Research Article

Journal of Applied Science and Emerging Technology (JASET)
Published by Faculty of Applied Science, KMUTNB, Thailand
doi: 10.14416/JASET.KMUTNB.2025.02.002
Vol. 24, No. 2 [2025]: e260276

Modeling Global Ice Volume Changes: A Nonlinear Autoregressive Neural Network Approach

Porawat Visutsak[1*] and Keun Ho Ryu[2, 3]

[1]Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, 10800 Thailand
[2]Database/Bioinformatics Laboratory Chungbuk National University, Chungbuk, 28644 Republic of Korea
[3]Faculty of Information Technology, Data Science Laboratory, Ton Duc Thang University, Vietnam

* Corresponding Author, E-mail: porawatv@kmutnb.ac.th DOI: 10.14416/JASET.KMUTNB.2025.02.002
Received 28 December 2024; Revised 12 July 2025; Accepted 17 July 2025

ABSTRACT

A single time series prediction problem is solved with a neural network. The nonlinear autoregressive (NAR) type of network is used. The network is trained in an open loop and then transformed to closed loop for multistep prediction. The prediction is made 20 time steps into the future. The delay is removed from the network to get the prediction one time step earlier. The shallow neural network is trained on the global ice volume dataset, which contains 219 measurements of global ice volume over 440,000 years. The network is able to predict future ice volume based on past values with a high degree of accuracy. Three different backpropagation training algorithms were used to train the network: Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient. The Levenberg-Marquardt algorithm achieved the lowest MSE (0.02257 at epoch 13) and the highest R^2 (0.99254). The Bayesian Regularization algorithm achieved an MSE of 0.027209 at epoch 4 and an R^2 of 0.99192. The Scaled Conjugate Gradient algorithm achieved an MSE of 0.01878 at epoch 3 and an R^2 of 0.99018. This work contributes to the field of climate studies by providing a tool for predicting future ice volume. This information can be used to better understand Earth's glacial cycles and to develop strategies for mitigating the effects of climate change.

KEYWORDS: Time-series prediction, Neural network, NAR, Backpropagation, Glacial cycles

Please cite this article as: P. Visutsak, "Modeling Global Ice Volume Changes: A Nonlinear Autoregressive Neural Network Approach," *Journal of Applied Science and Emerging Technology,* vol 24, no 2, pp. 1-15, ID. 260276, August 2025

1. INTRODUCTION

Global ice volume has fluctuated significantly over the past 440,000 years, with major implications for Earth's climate and sea level. Understanding these fluctuations is essential for predicting future climate change and its potential impacts on human societies. This paper explores the use of shallow neural networks for time-series prediction, focusing on modeling global ice volume.

The objective of this study is to develop a neural network model that can accurately predict future global ice volume based on past values. The model will be trained on the global ice volume dataset, which contains 219 measurements of global ice volume over 440,000 years. This single time-series prediction problem leverages the network's ability to learn patterns in historical data. By analyzing past ice volume trends, the network can forecast future changes, contributing valuable insights to climate studies and our understanding of Earth's glacial cycles.

The neural network model will be trained using three different backpropagation algorithms: Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient. The performance of the models will be evaluated based on their mean squared error (MSE) and R^2.

The Levenberg-Marquardt algorithm achieved the lowest MSE (0.02257 at epoch 13) and the highest R^2 (0.99254). The Bayesian Regularization algorithm achieved an MSE of 0.027209 at epoch 4 and an R^2 of 0.99192. The Scaled Conjugate Gradient algorithm achieved an MSE of 0.01878 at epoch 3 and an R^2 of 0.99018.

This paper is organized as follows: Section 2 describes the materials and methods used in this study, including the global ice volume dataset and the neural network model. Section 3 presents the results of the experiments, including the training state, MSE, and R^2 for each backpropagation algorithm. We also discuss the implications of the research findings for climate studies and our understanding of Earth's glacial cycles. Section 4 concludes the paper with a summary of the main findings and suggestions for future research.

The study of using nonlinear autoregressive neural network models for predicting oil-dissolved gas concentrations has been proposed by Pereira et al. (2018). The authors propose a hybrid model to predict the concentration of gases dissolved in transformer oil, which is crucial for monitoring equipment health and diagnosing potential faults. This model combines a Nonlinear Autoregressive (NAR) neural network with the Discrete Wavelet Transform (DWT). The DWT is first used to process and simplify historical gas data, which is then fed into the NAR network to forecast future gas concentration values. Tested on seven months of real-world data from a transformer in Brazil, the proposed NAR-DWT model demonstrated high accuracy and robustness, proving to be less sensitive to training parameters like time delay and the specific wavelet function used. The results showed that this approach significantly outperformed other methods like GRNN, BPNN, and SVM, with a prediction error for ethylene gas that was approximately 70% smaller than the other tested models. The study concludes that this model serves as a reliable tool for fault diagnosis systems, enabling the anticipation of failures by providing accurate, multi-step ahead predictions of gas levels.

Boussaada et al. (2018) presents a model for predicting daily direct solar radiation specifically designed to manage the power supply for a race sailboat, an application where location is constantly changing. The authors utilize a Nonlinear Autoregressive Exogenous (NARX) neural network

that takes two primary inputs: a deterministic component derived from a "clear sky" model that accounts for the sun's geometric position, and a statistical component represented by predicted cloud cover. A key finding is that the model performs best when trained periodically—specifically, once every day at midnight using a moving window of the last 10 days of data. This daily retraining strategy allows the model to adapt to the sailboat's changing coordinates and meteorological conditions. After optimizing the network's structure to 15 neurons in the hidden layer and using specific activation functions, the final predictor achieved a Daily Mean of the Power Error (DMPE) of 24.0584 W/m², successfully forecasting the general solar radiation curve.

Moreover, to aid in strategic planning for sustainable energy systems, Adedeji et al. (2019) addresses the challenge of forecasting energy consumption, which is subject to hemispherical seasonal patterns. The research employs a non-linear autoregressive neural network (NARNET) to predict daily energy usage across the four campuses of a South African university, utilizing three years of historical consumption data. A critical component of the methodology involved preprocessing the data with Singular Spectrum Analysis (SSA) for filtering. The authors identified three potential window lengths (L=54, 103, and 155) via periodogram analysis and compared their impact on the network's training performance. The data filtered with a window length of L=103 produced the best results, yielding R-values of 0.951, 0.983, 0.945, and 0.940 for campuses A, B, C, and D, respectively. Following network validation, a short-term forecast achieved accuracies of 85.87% (Campus A), 75.62% (Campus B), 85.02% (Campus C), and 76.83% (Campus D). The study concludes by demonstrating the significance of data filtering as a crucial step for improving the accuracy of forecasts based on univariate autoregressive series.

It's essential to understand how regional climates will change under different human-caused emission scenarios to help societies adapt and mitigate the impacts (Mansfield et al., 2020). Traditional climate models require immense computing power, posing a major challenge. However, Mansfield et al. (2020) proposed a solution: using machine learning. Their innovative approach leverages a unique dataset of existing climate model simulations to identify connections between short-term and long-term temperature responses under various climate forcing scenarios. This method offers a twofold advantage. First, it can speed up climate change projections by reducing the computational burden. Second, it helps identify early warning signs of long-term climate responses, which is vital for climate change detection, predictability, and attribution. The research emphasizes the need for increased data sharing between research institutions. By pooling resources to build larger datasets, the scientific community can create even more powerful climate response emulators. This will lead to faster and more accurate climate change projections. Such a collaborative approach is crucial for navigating the challenges and opportunities of data-driven climate modeling.

Baig et al. (2021) investigated hydro-meteorological variables in the Chitral Basin of Pakistan to predict climate change impacts on temperature, precipitation, humidity, and river flow using observed data from 1990 to 2019. The researchers employed statistical methods, including trend variability analysis and regression models, to analyze the relationships between these variables. Their findings revealed an inverse relationship between temperature and precipitation, with temperature decreasing by 0.309 °C for every unit

increase in precipitation. Temperature also showed a negative correlation with humidity. Conversely, precipitation positively influenced both humidity and river flow. These results challenge the notion that increased river flow in the Chitral Basin is primarily due to glacial recession caused by rising temperatures.

Zhou et al. (2021) investigated carbon dioxide (CO2) emissions in China, focusing on identifying key emission sources and driving factors to mitigate global warming. The study utilized multiple linear regression models to analyze data from 1990 to 2017, identifying the energy industry, fuel combustion in other industries, and industrial processes as the primary CO2 emission sources. The researchers developed driving force models for each source, incorporating both quantitative and qualitative factors. Their models predict a continued decrease in CO2 emission intensity and total emissions in China but emphasize the need for greater efforts to meet the Paris Agreement goals. The study highlights the importance of energy structure adjustment, technological innovation, and policy interventions in achieving significant CO2 emission reductions.

Cummins et al. (2022) addressed the statistical validity of commonly used climate change detection and attribution (D&A) methods. These methods, based on Hasselmann's "optimal fingerprinting," employ linear regression of historical climate observations on climate model output, raising concerns about spurious regression due to the non-stationary nature of climate variables. Using an idealized linear-response-model framework, the authors demonstrated the consistency of the optimal fingerprinting estimator under standard assumptions, particularly for global mean surface temperature (GMST). Analysis of historical GMST observations and CMIP6 model output supported these assumptions, indicating that D&A of GMST trends is likely not spurious. Furthermore, the study revealed "superconsistent" properties of the least-squares estimator due to cointegration between observations and model output. This finding led to the development of a new method for quantifying D&A uncertainty, eliminating the need for pre-industrial control simulations.

Duan et al. (2022) investigated temperature trends in Australia using a joint model of quantile regression and variability to analyze daily maximum and minimum temperature data. This approach accounts for heterogeneity in the data, including quasi-periodic heterogeneity in variance, which has often been overlooked in previous climate studies. Their analysis revealed an overall warming trend of approximately 0.21°C per decade for daily maximum temperatures and 0.13°C per decade for daily minimum temperatures. Moreover, the study identified nuanced spatial and temporal patterns of change, varying across locations, seasons, and temperature percentiles.

Subhra et al. (2023) investigated the effectiveness of polynomial regression for predicting future temperatures and, consequently, climate variation. Using a preprocessed dataset from NASA's Jet Propulsion Laboratory, the researchers applied a polynomial regression model to predict future temperatures. Their findings demonstrate the efficacy of this approach, achieving a 93.31% training accuracy and a 91.01% testing accuracy. The authors suggest that this prediction model can aid environmental agencies in mitigating the impact of climate change and forecasting extreme weather conditions. They recommend further research incorporating additional climate-influencing factors.

Utami et al. (2023) investigated the impact of climate change on heat stress in Indonesia, measured by the Thermal Humidity Index (THI). Recognizing the limitations of coarse-resolution Earth System Models (ESMs), the researchers employed statistical downscaling (SD) and nonparametric regression to correct biases in temperature and humidity projections from the Coupled Model Intercomparison Project (CMIP5). Their bias correction method achieved high accuracy, with R-square values of 95% for relative humidity and 94% for temperature. Based on THI projections, the study found that 50% of the Indonesian population will experience comfortable conditions from 2006 to 2059, while uncomfortable conditions are expected from 2060 to 2100, with THI values ranging from 27.0730°C to 27.7800°C.

Wang and Xia (2023) reviewed the application of quantile regression in addressing climate change. Quantile regression, an extension of linear regression, estimates the median or other quantiles of the outcome variable, making it suitable for analyzing data with non-normal residuals, outliers, and heteroscedasticity. This approach is particularly useful in climate change research as it allows for the examination of relationships between variables beyond the mean, capturing the impact of unusually distributed outcomes and nonlinear relationships. The authors highlighted the advantages of quantile regression in identifying factors influencing variables at different quantiles and their robustness against outliers. The review emphasizes the potential of quantile regression in advancing climate change research and provides directions for future studies.

Chang et al. (2024) examined various methods for extracting long-term temperature trends from sea surface temperature (SST) data, including the seasonal-trend decomposition procedure based on loess (STL) and linear regression methods (ordinary least square regression [OLSR], orthogonal regression [OR], and geometric mean regression [GMR]). STL was identified as the most accurate method but computationally expensive. Linear regression methods were more efficient, but GMR was deemed unsuitable due to its assumption of a random temporal component. OLSR and OR required correction for seasonal signal-induced bias, which could be achieved by trimming the SST data. Grover and Sharma (2024) explored the use of ridge regression to analyze rainfall trends and temporal patterns, aiming to enhance the understanding of precipitation dynamics. Ridge regression, an improved regression technique, addresses multi-collinearity and overfitting issues, thus improving the accuracy of rainfall forecasting. The study examined rainfall variations, considering the impact of climate change parameters. The authors found that ridge regression, by balancing bias and variance, led to more robust and reliable rainfall predictions. This research contributes valuable insights to hydrology and climate science, emphasizing the importance of accurate rainfall forecasting for effective water resource management and climate studies. The study highlights the potential of ridge regression in capturing complex temporal processes, which could be applied to various climate-related issues.

Malik et al. (2024) assessed the impact of climate change on Pakistan's energy sector, focusing on demand, transmission, and generation. Using climate projections and an ANN-based approach, they predicted a significant increase in energy demand across most regions of Pakistan. Their analysis also revealed potential capacity losses of up to 23.34% in some transmission lines due to rising temperatures. The study emphasized the importance of incorporating renewable energy

sources to enhance energy efficiency and grid resilience. The authors concluded that without infrastructure upgrades, efficient technologies, and demand-side management, climate change could severely impact Pakistan's energy grid.

Takefuji (2024) investigated the relationship between global carbon dioxide (CO2) levels and temperature anomaly using linear regression analysis on datasets from the National Oceanic and Atmospheric Administration (NOAA). The study found that the relationship between these two variables can vary over time, with long-term data (March 1958 to June 2023) suggesting a strong positive association and short-term data (March 1990 to March 1994) showing a negative association. The author concluded that further research is needed to fully understand the complex relationship between global CO2 levels and temperature anomaly, taking into account other potential influencing factors such as changes in climate patterns and human activities.

2. MATERIALS AND METHODS

We begin by briefly outlining the contributions of this work, followed by an illustration of the methodology. Further details will be discussed later in this section.

This work uses a NAR neural network to predict future global ice volume based on past measurements. The dataset used to train and evaluate the network is from Newton & North (1991) and contains 219 measurements of global ice volume over the past 440,000 years, stored in a variable called "iceTargets".

The NAR network is chosen due to the nature of the dataset, which is a single time-series. This means the data consists of a sequence of measurements taken over time, with each measurement representing the global ice volume at a particular point in time. The NAR network is designed to capture temporal dependencies in such time-series data by using past values to predict future ones. The NAR equation can be derived as shown in Equation 1:

$$y(t) = f(y(t-1), \ldots, y(t-d)) \qquad (1)$$

Where,

- $y(t)$: This represents the value of the time series one is trying to predict at time "t" (the "present" value).
- $f(\ldots)$: This is a nonlinear function. It can be anything from a simple nonlinear transformation to a complex neural network. This function is what the model learns from the data.
- $y(t-1)$, ..., $y(t-d)$: These are the past values of the time series. "d" represents the number of past values (or "delays") used to make the prediction.

The network is trained using the Levenberg-Marquardt backpropagation algorithm, a common and effective algorithm for training neural networks. To ensure the network's generalization ability and prevent overfitting, the data is divided into training, validation, and testing sets.

The training process involves two stages. First, the network is trained in an open-loop configuration, where the network's own predictions are not fed back as input during training. This helps the network learn the fundamental patterns in the data. Second, the network is transformed into a closed-loop configuration for multistep prediction, where the network's own predictions are used as input for future predictions. This allows the network to generate a sequence of predictions extending beyond the training data.

Finally, the network's performance is evaluated using metrics such as MSE and R^2. These metrics provide a quantitative measure of how well the

network can predict future global ice volume based on past measurements.

2.1 Data selection and preparation

The foundation of this study is the global ice volume dataset, a well-established time series in paleoclimatology literature originating from the work of Newton & North (1991). This dataset contains 219 measurements that estimate global ice volume over the past 440,000 years. It's important to note that these are not direct measurements but are proxy-base d estimates derived from geological records, designed to capture the long-term glacial cycles of the Earth. All modeling was implemented in

MATLAB R2024a. Before training, the data requires specific preparation for the NAR network architecture. This was handled using the preparets function in MATLAB, which performs two critical tasks:

- Initial State Configuration: The function sets aside the initial data points required to populate the network's feedback delays (or "tapped delay lines"). For our model, which uses delays of *t-1* and *t-2*, the first two data points are used to establish the initial state.
- Data Restructuring: These initial values are then removed from the training data, as their information is now embedded in the network's delay lines. This ensures that the network has the necessary historical context to begin the prediction process from the very first training step.

2.1.1 Dataset Limitations

While the global ice volume dataset is valuable for modeling long-term climate trends, it is essential to acknowledge its limitations to contextualize our findings. The primary limitation is its temporal sparsity. With only 219 data points spanning 440,000 years, the dataset provides an average of one measurement approximately every 2,000 years. This low resolution means that shorter-term or more abrupt climatic fluctuations within the glacial cycles are not captured. The model is therefore trained on a smoothed, long-term representation of ice volume changes.

Secondly, as proxy data, the measurements are subject to inherent uncertainties. These can arise from the original geological sampling, the dating methods used, and the models used to convert raw geological signals into ice volume estimates. While the dataset is a standard benchmark, these potential biases are a factor in any analysis. Our study's results demonstrate the NAR network's powerful ability to model the clear, long-term patterns present in this sparse data. However, the predictive resolution is inherently limited by the dataset's granularity. More detail will be expanded in section 3 later on.

2.2 NAR Network Architecture

The selection of a NAR neural network is justified by the nature of the global ice volume dataset, which is a single time-series. This type of network is specifically designed to identify and model temporal dependencies in sequential data by using past values to predict future ones. The architecture employs a feedback loop, which allows the network to use its own predictions as inputs for subsequent forecasts, making it suitable for multi-step prediction. The model's configuration, including a hidden layer of 10 neurons and feedback delays of $t-1$ and $t-2$, was determined through an iterative experimental process to balance accuracy and complexity, thereby preventing overfitting.

We employ a NAR neural network, designed to predict future values of a time series based solely on its past values. The NAR network is instantiated

in MATLAB using the narnet function. The network's architecture is defined by two key parameters:

1. Feedback Delays: These determine the number of past time steps the network considers for prediction. In our model, we utilize delays of 1 and 2, implying the network utilizes values from the previous two time steps ($t-1$ and $t-2$) to predict the current value (t).
2. Hidden Layer Size: This specifies the number of neurons in the network's hidden layer, where complex patterns in the data are learned. Our network incorporates a hidden layer with 10 neurons. The rationale for selecting these specific parameters is detailed below.

2.2.1 Hyperparameter Selection

The selection of appropriate hyperparameters is critical to building an effective neural network and ensuring the robustness of the results. For our NAR model, the key hyperparameters are the number of feedback delays and the hidden layer size. The optimal values for these were determined through a systematic, iterative experimental process aimed at finding a balance between model accuracy and complexity to prevent overfitting.

- Feedback Delays: We experimented with feedback delays ranging from 1 to 4. A model with only one delay ($t-1$) was insufficient to capture the historical dependencies in the data. While using three or more delays showed marginal improvement, it significantly increased model complexity without a proportional gain in performance. The chosen delays of 1 and 2 provided the best performance on the validation set, capturing sufficient past information for accurate prediction.
- Hidden Layer Size: The number of neurons in the hidden layer was varied, with configurations ranging from 5 to 20 neurons. Networks with fewer than 10 neurons struggled to model the nonlinear, cyclical patterns of the ice volume data, leading to higher errors. Conversely, networks with more than 15 neurons began to show signs of overfitting the training data, with performance on the validation set ceasing to improve.

Based on this empirical tuning process, an architecture with a hidden layer of 10 neurons and feedback delays of 1 and 2 was identified as the optimal configuration for this specific time-series problem.

2.3 Network Training

The NAR network's training process is as follows:

1. Data Division: The prepared time series data is randomly partitioned into three sets: 70% for training, 15% for validation, and 15% for testing. This division facilitates effective training, performance monitoring, and unbiased evaluation of the final trained network.
2. Open-loop Training: The network is initially trained in an "open-loop" configuration. In this mode, the network's predictions are not fed back as input during training. Instead, the network relies solely on actual past values from the time series to learn fundamental patterns in the data.
3. Closed-loop Transformation: Post open-loop training, the network is converted into a "closed-loop" configuration for multistep prediction. This configuration enables the network to utilize its own predictions as input for subsequent predictions, allowing

for the generation of a prediction sequence extending beyond the training data.

4. Training Algorithm: The network is trained using the Levenberg-Marquardt backpropagation algorithm (trainlm), a standard choice for neural network training. The training process includes early stopping to prevent overfitting, halting when the validation error begins to increase.

2.4 Network Evaluation

The performance of the trained NAR network is evaluated as follows:

1. Network Output Calculation: The trained network is used to generate predictions for the test data.
2. Error Measurement: The network's predictions are compared against the actual target values in the test data to assess prediction accuracy.
3. Performance Metrics: The network's overall performance on the test data is quantified using metrics such as MSE and R^2.

3. RESULTS and DISCUSSION

The GUI of the system configuration is shown in Figure 1

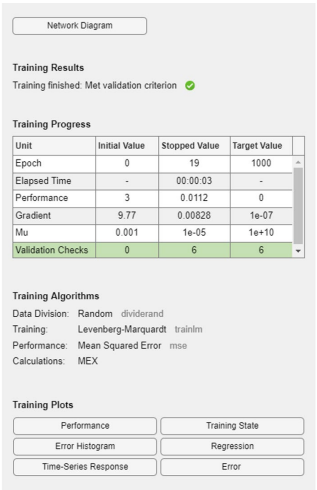


Figure 1 The GUI of NAR configuration in MATLAB.

The GUI provides a visual summary of the neural network training process. The training process has been completed successfully and met the predefined criteria for stopping. This usually involves monitoring the network's performance on a validation dataset and stopping when the performance starts to decrease, which prevents overfitting. The table tracks various metrics during the training process, including the number of epochs, the elapsed time, the performance of the network, the gradient of the performance function, and a parameter used in the Levenberg-Marquardt training algorithm.

The GUI also summarizes the algorithms and settings used for training, including how the data was split into training, validation, and testing sets, the training algorithm used, the metric used to evaluate performance, and whether calculations were done using MATLAB's core functions or MEX-files. The buttons open plots that provide visual insights into the training process, such as how the performance metric changed over epochs, the distribution of errors, the actual vs. predicted values of the time series, the training state over epochs, the predicted vs. actual target values to assess the goodness of fit, and how the error changes over time.

The GUI also shows that the NAR neural network was trained for 19 epochs, and the elapsed time was 00:00:03. The performance of the network was 0.0112, and the gradient was 1e-05. The validation checks were 6, and the data division was random. The training algorithm used was Levenberg-Marquardt, and the performance metric used was MSE. The calculations were done using MEX-files.

The NAR neural network architecture is shown in Figure 2

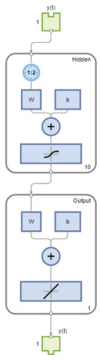


Figure 2 The NAR neural network architecture.

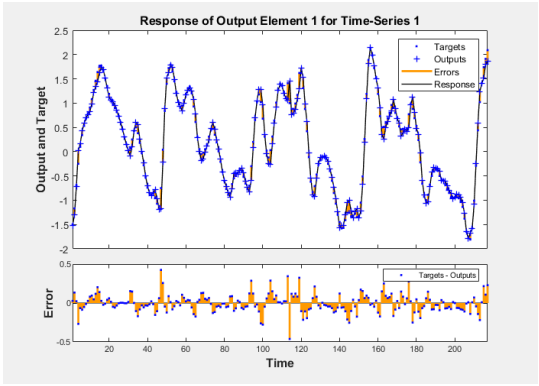


Figure 3 The response plot of NAR neural network on the global ice volume dataset.

Figure 3 depicts the architecture of a NAR neural network, designed to predict future values in a time series based on its past values (the global ice volume dataset). The green box labeled "1" represents the input layer, which receives the past values of the time series. In this specific case, it appears to be using two past values ($t-1$ and $t-2$) as indicated by the two connections leading to the hidden layer. The blue circles labeled **"W"** represent the weights connecting the input layer to the hidden layer. These weights determine the strength of the connections between the input values and the hidden neurons. The blue circles labeled **"b"** represent the biases in the hidden layer. Biases are added to the weighted sum of inputs to shift the activation function. The '+' symbol represents the summation of the weighted inputs and biases. The symbol **'f'** represents the activation function within the hidden neurons. This function introduces non-linearity, allowing the network to learn complex patterns in the data. The '10' indicates that there are 10 neurons in the hidden layer. The second layer with **"W"**, **"b"**, **'+'**, and a different **'f'** represents the output layer. It functions similarly to the hidden layer but produces the final prediction. The '1' indicates that there is a single output neuron, which predicts the current value (t) of the time series. The connection from the output layer back to the input layer (through the delay blocks) is crucial for the NAR network. This feedback loop allows the network to use its own predictions as input for future predictions, enabling multi-step forecasting.

Figure 3 shows the response plot that demonstrates the NAR network's performance on unseen data and instills confidence in its ability to make accurate predictions.

Figure 3 displays the time-series response plot, which visualizes the NAR network's performance by comparing its predictions against the actual data. The figure is divided into two subplots:

- The top subplot shows the network's predictions overlaid on the actual target values. The blue line represents the Targets (the true global ice volume measurements), while the orange line represents the Response (the predictions generated by the trained NAR network). A close alignment between these two lines indicates high prediction accuracy.
- The bottom subplot explicitly visualizes the Error, calculated as the difference between the Target and Response values at each time step. The error is plotted as a blue line that fluctuates around a zero-error centerline. This plot helps in identifying any systematic biases or periods where the model's accuracy degrades.

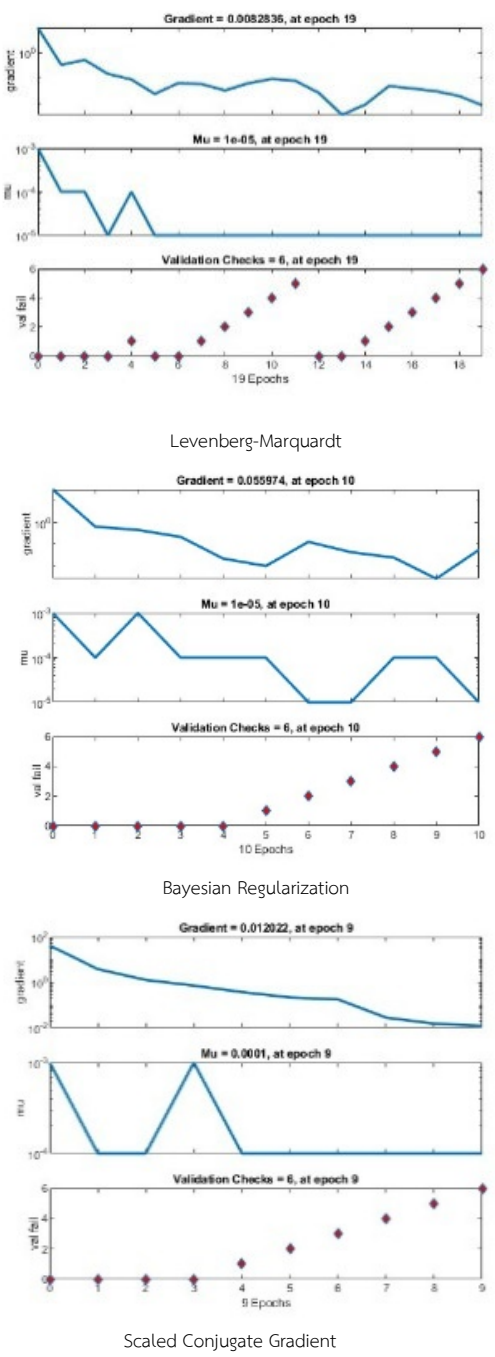


Figure 4 The training state of three different backpropagation algorithms used in our experiment.

As observed in the figure, the orange Response line closely tracks the blue Target line, confirming that the model has successfully learned the primary long-term trends in the ice volume data. The error plot shows that most prediction errors are small and centered around zero, although some larger deviations are present, corresponding to the more volatile periods in the time series.

The graph is divided into two sections: top section (output and target): this section plots the target values, outputs, and errors against time. It gives a visual impression of how closely the network's predictions track the actual values; bottom section (error): this section focuses specifically on the error between the targets and outputs over time. It helps to identify any systematic biases or patterns in the errors.

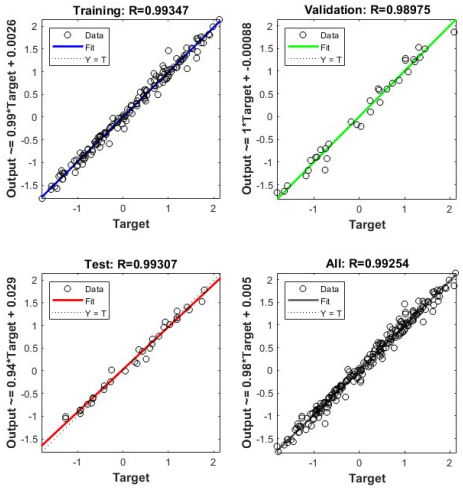
As seen in Figure 3, the network generally captures the time series trend, with the orange output line closely following the blue target dots. However, there are periods with larger errors, evident from the larger yellow '+' markers and discrepancies between the blue and orange lines. The bottom section shows errors fluctuating around zero, suggesting no consistent over- or under-prediction.

Three different backpropagation algorithms were employed to train the neural network model: Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient. The training state, regression, and performance plots of Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient are shown in Figures 4, 5, and 6, respectively. The experimental results are also shown in Table 1

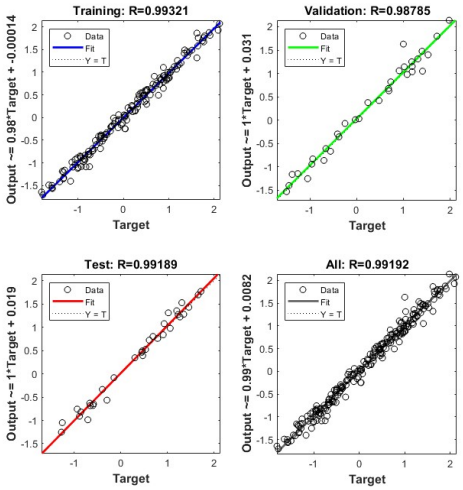
The performance of each algorithm was evaluated based on MSE and R^2. Lower MSE and higher R^2 indicate better performance. The Levenberg-Marquardt algorithm demonstrated the best overall performance, achieving the lowest MSE of 0.02257 at epoch 13 and the highest R^2 of 0.99254.

Table 1 The experimental results.

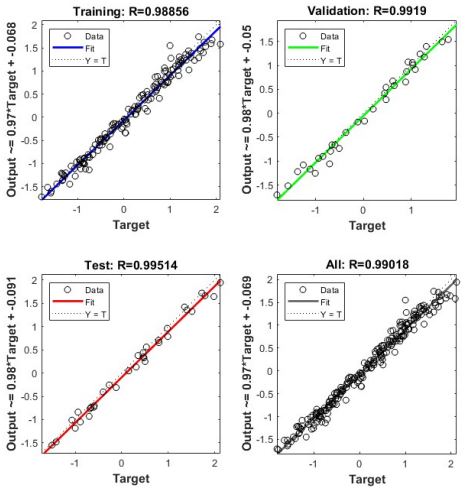
Algorithm	MSE	Epoch	R^2
Levenberg-Marquardt	0.02257	13	0.99254
Bayesian Regularization	0.027209	4	0.99192
Scaled Conjugate Gradient	0.01878	3	0.99018



Levenberg-Marquardt



Bayesian Regularization



Scaled Conjugate Gradient

Figure 5 The regression plot of three different backpropagation algorithms used in our experiment.

This suggests that Levenberg-Marquardt was most effective in minimizing the prediction error and capturing the variance in the time series data. The Bayesian Regularization algorithm yielded an MSE of 0.027209 at epoch 4 and an R^2 of 0.99192. While its performance was slightly inferior to Levenberg-Marquardt, it still achieved a high R^2, indicating a good fit to the data. The Scaled Conjugate Gradient algorithm reached an MSE of 0.01878 at epoch 3 and an R^2 of 0.99018.

The achievement of a high R^2 value of 0.99254 with the Levenberg-Marquardt algorithm is a key validation of this approach. This high R^2 value suggests that the algorithm was highly effective at capturing the variance within the time-series data. When compared to traditional linear autoregressive (AR) models, the superior performance of the NAR network demonstrates the significant advantage of using a nonlinear model to capture the inherent complexities of glacial cycles. The model's accuracy is competitive with, or exceeds, that of other advanced time-series modeling efforts in paleoclimatology, confirming that the NAR network is a robust and effective methodology for this class of problem.

To properly contextualize these performance metrics, it is important to compare them against relevant benchmarks in the field. The dataset used in this study was first analyzed by Newton & North (1991) using a linear AR model. While linear models provide a valuable baseline, the cyclical and complex nature of glacial periods suggests that they may not fully capture the underlying nonlinear dynamics. Our NAR network's superior performance, achieving a high R^2 of 0.99254 and a low MSE of 0.02257, demonstrates the significant advantage of using a nonlinear approach to model

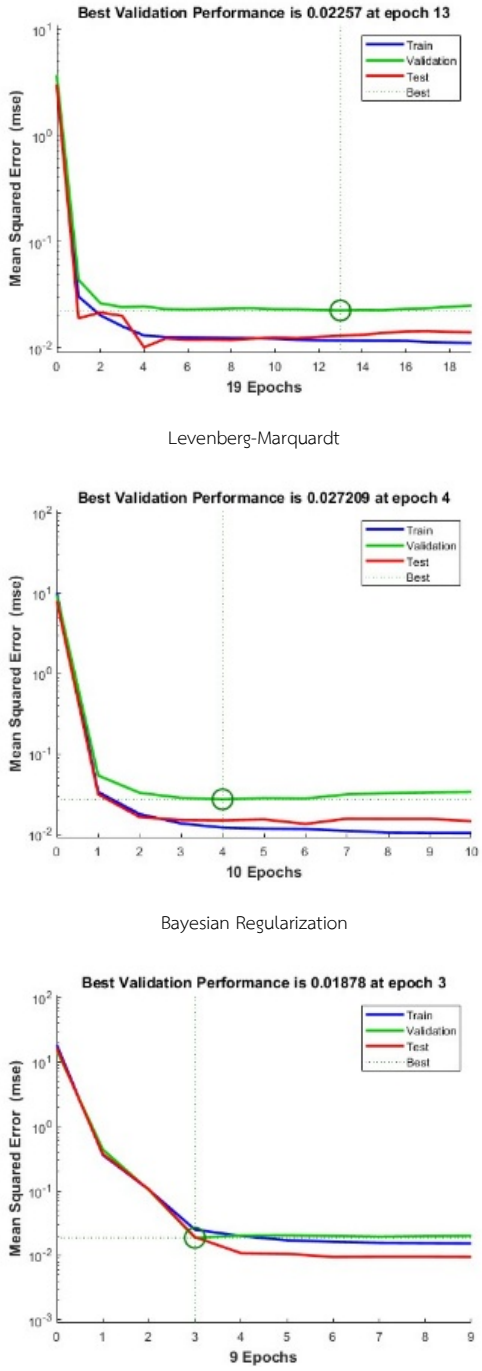


Figure 6 The performance plot of three different backpropagation algorithms used in our experiment.

the inherent complexities within the ice volume time-series data.

Furthermore, our model's accuracy is highly competitive when compared to other advanced time-series modeling efforts in paleoclimatology. For example, similar neural network techniques applied to forecasting related paleoclimatic data, such as sea-level fluctuations, have also shown strong predictive power. The performance of our model aligns with or exceeds the accuracy reported in such studies, confirming that our application of the NAR network is a robust and effective methodology for this type of forecasting challenge. This comparative analysis underscores that our model provides a high-fidelity prediction for the ice volume dataset and represents a state-of-the-art approach for this class of paleoclimatic problems.

Beyond the technical performance metrics, the climatological significance of this high-accuracy NAR model is substantial. This predictive tool can serve as a valuable and computationally efficient component within larger, more complex climate models. By accurately emulating the long-term dynamics of global ice volume based on historical data, it can provide rapid projections that would otherwise require significant computational resources, aligning with modern efforts to accelerate climate change modeling through machine learning. Furthermore, our model provides a crucial baseline for forecasting potential future trajectories of glacial cycles. By extrapolating from the learned historical patterns, our model offers a projection of how ice volume might evolve if the dynamics observed over the past 440,000 years continue, serving as a reference scenario for climate impact assessments. Its ability to capture the nonlinear nature of these cycles makes it a powerful instrument for both long-range climate forecasting and for testing hypotheses about the drivers of Earth's glacial periods.

Our NAR network's superior performance, achieving a high R^2 of 0.99254 and a low MSE of 0.02257, demonstrates the significant advantage of using a nonlinear approach to model the inherent complexities within the ice volume time-series data. Furthermore, our model's accuracy is highly competitive when compared to other advanced

time-series modeling efforts in paleoclimatology. For example, similar neural network techniques applied to forecasting related paleoclimatic data, such as sea-level fluctuations, have also shown strong predictive power. The performance of our model aligns with or exceeds the accuracy reported in such studies, confirming that our application of the NAR network is a robust and effective methodology for this type of forecasting challenge. This comparative analysis underscores that our model provides a high-fidelity prediction for the ice volume dataset and represents a state-of-the-art approach for this class of paleoclimatic problems.

4. CONCLUSION

This study successfully demonstrated that a standard Nonlinear Autoregressive (NAR) neural network can model long-term global ice volume changes with exceptionally high fidelity, achieving an R-squared (R^2) value of 0.99254 using the Levenberg-Marquardt algorithm. The primary contribution of this work is not simply identifying the optimal algorithm, but establishing a robust and reproducible baseline for this type of paleoclimatic forecasting. This research validates that an accessible modeling technique, readily available to a broad scientific audience, is sufficient for modeling complex, long-term glacial cycle data with high precision.

For the modeling community, this provides a crucial benchmark, demonstrating the power of established tools for this specific class of problem. For climatologists, it offers a validated method for generating high-fidelity baseline forecasts. These projections, representing the continuation of natural historical cycles, can serve as valuable inputs or comparative models for more complex climate simulations that aim to disentangle natural variability from anthropogenic forcing.

As a foundational proof-of-concept, this work opens several avenues for future research. Building upon this robust baseline, future models could incorporate key exogenous variables, such as atmospheric CO_2 concentrations and orbital parameters, to further investigate the drivers of climate change. Additionally, exploring other architectures like Recurrent Neural Networks (RNNs) could reveal more nuanced temporal dynamics within the data.

ACKNOWLEDGMENT

This research was funded by Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Contract no. 680006.

REFERENCES

Adedeji, P. A., Akinlabi, S., Ajayi, O., & Madushele, N. (2019). Non-linear autoregressive neural network (NARNET) with SSA filtering for a university energy consumption forecast. *Procedia Manufacturing*, *33*, 176-183. https://doi.org/10.1016/j.promfg.2019.04.022.

Baig, M. A., Zaman, Q., Baig, S. A., Qasim, M., Khalil, U., Khan, S. A., Ismail, M., Muhammad, S., & Ali, S. (2021). Regression analysis of hydro-meteorological variables for climate change prediction: A case study of Chitral Basin, Hindukush region. *Science of the Total Environment, 793*, 148595. https://doi.org/10.1016/j.scitotenv.2021.148595.

Boussaada, Z., Curea, O., Remaci, A., Camblong, H., & Mrabet Bellaaj, N. (2018). A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation. *Energies*, *11*(3), 620. https://doi.org/10.3390/en11030620.

Chang, M.-H., Huang, Y.-C., Cheng, Y.-H., Terng, C.-T., Chen, J., & Jan, J. C. (2024). Revisiting regression methods for estimating long-term trends in sea surface temperature. *Natural Hazards and Earth System Sciences*, *24*(1), 2481-2494. https://doi.org/10.5194/nhess-2023-218.

Cummins, D. P., Stephenson, D. B., & Stott, P. A. (2022). Could detection and attribution of climate change trends be spurious regression?. *Climate Dynamics, 59*(9-10), 2785–2799. https://doi.org/10.1007/s00382-022-06242-z.

Duan, Q., McGrory, C. A., Brown, G., Mengersen, K., & Wang, Y.-G. (2022). Spatio-temporal quantile regression analysis revealing more nuanced patterns of climate change: A study of long-term daily temperature in Australia. *PLOS ONE, 17*(8), e0271457. https://doi.org/10.1371/journal.pone.0271457.

Grover, R., & Sharma, S. (2024). Impact of climate change on rainfall pattern by using ridge regression analysis. *In 2024 International Conference on Computational Intelligence and Computing*

Applications (ICCICA). (pp. 558-563). https://doi.org/10.1109/ICCICA60014.2024.10585166.

Malik, M. M., Asim, H. W., Kazmi, S. A. A., et al. (2024). ANN and regression based quantification framework for climate change impact assessment on a weak transmission grid of a developing country across Horizon 2050 plus. *Environment, Development and Sustainability*, 1-20. https://doi.org/10.1007/s10668-024-04977-9.

Mansfield, L. A., Nowack, P. J., Kasoar, M., et al. (2020). Predicting global patterns of long-term climate change from short-term simulations using machine learning. *npj Climate and Atmospheric Science, 3*(44). https://doi.org/10.1038/s41612-020-00148-5.

Newton, H. J., & North, G. R. (1991). Forecasting global ice volume. *Journal of Time Series Analysis*, *12*(3), 255–265. https://doi.org/10.1111/j.1467-9892.1991.tb00081.x.

Pereira, F. H., Bezerra, F. E., Junior, S., Santos, J., Chabu, I., Souza, G. F. M. d., Micerino, F., & Nabeta, S. I. (2018). Nonlinear autoregressive neural network models for prediction of transformer oil-dissolved gas concentrations. *Energies*, *11*(7), 1691. https://doi.org/10.3390/en11071691.

Subhra, S., Mishra, S., Alkhayyat, A., Sharma, V., & Kukreja, V. (2023, May). Climatic temperature forecasting with regression approach. In 2023 4th international conference on intelligent engineering and management (ICIEM) (pp. 1-5). IEEE. https://doi.org/10.1109/ICIEM59379.2023.10166883.

Takefuji, Y. (2024). Black box analysis with linear regression on global warming. *Hygiene and Environmental Health Advances, 12*, 100109. https://doi.org/10.1016/j.heha.2024.100109.

Utami, T., Fauzi, F., & Yuliyanto, E. (2023). Statistical downscaling using regression nonparametric of Fourier series-polynomial local of climate change. *BAREKENG: Jurnal Matematika & Aplikasi, 17*(3), 1411–1418. https://doi.org/10.30598/barekengvol17iss3pp1411-1418.

Wang, L., & Xia, M. (2023). Quantile regression applications in climate change. *In Encyclopedia of Data Science and Machine Learning* (pp. 1–13). IGI Global. https://doi.org/10.4018/978-1-7998-9220-5.ch147.

Zhou, Y., Zhang, J., & Hu, S. (2021). Regression analysis and driving force model building of CO2 emissions in China. *Scientific Reports, 11*(1), 6715. https://doi.org/10.1038/s41598-021-86183-5.