economic, and other applied science research. numerous linear regression is a statistical model employed to forecast values, ascertain the relationship between a response variable and numerous explanatory variables, and illustrate the correlation among these explanatory variables and the response variable. It can be regarded as an extension of simple linear regression; nevertheless, it is more intricate as it involves several explanatory variables. Parameters are calculated by the ordinary least squares (OLS) method, and the model's quality can be evaluated using the coefficient of determination.[5, 6].

Multiple linear regression consists of m explanatory variables, where each explanatory variable contains n observations [6], as follows:

$$\begin{array}{ccccc} y_i & x_{i1} & x_{i2} & \cdots & x_{im} \\ y_1 & x_{11} & x_{12} & \cdots & x_{1m} \\ y_2 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_n & x_{n1} & x_{n2} & \cdots & x_{nm} \end{array}$$

where:
$i : 1, 2, \ldots, n$ Number of observations.

$j : 1, 2, \ldots, m$ Number of variables.

Create a mathematical model that allows forecasting the explanatory values of the response variable (y) based on the values entered for the explanatory variables $(x_1, x_2, \ldots, x_n)$ in the multiple regression analysis as follows in Equation 1.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{1}$$

where:

y: Response variable.

X_i: Explanatory variables that affect the response variable.

β_i: Parameters that express the effect of each explanatory variable on the response variable.

ϵ: Residuals is the difference between the actual values and the values forecasting by the model.

Multiple Linear Regression Assumptions [6]:

1. The response variable is normally distributed with a mean μ_i and variance σ^2 , i.e., the:
$y_i \sim N(\mu_i, \sigma^2)$; $i = 1, 2, \ldots, n$;

2. Linear model, i.e., the relationship between the response variable and the explanatory variables is linear. We can express the linear model as follows in Equation 2.

$$y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i \tag{2}$$

$i = 1, 2, \ldots, n$

where:

$\beta_0, \beta_1, \beta_2 \ldots, \beta_p$: Regression parameters.

p: Number of independent variables.

ϵ_i : Residuals. We will assume a zero mean and a constant variance. These are errors that are independent of each other and are expressed mathematically as:

$\epsilon \sim N(0, \sigma^2 I)$
$cov(\epsilon_i, \epsilon_j) = 0$

Regression parameters β are calculated using the Ordinary Least Square (OLS) method, which attempts to find values that minimize the error between the actual and forecasted values such as in Equation 3 [6].

$$\widehat{\beta} = (X'X)^{-1} X'y \tag{3}$$

where:

$\widehat{\beta}$: The vector of unknown parameters with dimensions $(M \times 1)$.

y: The vector of observations of the estimated response variable with dimensions $(P \times 1)$.

X: The matrix of observations of the explanatory variables with dimensions $(P \times M)$.

The multiple linear regression model is forecasted on the mathematical assumption of a linear connection between each explanatory variable and the response variable. For this model to function effectively, it is essential to assume that there is no substantial connection among the many explanatory factors, as we are assessing their impact on the response variable rather than on one another. The existence of a perfect or imperfect linear correlation among two or more explanatory variables in a multiple regression model causes multicollinearity, resulting in unstable estimates. To resolve this issue, it is essential to determine the variables contributing to it. To identify the issue of multicollinearity within a model, it is essential to compute the following:

Variance Inflation Factor (VIF)

The variance inflation factor (VIF) measure is an important method for detecting multicollinearity. It measures the variance inflation of the estimated parameters when there is a linear relationship between the explanatory variables in the model. When the explanatory variables are orthogonal, i.e., there is no correlation between them, this means $(VIF = 1)$ that the value is greater than one, indicating a linear relationship between those variables. If this value is greater than one $(VIF_j > 10 \; where \; j = 1, 2, \ldots, n)$, it indicates a multicollinearity problem. This can be determined using the coefficient of determination. The following formula is used to find VIF the values such as in Equation 4 [6]:

$$VIF = \frac{1}{(1 - R_j^2)} \tag{4}$$

$j = 1, 2, \ldots, n$

where:
n: Number of explanatory variables.

R_j^2: Coefficient of determination of the explanatory variable X_j extracted from a regression x_j on the remaining explanatory variables in the model.

The traditional least squares (OLS) method is used to estimate a multiple linear regression model as follows in Equation 5.

$$\widehat{y} = X\beta + \epsilon \tag{5}$$

where:
$\widehat{y}$: Vector of observations of the estimated response variable with dimensions $(P \times 1)$.

X : Matrix of observations of the explanatory variables with dimensions $(P \times M)$.

ϵ : A vector of residuals with dimensions $(P \times 1)$.

Assuming that $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2 I$

β : A vector of unknown parameters with dimensions $(M \times 1)$, the OLS method is used according to the following formula in Equation 6.

$$\widehat{\beta} = (X'X)^{-1}X'y \tag{6}$$

In the case of a multicollinearity problem, the estimators of the linear model parameters $(\widehat{\beta})$ and its variance will not be representative of the problem under study. Therefore, it must be addressed using the Ridge Regression Method. The multicollinearity problem occurs when the relationship between the explanatory variables is strongly correlated, and RR estimators address the problem. When the variance of the estimated parameters is large, the RR method is used, and multicollinearity is almost complete.

2.2. Ridge Regression (RR)

In a linear regression model with numerous connected variables, the coefficients may become ill-defined and display significant variance [7]. Multicollinearity arises in multiple linear regression models when there exists a significant connection among the explanatory variables. This association results in heightened variance in the estimations of the regression parameters, yielding unstable estimates and substantial standard errors. The estimations of the linear model parameters and their variance will be unrepresentative of the phenomenon being examined, rendering the conclusions of the ordinary least squares technique erroneous. To resolve the issue of multicollinearity, it is essential to identify the variable or variables responsible for this complication. Numerous techniques have been suggested by various scholars, including the Ridge Regression method introduced by Horel and Kennard (1970) [8]. This strategy seeks to resolve this issue and achieve a technique that provides the most accurate estimates. The Ridge Regression approach is advantageous for simulation, as its estimates yield the minimal squared errors. The RR method elucidates that incorporating a minor positive value into the principal diagonal of the matrix $(X'X)$, diminishes the variance of the estimated parameters by mitigating some bias; however, the mean square error (MSE) of the RR estimators will be inferior to that of the conventional least squares method. The equation for the RR estimators is as follows in Equation 7.

$$\beta_{RR} = (X'X + kI)^{-1}X'y \tag{7}$$

where:

I : The identity matrix,$0 \leq k \leq 1$ and when the value is $k = 0$, the estimators $\widehat{\beta}$ will be converted to OLS method estimators. The above equation (7) shows us that the value of k is added to the main diagonal of the information matrix $(X'X)$. There are several methods for choosing the value of k. The two primary techniques for determining a Ridge parameter are the Ridge Trace method and the k-value estimation approach. The subsequent section elucidates the two methodologies [5]:

1- Ridge Trace Technique

The graph is utilized in the Ridge Trace method, wherein coordinates are displayed to get the optimal k values. The vertical axis denotes the estimated parameters of the RR method for any k value, with the bias factor progressively increasing from zero within the designated range, while the horizontal axis specifies k values between (0, 1). The researcher determines the suitable k value to ensure that all estimations remain constant as k increases, hence maintaining the stability of the function.

2-Two methods for estimating k: The Ridge Parameter estimate approach is neutral and unbiased, paralleling the Ridge Trace method, as the optimal Ridge Parameter value is contingent upon the values of the response variable. Determining the Ridge Regression (k) value substantially decreases variance while augmenting the squared bias value. When k exceeds zero, we acquire biased estimates deemed more stable. Consequently, we depend on the values of k at the stability point.

2.3. Bayesian Ridge Regression (BRR)

While RR regression mitigates multicollinearity by incorporating a minor positive value into the principal diagonal of the matrix$(X'X)$, it solely yields point estimates of the parameters β , lacking information regarding the confidence level or corresponding variance. Conversely, Bayesian ridge regression (BRR) is a Bayesian adaptation of conventional ridge regression that considers model parameters β as random variables instead of fixed constants. It is based on Bayesian theory, which integrates the data (conditional likelihood distribution) with a prior distribution, hence minimizing bias in parameter estimate. This is a linear regression model that employs probability theory to derive probability distributions for the parameters, utilizing prior and posterior information instead of ordinary least squares (OLS) for parameter estimation β. The primary objective of Bayesian analysis is to revise the probability distributions of parameters by integrating knowledge derived from observational data. The characteristics of the resultant distribution underpin estimation and inference. BRR offers a robust framework for estimating parameters with established uncertainty, autonomously determines the regularization value, and integrates prior knowledge, rendering it a potent and dependable instrument in linear regression analysis. It is not merely a regularization technique, but a comprehensive probabilistic model that offers integrated estimation of residuals and regularization parameters, while accommodating existing prior knowledge and systematically mitigating statistical risk. Consequently, BRR offers enhanced interpretive depth and statistical dependability in forecasting relative to conventional RR [1], [9],[10]. The mathematical formula for BRR is as follows in Equation 8.

$$P(\beta|y,X) = P(y|X,\beta,\sigma^2)P(\beta) \tag{8}$$

where:

$P(\beta|y,X)$: The posterior distribution.

$P(y|X,\beta,\sigma^2)$: The conditional distribution of the probability function.

$P(\beta)$: The initial distribution.

That is, Bayesian theory calculates the posterior distribution, which is the multiplication of the conditional distribution of the probability function by the initial distribution. Thus, parameters are derived from the posterior distribution, where the expected value provides an estimate of the parameters, and the posterior variance expresses the uncertainty in the estimates [11]. One of the main goals of Bayesian analysis is to calculate the posterior distribution, which is the distribution of the parameters updated using the data. It consists of the following quantities:

1. The conditional distribution of the probability function is found:

Since the random error is normally distributed with a mean of zero and variance $\sigma^2 I$, so $\epsilon \sim N(0,\sigma^2 I)$, the random variable follows a normal distribution, distributed as follows [10], [12], [13]:
$P(y|X,\beta,\sigma^2) \sim N(X\beta,\sigma^2 I)$

Therefore, the probability density function of the random variable y is as in Equation 9.

$$P(y|X,\beta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{\frac{-1}{2\sigma^2}(y-X\beta)'(y-X\beta)} \tag{9}$$

where:
$P(y|X,\beta,\sigma^2)$: Conditional distribution of the probability function

σ^2: Variance of errors

n: Number of observations

β: Vector of parameters

X: Matrix of observations of the explanatory variables

y: Vector of observations of the response variable

2- Finding the Prior Distribution.

In this section, an argument will be made for choosing a particular metric, since locally uniform priors can be considered uninformative regarding the parameters. It is important to keep in mind that one cannot be in a state of complete ignorance, and furthermore, the phrase ”knowing little in advance” can only have meaning in

relation to the information provided by the experiment. Therefore, the key issue is how to choose a prior that provides little information compared to what the experiment is expected to provide[14]. This model assumes that the parameters β have a Gaussian prior distribution of the form $\beta \sim N(0, \tau^2)$, This is consistent with the formula used in [9],where the initial distribution is expressed in terms of precision rather than variance, as follows:$\lambda = \frac{1}{\tau^2}$,The variance of the residuals in the likelihood function, symbolized by σ^2 , can also be expressed in terms of precision as: $\alpha = \frac{1}{\sigma^2}$. Thus, the two symbols α and λ used in some practical applications directly correspond to the theoretical parameters σ^2 and τ^2 used in the statistical model [9].
In Bayesian regression, a Gaussian prior distribution is assumed on the parameters β to control for complexity and avoid overfitting [12].

Therefore:
$\beta \sim N(0, \tau^2)$
where:
τ^2 : The variance of the initial distribution represents the variance in the initial distribution of parameters.
This assumption is used to control the values of the model parameters and avoid overfitting. That is, the initial distribution of the parameters β is a normal distribution with a mean of zero and variance τ^2. It can be expressed such as in Equation 10.

$$P(\beta) = \frac{1}{(2\pi\tau^2)^{\frac{p}{2}}} e^{-\frac{1}{2\tau^2}\beta'\beta} \tag{10}$$

where:
p: Number of parameters.
β : Vector of parameters.

3. Posterior Distribution:
It is the multiplication of the conditional probability distribution by the initial distribution to obtain the best Bayesian estimator for the parameters such as in Equation 11 [9].

$$P(\beta|y, X) = P(y|X, \beta, \sigma^2)P(\beta) \tag{11}$$

The log of both sides of the equation will be taken such as in Equation 12.

$$LogP(\beta|y, X) = LogP(y|X, \beta, \sigma^2)LogP(\beta) \tag{12}$$

Since $P(y|X, \beta, \sigma^2)$ and $P(\beta)$ the distribution is Gaussian ,so the posterior distribution is also Gaussian [9].

Then, the magnitude will be found $LogP(y|X, \beta, \sigma^2)$, followed by the magnitude $LogP(\beta)$, and then substituted into the mathematical formula for the posterior distribution to find the best Bayesian estimator.
Note that:

$$P(y|X, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{\frac{-1}{2\sigma^2}(y-X\beta)'(y-X\beta)} \tag{13}$$

After taking the log of both sides of the equation 13.

$$LogP(y|X, \beta, \sigma^2) = Log\left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{\frac{-1}{2\sigma^2}(y-X\beta)'(y-X\beta)}\right) \tag{14}$$

$$LogP(y|X, \beta, \sigma^2) = Log(2\pi\sigma^2)^{-\frac{n}{2}} + Loge^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)} \tag{15}$$

$$LogP(y|X, \beta, \sigma^2) = -\frac{n}{2}Log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \tag{16}$$

The second half is also found in the mathematical formula for the posterior distribution $LogP(\beta)$ After obtaining $P(\beta)$ and equal to as follows such as in Equation 17.

$$P(\beta) = \frac{1}{(2\pi\tau^2)^{\frac{p}{2}}} e^{-\frac{1}{2\tau^2}\beta'\beta} \tag{17}$$

In the next step, the Log of both sides of the equation will be taken for Equation 17.

$$LogP(\beta) = Log\left(\frac{1}{(2\pi\tau^2)^{\frac{p}{2}}} e^{-\frac{1}{2\tau^2}\beta'\beta}\right) \tag{18}$$

$$LogP(\beta) = Log(2\pi\tau^2)^{-\frac{p}{2}} + Loge^{-\frac{1}{2\tau^2}\beta'\beta} \tag{19}$$

$$LogP(\beta) = -\frac{p}{2}Log(2\pi\tau^2) - \frac{1}{2\tau^2}\beta'\beta \tag{20}$$

The value of $LogP(\beta)$ and the value of $LogP(y|X,\beta,\sigma^2)$ will be substituted in the mathematical formula for the posterior distribution, which is as follows in Equation 21.

$$LogP(\beta|y,X) = LogP(y|X,\beta,\sigma^2)LogP(\beta) \tag{21}$$

$$LogP(\beta|y,X) = -\frac{n}{2}Log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta) - \frac{p}{2}Log(2\pi\tau^2) - \frac{1}{2\tau^2}\beta'\beta \tag{22}$$

By taking the derivative of both sides of the Equation 22, we get:

$$\frac{\partial \log P(\beta \mid y,X)}{\partial\beta} = \frac{\partial}{\partial\beta}\left(-\frac{n}{2}Log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta) - \frac{p}{2}Log(2\pi\tau^2) - \frac{1}{2\tau^2}\beta'\beta\right) \tag{23}$$

$$\frac{\partial \log P(\beta \mid y,X)}{\partial\beta} = 0 + \frac{X'(y-X\beta)}{\sigma^2} - 0 - \frac{1}{\tau^2}\beta \tag{24}$$

The derivative on the left-hand side will be made equal to zero.

$$0 = \frac{X'(y-X\beta)}{\sigma^2} - \frac{1}{\tau^2}\beta \tag{25}$$

After multiplying both sides of the Equation 25 by σ^2 , we get

$$0 = X'(y-X\beta) - \frac{\sigma^2}{\tau^2}\beta \tag{26}$$

$$0 = X'y - X'X\beta - \frac{\sigma^2}{\tau^2}\beta \tag{27}$$

$$X'X\beta + \frac{\sigma^2}{\tau^2}\beta = X'y \tag{28}$$

$$\left(X'X + \frac{\sigma^2}{\tau^2}\right)\beta = X'y \tag{29}$$

$$\beta_{BRR} = \left(X'X + \frac{\sigma^2}{\tau^2}\right)^{-1} X'y \tag{30}$$

This result, obtained by taking the log and then taking the derivative, is the best Bayesian estimator and is considered an estimate of the parameters β.

When we compare the best Bayesian estimator with the estimated formula for the Ridge regression, we note that $\lambda = \frac{\sigma^2}{\tau^2}$, which means that

$$\beta_{bayes} = \left(X'X + \lambda I\right)^{-1} X'y \tag{31}$$

Equations 11 to 31 illustrate the mathematical basis of Bayesian ridge regression (BRR), where the optimal estimator of the model's parameter β is derived using Bayesian inference techniques. Starting from the likelihood function and applying Bayes' rule, Equation 30 shows that the optimal Bayesian estimator of the parameters is given by $\beta_{BRR} = \left(X'X + \frac{\sigma^2}{\tau^2}\right)^{-1} X'y$. This result indicates that BRR is formally equivalent to ridge regression with the introduction of a regularization parameter $\lambda = \frac{\sigma^2}{\tau^2}$, as shown in Equation 31. This similarity between the two models demonstrates that BRR not only provides regularized parameter estimates but also adds a statistical dimension by incorporating prior information about the probability distribution of the parameters. In practice, these theoretical properties are directly reflected in the results of the current study, as the use of BRR helped address the problem of multicollinearity and reduce the risk of overfitting, leading to improved model stability and forecasting accuracy. The obtained experimental results confirm the superiority of BRR over ordinary linear regression, especially when dealing with complex and interconnected climate data, such as in the case of evaporation studies.

3. Echo State Network (ESN) and Time Series

Time series are characterized as a sequence of observations of a system over a designated time interval. These observations frequently exhibit irregularity, chaos, and nonlinearity. These attributes result in issues like as nonlinearity and uncertainty, which hinder future forecasting. Given that time series forecasting predominantly depends on previous data to forecast future values, the advancement of precise methodologies for managing this data is crucial in climate and environmental research. Numerous neural network models have been employed in this context, with the echo state network (ESN) being one of the most distinguished, recognized for its efficacy in managing intricate time series, particularly those marked by nonlinearity and substantial data interference. ESN is a contemporary neural network methodology engineered to effectively process sequence-oriented temporal data. The network utilizes an architecture including a dynamic reservoir comprised of many recurrent neurons with untrained random connections. The principal advantage of ESN resides in the straightforwardness of their training methodology. The input and reservoir weights are produced randomly and remain static, whereas only the output weights undergo training. This methodology differentiates them from conventional neural networks, which necessitate complete training of all parameters. This enhances the efficiency of ESN and reduces their vulnerability to optimization challenges and time complexity. Given that the dimensionality of the input variables to the model is p and the number of nodes or hidden units in the reservoir is m, the Echo State Network (ESN) design facilitates the development of an efficient model with substantial computational capacity to manage the dynamic properties of time series. The network functions as a long-range learning model, particularly adept at analyzing intricate systems like climate systems. The exceptional efficacy of ESNs in forecasting nonlinear and highly variable data renders them an appropriate selection for climate forecasting research, particularly when integrated with other statistical models like Bayesian regression (BRR) to create a hybrid model that enhances forecasting precision. Consequently, ESN is extensively utilized in various domains, including time series forecasting [15] , [18].The State Equation (SE) and the Observation Equation (OE) can be expressed as follows in Equations 32 and 33.

$$Z_t = AZ_{t-1} + Bu_{t-1} + C'a_t \tag{32}$$

$$Y_t = CZ_t \tag{33}$$

$r = max(g, j)$

g: is the number of the delayed series of the variable Z_t, j : is the number of the delayed series of errors, Z_t : is the state vector of dimension r, u_t: is the vector of the delayed series of errors, a_t : is the transposed vector of the current errors, Equations (32) and (33) are ambiguous in implementation, and for the sake of simplicity, their explanation can be formulated as follows in Equations 34 and 35.

$$Z_t = AZ_{t-1} + C'a_t \tag{34}$$

$$\widehat{y}_t = CZ_t \tag{35}$$

Z_t: is the m-dimensional state vector, A is the $(m \times m)$ dimension state transition matrix, C is the $(m \times 1)$ dimension width transition matrix, $\widehat{y}_t$:width is the transposed vector representing the output series, m is the number of lagged series terms of Z_t , and each lagged series of the residual series a_t is one of the right-hand side of the BRR-ESN data model equation after simplifying it and keeping Z_t only the left-hand side.

$$Z_t = [Z_{1,t} \; Z_{2,t} \; \ldots \; Z_{m,t}]' \tag{36}$$

$$A = \begin{vmatrix} P_1 & P_2 & P_3 & \ldots & P_m \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{vmatrix}_{m \times m} \tag{37}$$

$$C = [1 \; 0 \; 0 \; \ldots \; 0]_{1 \times m} \tag{38}$$

$P_1 \; P_2 \; P_3 \; \ldots \; P_m$ all parameter values for the lagged series Z_t and all time-lagged residual series a_t are included on the right-hand side of the equation for the BRR-ESN data model after simplification, retaining Z_t only the left-hand side. The matrix A, the row vector C, and the variables in the BRR-ESN data model equation will define the input variables and structure of the ESN used in this research. The Echo State Network (ESN) is a neural network model that utilizes linear regression for training, endowing it with a higher capacity to address data production challenges compared to typical neural networks that depend on gradient methods. ESN exhibit superior efficiency and enhanced forecasting accuracy relative to conventional networks, in addition to accelerated learning rates [19]. Moreover, the computational expenses linked to training an ESN are considerably reduced compared to the sequential mathematical procedures employed in conventional networks, such gradient-based techniques or global optimization approaches. The efficiency and flexibility of the ESN stem from its uncomplicated internal architecture and the absence of a requirement to update all weights, in contrast to conventional recurrent neural networks (RNNs). The ESN modelling procedure is delineated by the subsequent two equations 39 and 40.

$$Z_t = f(W_z Z_{t-1} + W_{in} X_t) \tag{39}$$

$$\widehat{y}_t = W_{out} Z_t \tag{40}$$

t represents the number of time stages, X_t represents the transposed input vector of dimension $(t \times 1)$, Z_t represents the transposed matrix of internal states of dimension $(m \times t)$ that resides in the hidden layer, and $\widehat{y}_t$ represents the transposed output vector of dimension $(t \times 1)$ of the neural network. W_{in} represent the weights in the input layer, W_z represents the weights in the hidden layer called the reservoir, W_{out} represent the weights in the output layer. The output weights can be calculated using the general inverse method with the following formula:

$$W_{out} = \widehat{y}_t' Z_t^+ = y_t'(Z_t' Z_t)^{-1} Z_t' \tag{41}$$

$$W_{in} = C', \; W_{out} = C, \; W_z = A \tag{42}$$

$$z = W_Z Z_{t-1} + W_{in} X_t \tag{43}$$

It is a non-linear function, and in this research the tangent function and its formula were used:

$$f(z) = tanh(z) = \frac{1 - e^{-2z}}{1 + e^{-2z}} \tag{44}$$

4. Hybrid BRR-ESN Model

The Bayesian Ridge Regression (BRR) and Echo State Network (ESN) models were employed to project future alterations and assess the influence of climate on the evaporation time series. The Bayesian Ridge Regression (BRR) model was selected as a sophisticated statistical method because of its efficacy in mitigating multicollinearity and overfitting concerns that might compromise traditional regression models. The BRR model integrates a prior information into the estimating process, enhancing model performance when addressing complicated and heterogeneous data, such as climate data. The BRR model was developed in this study to create a solid basis for the future advancement of the BRR-ESN hybrid model, assuring compatibility with the study data's characteristics and yielding more precise and dependable forecasting outcomes. The optimal Bayesian regression (BRR) model was constructed and chosen following multiple efforts to enhance the fit to the study data. The research predominantly utilized meteorological data, encompassing several climatic parameters and evaporation metrics from the chosen agricultural meteorological station. Prior literature suggests that meteorological data frequently exhibit nonlinearity, rendering linear models like BRR less effective for forecasting. Consequently, to more effectively manage the intricacies of these nonlinear data, sophisticated nonlinear models, such as the hybrid BRR-ESN model, were introduced to enhance forecasting performance. The overarching structure of the study comprises the subsequent steps:

1. Utilising the Time Stratified (TS) approach to category the data into two primary seasons: the hot season and the cold season.
2. Partitioning the data for each season into two subsets: a training subset and a testing subset.
3. Employing the BRR model to analyze the training set data.
4. Using an ESN network based on the outputs of the BRR model, by structuring the network inputs through a matrix A that simulates the structure of the BRR model parameters and their estimated values, a vector C that adjusts the dimensions of the predictive vector, a vector of residuals, and an input data matrix that is the same as the explanatory variables corresponding to the relevant parameters in the BRR model. The BRR-ESN hybrid model integrates Bayesian regression with echo networks to enhance the precision of evaporation forecasting. The general framework of the study can be formulated as shown in Figure (1) below.

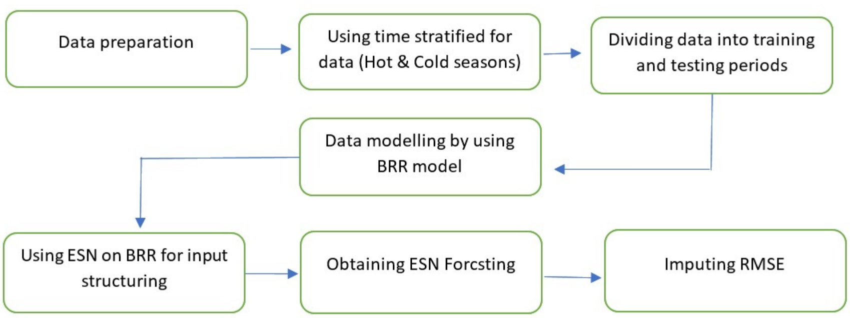

Figure 1. General framework of the study.

5. Time stratified (TS)

Time stratified (TS) is an effective method for analyzing time series data. It facilitates the organization of data based on seasonal fluctuations, which are critical determinants of time series dynamics and forecasting outcomes. This method seeks to categories data based on specific seasonal time strata, facilitating a more profound comprehension of the impact of seasons or cyclical temporal patterns on the variable under investigation. Time series analysis can be efficiently utilized for diverse time series, given that these series display consistent seasonal trends occurring under analogous correlation and influence conditions. This strategy classifies data into more homogeneous groupings, so diminishing random variance and improving the precision of statistical and forecasting modelling. This strategy enhances forecasting model performance by distinguishing repeating seasonal patterns from overall behavior, resulting in more trustworthy outcomes than those derived from aggregate, unclassified data [20] , [21].

6. Error measurements

The root mean square error (RMSE) is a metric used to express ideas about the accuracy of data prediction. RMSE [22]is written as follows in Equation 45.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} a_t^2}{n}} \tag{45}$$

$t = 1, 2, \ldots, n, \quad a_t = y_t - \widehat{y}_t$, n: number of observations, y_t :true value at time t, $\widehat{y}_t$:forecasted value of observation$y$at time t, a_t:error series at time t.

The mean absolute percentage error (MAPE) is one of the most popular measures of the forecast accuracy. It is recommended in most textbooks [23], Let y_t and $\widehat{y}_t$ denote the actual and forecast values at data point t, respectively. Then, MAPE is defined as follows in Equation 46.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{y_t - \widehat{y}_t}{y_t} 100 \tag{46}$$

where n is the number of observations.

7. Findings and Inferences

7.1. Data Characterization and Research Framework

This work employed hybrid evaporation forecasting models (BRR-ESN) to develop an accurate model for the city of Mosul. A comprehensive examination of the city's climatic data was performed from 1-November-2012, to 30-September-2022, to elucidate weather patterns and circumstances, thereby assessing the implications of climate change on evaporation. Climatic data were sourced from the Agricultural Meteorology Center at the Mosul Station in Nineveh Governorate, positioned at a longitude of 43.16°E and a latitude of 36.33°N. In this study, ten climatic variables were employed as explanatory factors, These include maximum temperature x_1, minimum temperature x_2, average temperature x_3, maximum relative humidity x_4, minimum relative humidity x_5, average relative humidity x_6, total solar radiation x_7, average wind speed x_8, maximum wind speed x_9, and wind direction x_{10}. Evaporation denoted by y was considered as the response variable. The variable of evaporation quantities for the hot season for the training and testing periods can be described in Figure (2) below. The variable of evaporation quantities for the cold season for the training and testing periods can be described in Figure (3) below.

The study focused on data preparation and forecasting modelling, utilizing forecasting models to depict the behavior of climate data in Mosul during summer and winter seasons. The objective was to forecast evaporation quantities employing the hybrid BRR-ESN method, influenced by various climatic factors. A seasonal classification

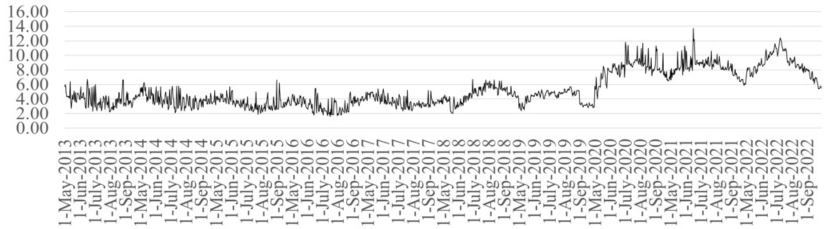

Figure 2. Description of the evaporation quantities variable for the hot season for the training and testing periods.

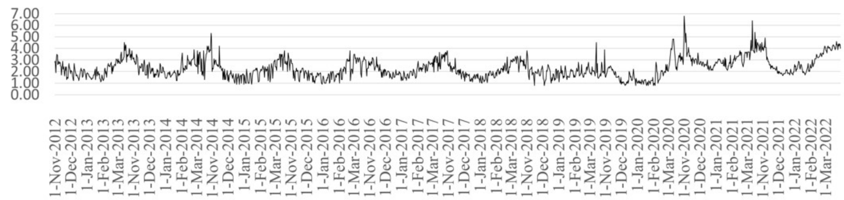

Figure 3. Description of the evaporation quantities variable for the cold season for the training and testing periods.

approach was implemented to address the variability of climatic data in the city, hence enhancing data consistency and improving forecast efficiency for the daily period from November 1, 2012, to September 30, 2022. The hot season data encompassed the months of May to September from 2012 to 2022, whilst the cold season data included the months from November to March for the same time. For the hot season, 1,200 observations were designated for training data and 300 for testing data; similarly, for the cold season, 1,200 observations were allotted for training data and 300 for testing data. The data were formulated and structured according to the seasonal multiplier model utilizing (BRR-ESN), with the study presuming a cyclical seasonality in which each seasonal cycle encompasses five months, hence establishing the value of the seasonal period as (s=5).

8. Bayesian Ridge Regression Model

This research examines the outcomes of employing Bayesian ridge regression on time series data concerning the impact of climatic variables on evaporation, the primary response variable in the study, and the climatic factors affecting it from 2012 to 2022. The examination of climatic conditions and evaporation data indicated substantial relationships among certain variables. The concurrent and interrelated characteristics of climate variables, including temperature, humidity, and wind, resulted in significant correlations among them, causing pronounced multicollinearity in the design matrix, rendering the explanatory variables insufficiently independent. Multicollinearity undermines the model, resulting in bias and instability in parameter estimates, as the estimated values exhibit heightened sensitivity to minor data fluctuations. Research demonstrates that traditional linear regression models in these instances yield erroneous and exaggerated coefficients, thereby affecting the model's trustworthiness. The Bayesian ridge regression (BRR) model was employed to tackle this issue by incorporating a Bayesian regularization term, thereby stabilizing parameter values and diminishing their variation. This enhances model performance and elevates forecast accuracy, particularly in settings marked by numerous correlations and erratic variables. BRR additionally facilitates the provision of probability distributions for parameters, thereby offering a more profound comprehension of the uncertainty linked to estimation and augmenting its trustworthiness as a sophisticated forecasting model in contrast to conventional linear models. The preliminary use of BRR facilitated the creation of a hybrid model incorporating an echo state network (ESN), which aligns with the intricate dynamic properties of climate data. To identify multicollinearity among the examined variables, the Simple Linear Correlation Coefficient (SLC), the Pearson Correlation Coefficient (PCC), and the Variance Inflation Factor

(VIF) values were utilized. The results demonstrated significant correlations among certain variables, suggesting the possibility of multicollinearity that impacts the stability and precision of forecasting models. Specifically, multicollinearity was detected between the variables (x_1, x_2, x_3) and (x_4, x_6)in the data for both the hot and cold seasons, as illustrated in the subsequent tables. These tables show the values of the correlation coefficients and variance inflation factors, supporting the need to use advanced regression models capable of dealing with this problem, such as the Bayesian Ridge Regression model.

$$r_{xy} = \begin{array}{c|ccccccccccc|} & y & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \\ y & 1 & 0.097 & 0.035 & 0.072 & -0.039 & -0.040 & -0.044 & 0.042 & 0.837 & 0.252 & -0.0100 \\ x_1 & & 1 & 0.760 & 0.945 & -0.592 & -0.643 & -0.687 & 0.261 & -0.062 & -0.187 & -0.009 \\ x_2 & & & 1 & 0.931 & -0.565 & -0.541 & -0.629 & 0.248 & 0.027 & -0.005 & 0.015 \\ x_3 & & & & 1 & -0.617 & -0.633 & -0.703 & 0.271 & -0.021 & -0.107 & 0.002 \\ x_4 & & & & & 1 & 0.498 & 0.951 & -0.120 & 0.104 & 0.141 & -0.042 \\ x_5 & & & & & & 1 & 0.741 & -0.435 & 0.106 & 0.135 & -0.030 \\ x_6 & & & & & & & 1 & -0.247 & 0.118 & 0.158 & -0.043 \\ x_7 & & & & & & & & 1 & 0.088 & 0.069 & 0.061 \\ x_8 & & & & & & & & & 1 & 0.470 & -0.199 \\ x_9 & & & & & & & & & & 1 & -0.295 \\ x_{10} & & & & & & & & & & & 1 \end{array}$$

Table (1) below indicate the highest values of the Variance Inflation Factor (VIF) among the ten variables $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$ as explanatory variables when y is the evaporation variable as the response variable.

Table 1. Variance Inflation Factor (VIF) values of the explanatory variables

	Variable	Parameters	T-Value	P-Value	VIF
Constant		2.6010	5.63	0.000	
	x_1	0.1068	3.18	0.002	9.28
	x_3	-0.0808	-2.15	0.032	9.28
Constant		2.6010	5.63	0.000	
	x_2	-0.1068	-3.18	0.002	7.55
	x_3	0.1328	3.91	0.000	7.55
Constant		4.7020	19.03	0.000	
	x_4	0.0057	0.38	0.701	10.54
	x_6	-0.0196	-0.86	0.392	10.54

Therefore, we will resort to using Bayesian Ridge Regression (BRR) to solve the problem, and the parameter values are as follows. The Bayesian Ridge Regression equation can also be written as follows in Equation 47.

$$\begin{aligned} \widehat{y} = -0.6822 + 0.0956x_1 - 0.1069x_2 - 0.0057x_3 - 0.0046x_4 - 0.0328x_5 \\ -0.0187x_6 - 0.0476x_7 + 2.7982x_8 - 0.0829x_9 + 0.0014x_{10} \end{aligned} \tag{47}$$

Table (2) below indicate the highest values of the Variance Inflation Factor (VIF) for the 10 variables $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$ as explanatory variables when y is the evaporation variable as the response variable.

$$r_{xy} = \begin{array}{r|ccccccccccc|} & y & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \\ y & 1 & 0.695 & 0.363 & 0.601 & -0.180 & -0.062 & -0.167 & 0.718 & 0.107 & 0.152 & 0.120 \\ x_1 & & 1 & 0.644 & 0.926 & -0.368 & -0.031 & -0.293 & 0.459 & -0.063 & -0.033 & 0.152 \\ x_2 & & & 1 & 0.886 & -0.302 & 0.001 & -0.227 & 0.109 & 0.191 & 0.223 & -0.181 \\ x_3 & & & & 1 & -0.373 & -0.018 & -0.290 & 0.333 & 0.056 & 0.090 & 0.003 \\ x_4 & & & & & 1 & 0.244 & 0.875 & 0.014 & 0.071 & 0.091 & -0.027 \\ x_5 & & & & & & 1 & 0.682 & 0.058 & 0.315 & 0.239 & -0.058 \\ x_6 & & & & & & & 1 & 0.039 & 0.211 & 0.187 & -0.049 \\ x_7 & & & & & & & & 1 & -0.001 & 0.007 & 0.217 \\ x_8 & & & & & & & & & 1 & 0.827 & -0.311 \\ x_9 & & & & & & & & & & 1 & -0.372 \\ x_{10} & & & & & & & & & & & 1 \end{array}$$

Table 2. Variance Inflation Factor (VIF) values of the explanatory variables

	Variable	Parameters	T-Value	P-Value	VIF
Constant		-0.0598	-0.92	0.360	
	x_1	0.1627	17.85	0.000	6.98
	x_3	-0.0599	-5.41	0.000	6.98
Constant		-0.0598	-0.92	0.360	
	x_2	-0.1627	-17.85	0.000	4.63
	x_3	0.26552	29.46	0.000	4.63
Constant		3.104	20.03	0.000	
	x_4	-0.00897	-2.50	0.013	4.27
	x_6	-0.00364	-0.67	0.502	4.27

Consequently, we will employ Bayesian Ridge Regression (BRR) to address the issue, with the following parameter values. The equation for Bayesian Ridge Regression can be expressed as follows in Equation 48.

$$\begin{aligned} \widehat{y} = -0.6423 + 0.0.0754x_1 - 0.0.0276x_2 + 0.0238x_3 + 0.0017x_4 - 0.0100x_5 \\ -0.0042x_6 + 0.1096x_7 + 0.0447x_8 + 0.0462x_9 - 0.0001x_{10} \end{aligned} \tag{48}$$

The VIF values were examined after estimating the parameters using BRR and it was found that the problem of multicollinearity was completely eliminated through its values indicating the absence of multicollinearity.

9. The hybrid BRR-ESN method

According to the ESN model, which is fundamentally derived from the state space model in Equations (34) and (35), along with Equation (46), the ESN inputs will be organized such that the A matrix and C row vector for the hot season will be as follows:

$$A = \begin{bmatrix} 0.0956 & -0.1069 & -0.0057 & -0.0046 & -0.0328 & -0.0187 & -0.0476 & 2.7982 & -0.0829 & 0.0014 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

and an X_{tr} matrix with dimensions (Number of variables $\times$ Number of observations)

Regarding the cold season, according to the ESN model, fundamentally derived from the state space model in equations (34) and (35), along with equation (47), the ESN inputs will be organized such that matrix A and row vector C will be as follows:

$$A = \begin{bmatrix} 0.0754 & -0.0276 & 0.0238 & 0.0017 & -0.0100 & -0.0042 & 0.1096 & 0.0447 & 0.0462 & -0.0001 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

and an X_{tr} matrix with dimension (Number of variables * Number of observations)

The outcomes of administering these tests on time series data are presented in the subsequent tables, The tables 3 and 4 below illustrates the Root Mean Square Error (RMSE) values for the BRR and BRR-ESN models across the two seasons (hot and cold) regarding training and testing data derived from 10 climatic variables.

Table 3. The RMSE values of BRR and BRR-ESN for training period

	BRR	**BRR-ESN**
Cold	0.3803	0.3802
Hot	0.8285	0.8258

Table 4. The RMSE values of BRR and BRR-ESN for testing period

	BRR	**BRR-ESN**
Cold	0.9144	0.4502
Hot	2.5831	1.6204

The tables 5 and 6 below illustrates the Mean Absolute Percentage Error (MAPE) values for the BRR and BRR-ESN models across the two seasons (hot and cold) regarding training and testing data derived from 10 climatic variables.

Table 5. The MAPE values of BRR and BRR-ESN for training period

	BRR	BRR-ESN
Cold	0.1453	0.0107
Hot	0.1689	0.0118

Table 6. The MAPE values of BRR and BRR-ESN for testing period

	BRR	BRR-ESN
Cold	0.2458	0.0262
Hot	0.2453	0.0374

The data derived from Tables 3 and 4 demonstrate that the suggested hybrid model (BRR-ESN) surpasses the conventional Bayesian Ridge Regression model. The hybrid model exhibited reduced root mean square error (RMSE) values, signifying a notable enhancement in forecasting accuracy. This superiority demonstrates the hybrid model's capacity to tackle the nonlinearities and heterogeneity present in climate data, fulfilling the study's main goal of enhancing climate forecasting performance.
The data derived from Tables 5 and 6 demonstrate that the suggested hybrid model (BRR-ESN) surpasses the conventional Bayesian Ridge Regression model. The hybrid model exhibited reduced the mean absolute percentage error (MAPE) values, signifying a notable enhancement in forecasting accuracy. This superiority demonstrates the hybrid model's capacity to tackle the nonlinearities and heterogeneity present in climate data, fulfilling the study's main goal of enhancing climate forecasting performance. Then the result of (MAPE) is confirm the result of (RMSE). The findings indicate that evaporation, as the dependent variable, is significantly influenced by variations in multiple independent climatic factors, including maximum and minimum temperatures and their averages, maximum and minimum relative humidity and their averages, solar radiation, wind direction, average wind speed, and the peak recorded wind speed. The implementation of the Time Stratification (TS) method enhanced the uniformity of seasonal data, thereby improving the forecasting performance of the hybrid model and validating its efficacy and efficiency in managing multivariate time series characterized by nonlinear and seasonal attributes.

Conclusions: Based on the findings in the ”Results and Discussion” section, it is evident that the hybrid model (BRR-ESN) significantly enhances forecast quality compared to conventional approaches employed for analyzing climate variables. This validates the model's exceptional efficacy in managing multivariate time series, especially in forecasting evaporation quantities, a variable influenced by various fluctuating climatic conditions. This method is distinguished as an effective and adaptable choice for the nonlinear and heterogeneous characteristics of data, improving the reliability and precision of forecasting in intricate climate scenarios.

REFERENCES

1. k. Murphy *Machine learning: a probabilistic perspective*, MIT press. 2012.
2. T. Kim, and B. King *Time series prediction using deep echo state networks*, Neural Computing and Applications, vol. 32, no. 23, pp. 17769–17787, 2020.
3. J. Viehweg, K. Worthmann, and P. Mader, *Parameterizing echo state networks for multi-step time series prediction*, Neurocomputing, vol. 522, pp.214–228, 2023.
4. Y. Peng, and et al, *A novel hybridization of echo state networks and multiplicative seasonal ARIMA model for mobile communication traffic series forecasting*, Neural Computing and Applicationsp, vol.24, pp. 883–890, 2014.
5. M.Kutner, and et al, *Applied linear statistical models*, McGraw-hill. 2005.
6. D. Montgomery, E. Peck, and G. Vining, *Introduction to linear regression analysis. John Wiley and Sons*, Inc., Hoboken, New Jersey, pp. 12–122, 2012.
7. T. Hastie, and et al, *The elements of statistical learning: data mining, inference, and prediction*, Springer, vol. 2, 2009.

8. A. Hoerl, and R.Kennard, *TRidge regression: Biased estimation for nonorthogonal problems*, Technometrics, vol. 12, no. 1 pp. 55–67, 1970.
9. D. Donoho, and N. Nasrabadi, *Pattern recognition and machine learning*, Springer, vol. 4, 2006.
10. S. Edition, *Bayesian data analysis*, CRC Press, 2013.
11. P. M.Lee, *Bayesian Statistics: an Introduction*, John Wiley and Sons, 2012.
12. T. Hastie, *The elements of statistical learning: data mining, inference, and prediction*, Springer, 2009.
13. N. Heard, *An introduction to Bayesian inference, methods and computation*, Springer, 2021.
14. G. Box, and G. Tiao, *Bayesian inference in statistical analysis*, John Wiley and Sons, 2011.
15. 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: 1613–1625, 2019.
16. Q. Ma, and et al, *Convolutional multitimescale echo state network*, IEEE Transactions on Cybernetics, vol.51, no.3, p: 1613–1625, 2019.
17. J. Qiao, and et al, *Growing echo-state network with multiple subreservoirs*, IEEE transactions on neural networks and learning systems, vol.28, no. 2, p: 391–404, 2016.
18. C. Yang, and et al, *Dynamical regularized echo state network for time series prediction*. Neural Computing and Applications, vol. 31, p: 6781–6794, 2019.
19. 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, p: 2413–2425, 2015.
20. B. Malig, and et al, *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, p: 745–753, 2016.
21. 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, Stata Users Group, 2014.
22. R. Hyndman, and A. Koehler, *Another look at measures of forecast accuracy*. International journal of forecasting, vol. 22, no. 4, p: 679–688, 2006.
23. Bowerman, Bruce L, O'Connell, Richard T, and Koehler, Anne B, *Forecasting, time series, and regression: an applied approach*. Thomson Brooks/Cole, 2005.