Supporting Information

Nonlinear auto-regressive with exogenous inputs neural network

The error autocorrelation function serves as a tool for assessing the performance of the neural network (NN). It elucidates the temporal relationship between prediction errors. In an ideal prediction model, the autocorrelation function would exhibit a single non-zero value at zero lag, indicating that forecast errors are entirely independent of each other. This suggests that the prediction errors exhibit no correlation with each other (white noise). In practical terms, achieving absolute disparity is unattainable. When determining the extent of autocorrelation, it is advisable to set the confidence interval to 95%. For 1,801 data points, the autocorrelation coefficient peaks at around 6.68×10^{-9}, with the maximum value occurring at zero lag, while other values remain within the confidence interval (Fig. 2).

Mean square error (MSE), also referred to as mean square deviation, is computed by determining the variance between actual and predicted values. The performance graphs illustrate the convergence of the solutions obtained using the proposed technique toward the minimum MSE values across all three datasets: training, testing, and validation (Fig. 3). The MSE values depicted in the graph span from 10^{-9} to 10^{-10}, indicating a consistent performance of the NN. This suggests that it performs well for the large datasets.

Error histogram (Fig. 4) illustrates the distribution of errors (difference between targets/actual values and outputs/predicted values). Since these error values represent the disparity between predicted and target values, they may be negative. Bins refer to the vertical bars displayed on the graph, dividing the total error range into smaller segments; in this instance, 20 bins are utilized. The y-axis indicates the number of samples from the dataset falling within a particular bin.

Hyper-parameter gradient quantifies the rate of change in all weights concerning the change in error. It reflects the slope of a function. A higher gradient indicates a steeper slope and facilitates faster learning by the model. However, if the slope is zero, the model ceases to learn. A parameter, mu, acts as the control parameter and it is also called training gain for the algorithm used in training the NN. It accelerates the training process and mitigates oscillations during weight updates. Validation checks are employed to halt the learning process of the NN. The frequency of these checks will be determined by the number of successive iterations of the NN. (See hyper-parameters depicted in Fig. 5)

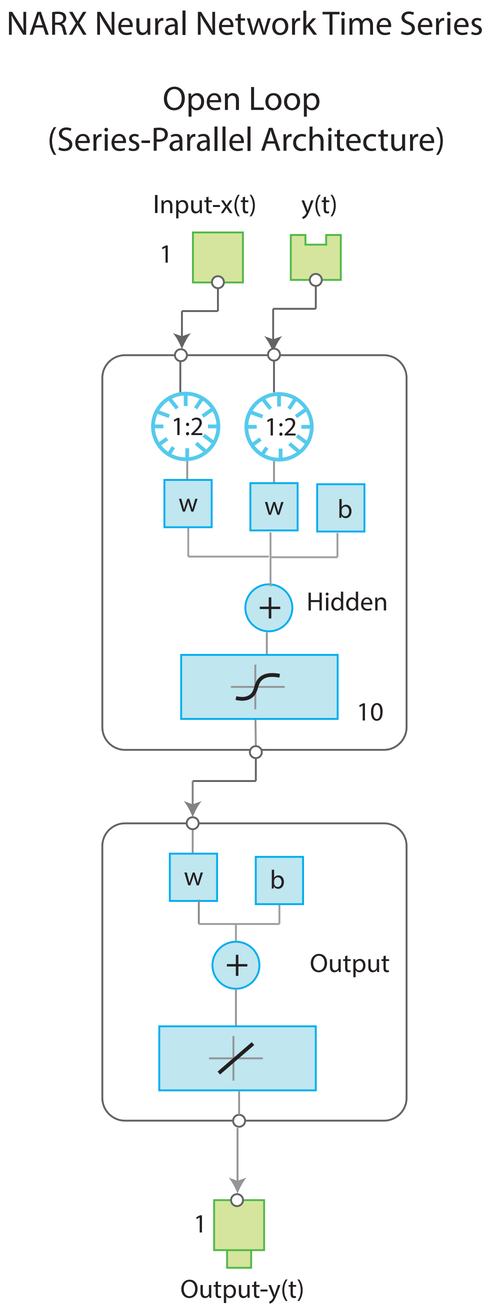

Fig. 1: Nonlinear auto-regressive with exogenous inputs neural network model.

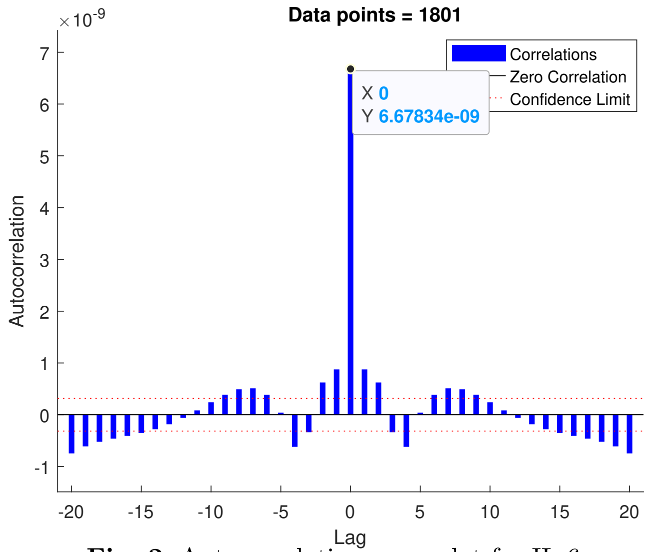

Fig. 2: Autocorrelation error plot for IL-6.

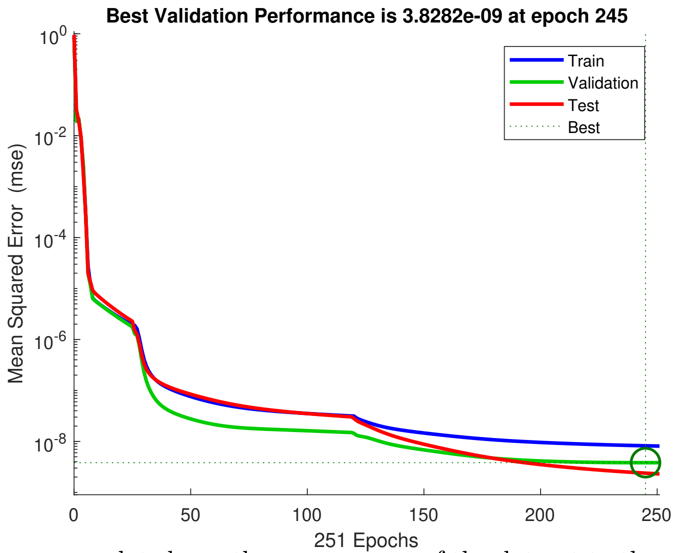

Fig. 3: Performance plot shows the convergence of the dataset to the minimum MSE.

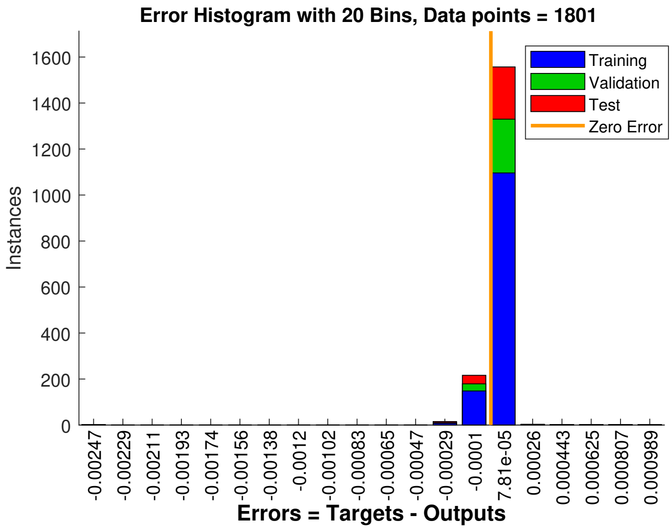

Fig. 4: Error histogram indicates the distribution of error of testing, validation, and training datasets.

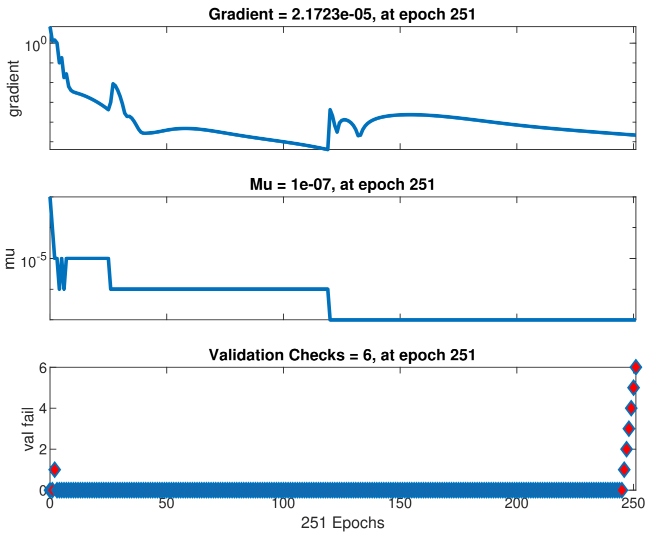

Fig. 5: Training graph depicted hyper-parameters like gradient, mu and validation.