

# Comparison of Machine Learning-Based Predictive Models of the Nutrient Loads Delivered from the Mississippi/Atchafalaya River Basin to the Gulf of Mexico

Y. Zhen, S. Yoo

To be published in "Water"

October 2024

Computational Science Initiative  
**Brookhaven National Laboratory**

**U.S. Department of Energy**

USDOE Office of Science (SC), Advanced Scientific Computing Research (ASCR)

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No.DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

## **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

# Comparison of machine learning based predictive models of the nutrient loads delivered from Mississippi/Atchafalaya River Basin to Gulf of Mexico

Yi Zhen<sup>1</sup>, Huan Feng<sup>2</sup>, Shinjae Yoo<sup>3</sup>

<sup>1</sup>Department of Natural Sciences, Southern University at New Orleans, New Orleans, LA 70126

<sup>2</sup>Department of Earth and Environmental Studies, Montclair State University, Montclair, NJ 07043

<sup>3</sup>Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973

**Abstract:** Predicting nutrient loads is essential to understand and manage one of environmental issues faced by the northern Gulf of Mexico-hypoxic zone which poses severe threats to Gulf's healthy ecosystem and economy. The development of hypoxia in Gulf of Mexico is strongly associated with eutrophication process initiated by excessive nutrient loads. Due to the complexities in the excessive nutrient loads to Gulf of Mexico, it is challenging to understand and predict underlying temporal variation of nutrient loads. The study was aimed at identifying an optimal predictive machine learning model to capture and predict nonlinear behavior of the nutrient loads delivered from the Mississippi/Atchafalaya River Basin (MARB) to Gulf of Mexico. For this purpose, monthly nutrient loads (N and P) in tons were used collecting from US Geological Survey (USGS) monitoring station 07373420 for a period of 1980 to 2020. The machine learning models including autoregressive integrated moving average (ARIMA), gaussian process regression (GPR), single-layer multilayer perceptron (MLP), and a long short-term memory (LSTM) with the single hidden layer were developed to predict the monthly nutrient loads and model performances were evaluated by standard assessment metrics- Root Mean Square Error (RMSE) and Correlation Coefficient (R). The residuals of predictive models have been examined by Durbin-Watson statistic. The results showed that MLP and LSTM persistently achieved the better accuracy in predicting monthly TN and TP loads compared to GPR and ARIMA. In addition, GPR models achieved slightly better test RMSE score than ARIMA models while their correlation coefficients are much lower than ARIMA models. Moreover, MLP performed slightly better than LSTM in predicting monthly TP loads while LSTM slightly outperformed for TN loads. Furthermore, it was found that the optimizer and number of inputs didn't show effects on the LSTM performance while they exhibit impact on MLP outcomes. This study explores the capability of machine learning models to accurately predict nonlinearly fluctuating nutrient loads delivered to Gulf of Mexico. Further efforts focus on im-proving the accuracy of forecasting using hybrid models which combine several machine learning models with superior predictive performance for nutrient fluxes throughout MARB.

Keywords: autoregressive integrated moving average (ARIMA); gaussian process regression (GPR); long short-term memory (LSTM); Mississippi/Atchafalaya River Basin (MARB); multilayer perceptron (MLP)

## 1. Introduction

Development of predictive models of nutrient loads may hold the key to understand and manage the dynamics of hypoxia in the northern Gulf of Mexico. Hypoxic zone (or dead zone) is an area where oxygen levels drop to 2 milligrams per liter or lower and most marine life can no longer survive due to depleted oxygen levels [1]. Hypoxia which occurs near the coastal Gulf of Mexico has attracted much attention because it poses a threat to the coastal ecosystem and economy by destroying coastal wetland habitats of numerous fish and wildlife species at an alarming rate [2]. Water quality is

deteriorated by increasing municipal and manufacturing needs [2]. The study shows that the formation of hypoxia in the northern Gulf of Mexico is closely related to the excess nutrients (N and P) from MARB [3]. It is desirable to forecast time series of nutrient loads to effectively manage and precisely control excess nutrient export from MARB. Time series of nutrient loads prediction, however, is complex and challenging due to its volatile and nonlinear fluctuations caused by the uncertain sources of nutrient loads relating to the changes in geomorphological natures of riverine and estuarine systems and anthropogenic activities [4-5].

Extensive efforts have been made to pinpoint the critical sources of N and P throughout the MARB, analyze trends of nutrient loads, identify the key influential factors in the nutrient fluxes, investigate the spatial correlations of the distributions of nutrient yields, and assess the effectiveness of nutrient reduction strategies using various statistical approaches [6-23]. The simulation techniques such as SPATIally Referenced Regression On Watershed attributes (SPARROW) models and The Soil and Water Assessment Tool (SWAT) are the most commonly used to describe loads/yields throughout the MARB, analyze the relative importance of various sources and evaluate the effectiveness of management practices [24-27]. ARIMA model has been prevalently applied in modeling economical and financial time series [28-31]. The working principle of ARIMA is to investigate the correlations among the time series observations and construct a model to describe the structures of the relationships. The patterns in the correlations are assumed stationary over the time and then used to predict the future values. However, assumption of stationarity poses limitations to the effectiveness of the ARIMA model because, in practice, the time series may be non-stationary [32]. Apart from its restrict assumption, it is hard for ARIMA model to capture nonlinear variations in the time series [32]. The advancement in machine learning architectures brings possibilities in developing sophisticated methods to predict time series with nonlinear behaviors. The time series forecasting problems have been addressed by various machine learning based approaches. For instance, algorithms of deep learning and random forests have been applied to predict stock time series [33]. Deep learning-based modeling has been reported to be under-performed gradient-boosted trees and random forests and it is difficult to train neural network. RNN-based approach is introduced to predict time series of stock returns [34]. RNN-based methods are per-formed for in predicting time series in financial market [34-35]. As a variant of RNN, LSTM is designed to cope with the vanishing gradient situation [36].

In this study, the computational frameworks including ARIMA, GPR, MLP, and LSTM were developed to predict monthly nutrient loads (N and P) from 1980 to 2020 and explore the optimal machine learning architecture for forecasts of time series of nutrient loads. Moreover, to make comparison to ARIMA and GPR, neural networks MLP and LSTM are restricted to be single hidden layered architecture. The study showed that the neural network LSTM and MLP persistently lead in forecasting competition as expected while GPR exhibits comparable performance better than multi-criteria selected ARIMA. Furthermore, although GPR achieves better accuracy than ARIMA, its R-squared score is the lowest while those of ARIMA are the highest. The article is organized as follows. Section 2 describes the study site, data sources and the mathematical formalism of ARIMA, GPR, MLP, and LSTM. Section 3 presents descriptive statistics of nutrient loads and predictive performances of ARIMA, GPR, MLP, and LSTM models of nutrient loads. Section 4 discusses the predictive models and concludes the paper.

## **2. Materials and methods**

Figure 1 shows the flowchart of the methodology, which will be explained in detail in the next subsections

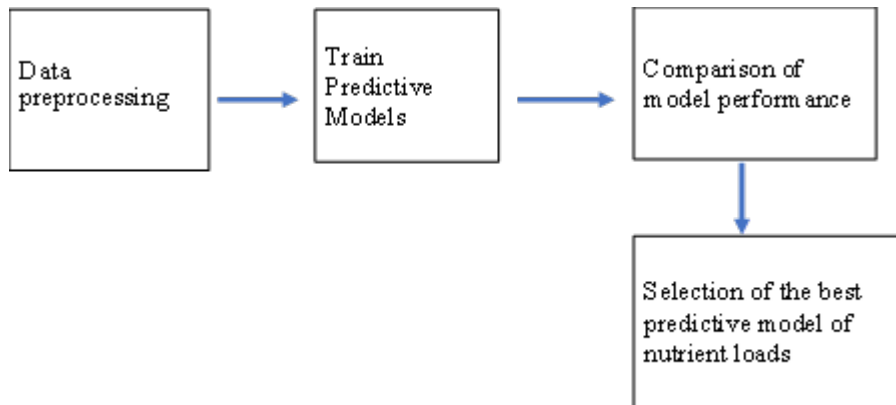


Figure 1. Schematic diagram of methodology

## 2.1 Study site and data

Monthly TN and TP (total nitrogen) and TP (total phosphorus) loads from US Geological Survey (USGS) gaging station 07373420 (Hydrologic unit: 08070100) were downloaded from the government public data source (<https://www.sciencebase.gov/catalog/item/61c08ec5d34ee9cd54ed3425>). The USGS 07373420 is located on the Mississippi River in West Feliciana Parish, Louisiana in the lower MARB, as shown in Figure 2. The nutrient loads data are presented monthly for a period of 1980 - 2020 and recorded based on the water year (12-month period from October 1 for a given year through September 30 of the following year). The monthly nutrient loads are reported in tons and used for characterization of trends and seasonal variation and development of predictive models.

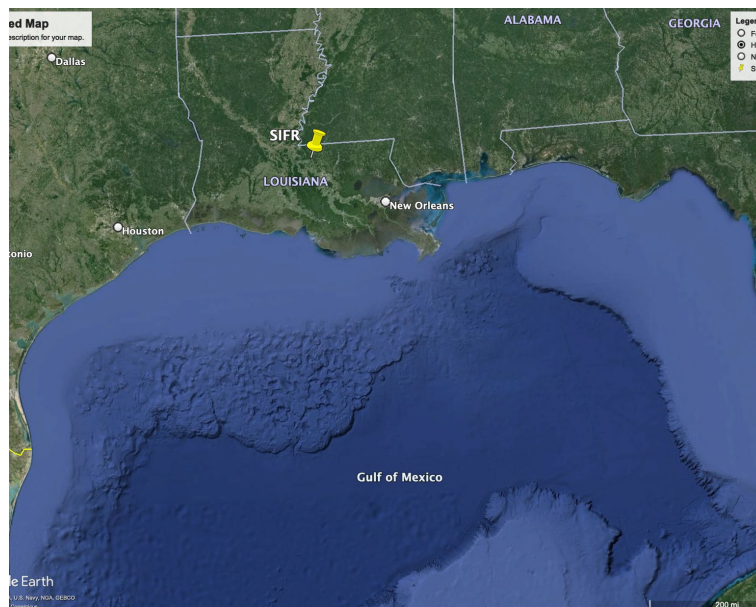


Figure 2. USGS water station 07373420 concerned in Mississippi/Atchafalaya River Basin indicated by yellow placeholder. Table 1: USGS water stations used in the study during 1980-2020 (<https://www.sciencebase.gov/catalog/item/61c08ec5d34ee9cd54ed3425>)

### 2.1.1 Data processing

The dataset was divided into training (80%) and testing data (20%). The training dataset was used for hyper-parameter tuning. The statistical software R was employed to process and analyze the data collected at the water sampling stations to identify optimal predictive ARIMA models of the nutrient loads. The analysis on MLP and LSTM uses python (3.10.8) programming environment along with TensorFlow and Keras APIs. The selections of optimal single-layer MLP and LSTM for forecasting nutrient loads are conducted in a Jupyter Notebook (6.5.2). Comparison among water quality data from different water stations were made.

### 2.1.2 Assessment Metric

ARIMA, GPR, single layer MLP, and single layer LSTM are implemented to predict the monthly TN and TP loads. Prediction accuracy is assessed by RMSE and R scores which measure the differences or residuals between actual and predicted values. The formula for computing RMSE and R are as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (1)$$

$$R = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}} \quad (2)$$

Where N is the total number of observations,  $y_i$  is the actual observation,  $\hat{y}_i$  is the predicted value,  $\bar{y}$  is the average observation, and  $\bar{\hat{y}}$  is average prediction. The RMSE score is used as a primary model selection criterion followed by R scores and significance of residuals due to that RMSE is affected by large errors and scaled in the same units as the forecast values (i.e., tons per month for this study). The smallest RMSE along with the greatest possible R and insignificant residuals indicate the best model. The randomness of the residuals is examined by Durbin-Watson statistic. Comparison of performance among predictive models ARIMA, GPR, MLP, and LSTM are made.

### 2.2. Description of ARIMA model

ARIMA [31] is a composite model of the time series observations. The model consists of autoregressive component of order p (AR(p)) which describes the linear dependencies of an observation on p lagged observations and moving average component of order q (MA(q)) which is used to account for the dependency between observations and q lagged error terms. Integrated step is to convert a non-stationary time series into a stationary by differencing the time ordered observations (d). The general ARIMA (p, d, q) model is expressed as

$$\phi(B)(1 - B)^d Y_t = \theta(B)e_t \quad (3)$$

Where  $Y_t$  is the observation, B is backshift operator  $BY_t = Y_{t-1}$ ,  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  is a non-seasonal AR (p),  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  is a non-seasonal MA(q), and  $e_t$  is the white noise term which follows Gaussian distribution with mean zero and variance  $\sigma_e^2 > 0$ . With seasonal time series data, the seasonal factors are introduced by a multiplicative seasonal ARIMA model. The

mathematical expression of a seasonal ARIMA model with seasonal period S denoted ARIMA(p, d, q) × (P, D, Q)<sub>s</sub> is given as,

$$\phi(B)(1 - \Phi_1 B^S - \dots - \Phi_P B^{SP})(1 - B)^d(1 - B^S)^D Y_t = \theta(B)(1 - \theta_1 B^S - \dots - \theta_Q B^{SQ})e_t$$

(4)

where P is the seasonal AR order, D is the seasonal differencing, Q is the seasonal MA order,  $\Phi$  is a seasonal AR coefficient, and  $\theta$  is a seasonal MA coefficient.

### 2.3. Description of GPR model

GPR is to use optimal probabilistic model in which the relationships between the predictor variables and response variables are described by Gaussian distribution [37]. The predictive model is developed based on the assumption that the function to be learned follows a Gaussian process. Explicitly, let x be the input variable, the prediction from GPR model can be found by:

$$f(x) \sim \mathcal{N}(\mu(x), k(x, x)) \quad (5)$$

where f is the function to be learned,  $\mathcal{N}$  is Gaussian distribution,  $\mu$  is the mean function, and k is variance matrix known as kernel function. The GPR model uses Bayesian posterior distribution to predict unseen observations.

### 2.4. Description of MLP model

As a feedforward neural network, MLP is used to learn the relationship between the input and output variables [38]. The architecture consists of an input layer where is fed with observations of input variables, hidden layers, and an output layer with value of output variables. The neurons in two consecutive layers are fully connected through weighted combinations and a nonlinear activation function will be applied to the values of the neurons before they are fed forward into the next layer in order to introduce non-linearity to the model. Explicitly, the value of the i-th neuron in the n-th layer  $z_i^n$  is related to the activations in the (n-1)-th layer by

$$z_i^n = \sigma(\sum_m w_{im}^n z_m^{n-1} + b_i^n) \quad (6)$$

where  $\sigma$  is activation function,  $w_{im}^n$  is a weight relating activation of m-th neuron in (n-1)-th layer to the i-th neuron in the n-th layer, and  $b_i^n$  is a bias for n-th layer. The learning process is to seek optimal weights and bias which minimize the error function of the predicted and actual observations.

$$\min_{w,b} \text{Error}(y, y) \quad (7)$$

where  $w$  are weights,  $b$  are biases,  $y$  are actual observations, and  $\hat{y}$  are outputs from MLP. In this study, the final optimal model architecture of the single layer MLP is selected by exploring a wide range of possibilities. Six models with 10, 30, 50, 100, 150, and 200 neurons are experimented with optimizers- adaptive moment estimation (Adam), root mean squared propagation (Rmsprop), and Nesterov-accelerated adaptive moment estimation (Nadam); the epoch number 10, 20, 50, 100, 150, 200, 250, 300, and 350; the batch sizes 1, 2, 4, 8, 16, 32; and the the number of inputs 1, 2, 3, 4, 5, 6. The default learning rate 0.01 and activation function Relu are used for entire study. Each combination of hyperparameters is executed 30 times.

## 2.5 Description of LSTM model

LSTM [39-41] is a variant of Recurrent Neural Network (RNN) capable of remembering the observations from earlier stages for the future use. LSTM model consists of input layer which takes in time sequence of data, output layer which generates predicted values, and LSTM layers in between to produce hidden states ( $h_t$ ) that is short term memory and cell state ( $c_t$ ) which is long term memory to encode the most information from the input sequence. The architecture of LSTM layer consists of four gates to update hidden state and cell state. Explicitly, for a given input at time  $t$  ( $x_t$ ), the hidden state at previous step ( $h_{t-1}$ ), weight ( $W$ ), and bias ( $b$ ), three gates: input gate  $i_t = \sigma(W_i x_t + W_{hi} h_{t-1} + b_i)$ , forget gate  $f_t = \sigma(W_f x_t + W_{hf} h_{t-1} + b_f)$  and change gate  $c_t = \sigma(W_c x_t + W_{hc} h_{t-1} + b_c)$ , where  $\sigma$  is the sigmoid function are used to dispose, filter, and add the information into the memory. Taking account of the results from three gates, the memory cell state  $c_t$  is updated by

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \quad (8)$$

where the operator  $\otimes$  is the element-wise product. Using the output gate  $o_t$  and current memory cell state  $c_t$ , the current hidden state  $h_t$  is given by

$$h_t = o_t \otimes \tanh(c_t) \quad (9)$$

where  $o_t = \sigma(W_o x_t + W_{ho} h_{t-1} + b_o)$  and  $\tanh$  represents hyperbolic tangent function. The iterations continue until the final corresponding output hidden state  $h_f$  is reached. Then, the final hidden state  $h_f$  is used to make a prediction from LSTM.

## 3.Results

### 3.1 Annual variation

The monthly loads of total nitrogen (TN) and total phosphorus (TP) at USGS monitoring station 07373420 from 1980 to 2020 are shown in Figure 3 and Figure 4, respectively. As the Figure 3 shows, the mean monthly TN loads from 1980 to 2020 observed at Station 07373420 was  $99323.61 \pm 23683.35$  tons/month with a range of 54308.33 to 153916.67 tons/month. The mean monthly TN loads slightly decreased from 109037.50 tons/month in 1980s to 105543.33 tons/month in 1990s, and significantly declined to 83910.83 tons/month in 2000s, and then slightly increased to 98472.73 tons/month in 2010s. The peak mean monthly loads appeared in the 1980s, showing a decreasing trend in the last three decades. From Figure 4, it can be seen that the mean monthly TP loads from 1980 to 2020 observed at Station 07373420 was  $9830.77 \pm 2217.51$  tons/month with a range of 4791.67 to 16003.33 tons/month. The mean monthly TP loads increased from 8869.00 tons/month in 1980s to 9263.42 tons/month in 1990s, and slightly continuously increased to 9749.67 tons/month in 2000s, and then significantly increased to 11294.62 tons/month in 2010s. The peak mean monthly loads appeared in the 2010s, showing a persistent increasing trend in the last four decades. The pattern of variations in TN implies that excessive nutrient fluxes are manageable through collective efforts [2].

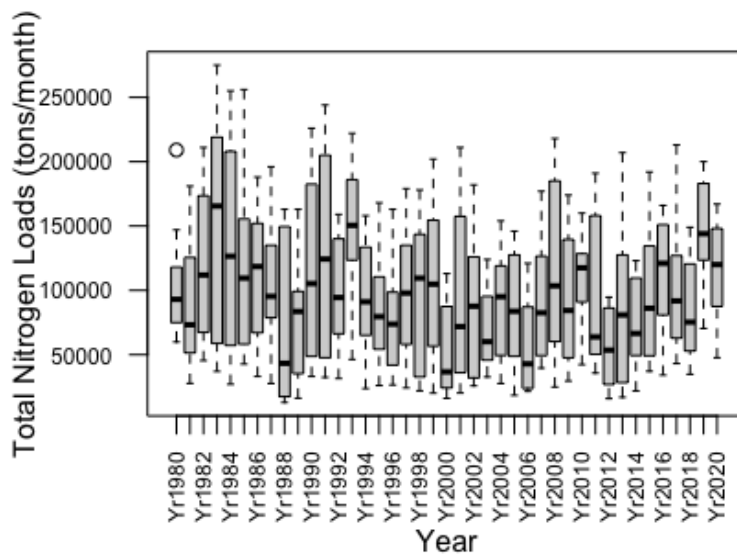


Figure 3. Boxplot of monthly total nitrogen from 1980 to 2020

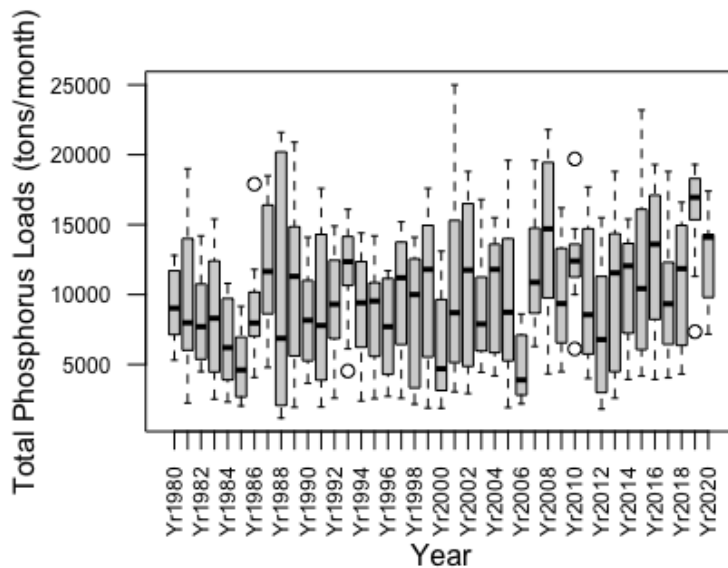


Figure 4. Boxplot of monthly total phosphorus from 1980 to 2020

### 3.2 Seasonal Variations

The seasonal variations in the mean monthly TN and TP loads at Station 07373420 are shown in Figure 5 and Figure 6. As shown in Figure 5 and Figure 6, the mean monthly nutrient loads of autumn were lowest while those in spring were highest. Among the four seasonal periods, the mean monthly TP nutrient loads was in the order of spring > winter > summer > autumn while the mean monthly TN loads in summer was slightly higher than that in winter. Similar to TP loads, TN loads in spring consistently exceeded that in the other three seasons. A relatively high mean monthly TN loads (154659.35 tons) was observed in spring, followed by summer (99542.28 tons), winter (94913.82 tons), and autumn (47773.98 tons) while the highest mean monthly TP loads in spring was 13273.90 tons, followed by winter (10680.24 tons), summer (9472.11 tons), and autumn (5896.83 tons). The mean monthly TN and TP loads for different seasons at Station 07373420 were calculated using the data from 1980 to 2020 sampled in springs, summers, autumns, and winters. The contributions of different season nutrient loads to the total nutrient loads varied considerably. TN nutrient loads in spring made up the largest proportion (39%), followed by summer (25%), winter (24%), and autumn (12%), while seasonal TP loads was in the sequence of spring (34%), winter (27%), summer (24%), and autumn

(15%).

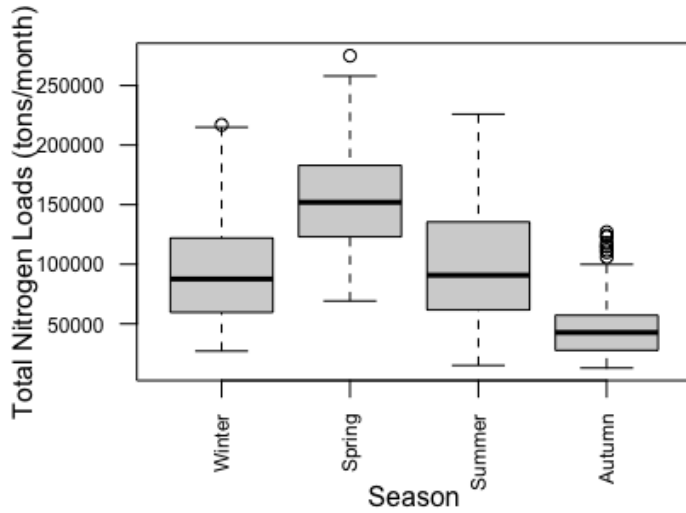


Figure 5. Box plot of monthly total nitrogen from 1980 to 2020 by seasons

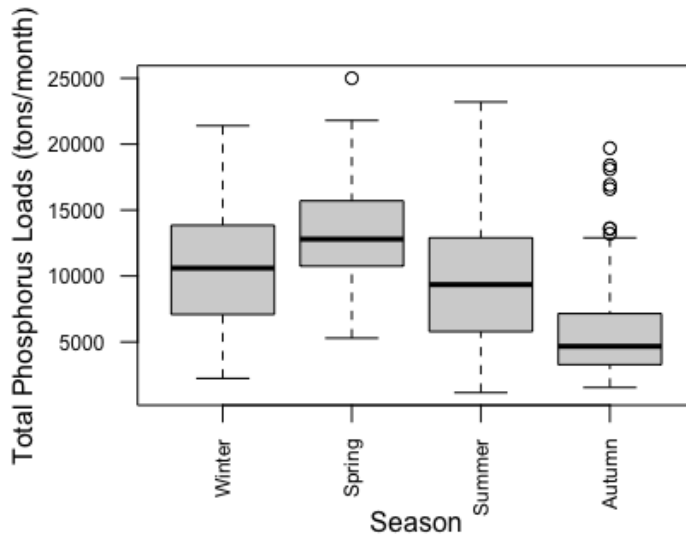


Figure 6. Box plot of monthly total phosphorus from 1980 to 2020 by seasons

### 3.3 Nutrient loads prediction

#### 3.3.1 ARIMA prediction

An autoregressive integrated moving average (ARIMA) models were constructed for the TN and TP nutrient loads forecast, respectively. For the monthly TN loads, Augmented Dickey-Fuller test detected stationarity in the data set ( $p = 0.01$ ). Therefore, ARIMA is suitable. The decomposition analysis suggests a seasonal component for the model. Based on several selecting criteria including results of the autocorrelation function (ACF), the partial autocorrelation function (PACF), significances of coefficients (SC), Akaike information criterion (AIC) values, residuals of the fitted model (RF) and test errors (TE) of all possible models, ARIMA  $(1,0,0) \times (0,1,1)_{12}$  is selected as the final model which is expressed as follows:

$$(1 - \phi_1 B)(1 - B^S)(X_t - \mu) = (1 + \theta_1^S B^{1,S})w_t \quad (9)$$

where B is backshift operator,  $B^S$  is seasonal backshift operator,  $\phi_1 = 0.6486$  is the coefficient of AR (1) term and  $\theta_1^S = -0.9261$  is the coefficient of seasonal MA (1) term. The coefficients in the final model are all highly significant ( $p < 0.0001$ ). The residuals

of ARIMA  $(1,0,0) \times (0,1,1)_{12}$  for TN loads are shown in figure 7. The residuals plot indicates that the multicriteria based model captures the time series attributes of observations. The monthly TN forecasting load is shown in figure 8 along with observations in the testing dataset.

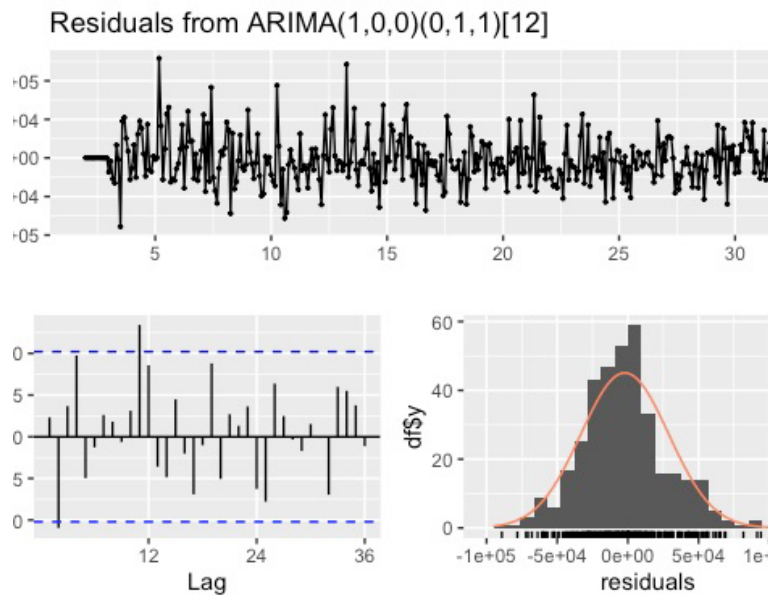


Figure 7. The residuals of ARIMA  $(1,0,0) \times (0,1,1)_{12}$  for TN loads

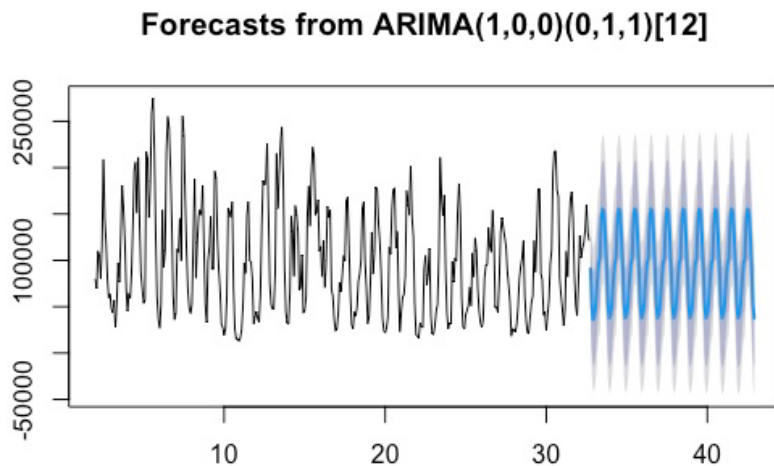


Figure 8. The predictions from ARIMA  $(1,0,0) \times (0,1,1)_{12}$  for TN loads. The blue line represents the predictions of TN loads.

For TP loads, Augmented Dickey-Fuller Test confirms the stationarity of the data set ( $p = 0.01$ ). The final model is defined as ARIMA  $(1,0,0) \times (5,1,0)_{12}$

$$(1 - \phi_1 B)(1 - \phi_1^S B^{1,S} - \dots - \phi_5^S B^{5,S}(1 - B^S))(X_t - \mu) = w_t \quad (10)$$

where B is backshift operator,  $B^S$  is seasonal backshift operator,  $\phi_1=0.5803$  is highly significant coefficient of AR (1) ( $p<0.0001$ ).  $\phi_1^S = -0.7636$  ( $p<0.0001$ ),  $\phi_2^S = -0.6014$  ( $p<0.0001$ ),  $\phi_3^S = -0.3649$  ( $p<0.0001$ ),  $\phi_4^S = -0.3152$  ( $p<0.0001$ ),  $\phi_5^S = -0.1508$  ( $p<0.001$ ) are the coefficient of seasonal AR(1), AR(2), AR(3), AR(4), AR(5) terms, respectively. The residuals of ARIMA  $(1,0,0) \times (5,1,0)_{12}$  are shown in figure 8. From figure 9, it can be seen that the model captured the characteristics of time series observations of monthly TP loads. The conclusion of non-correlated residuals is

supported by the Durbin-Watson statistic of 1.7801 which falls between 1.5 and 2.5. The forecasting values of loads are presented along with original observations in the testing dataset in figure 10.

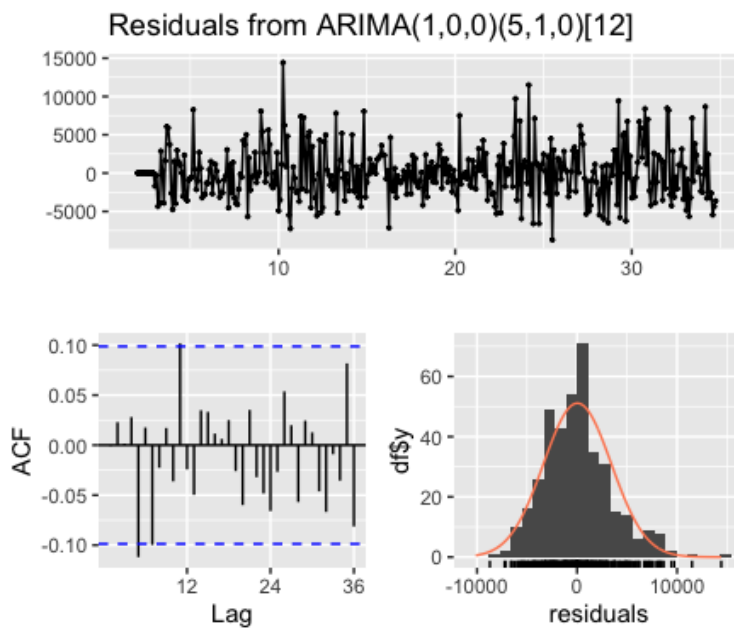
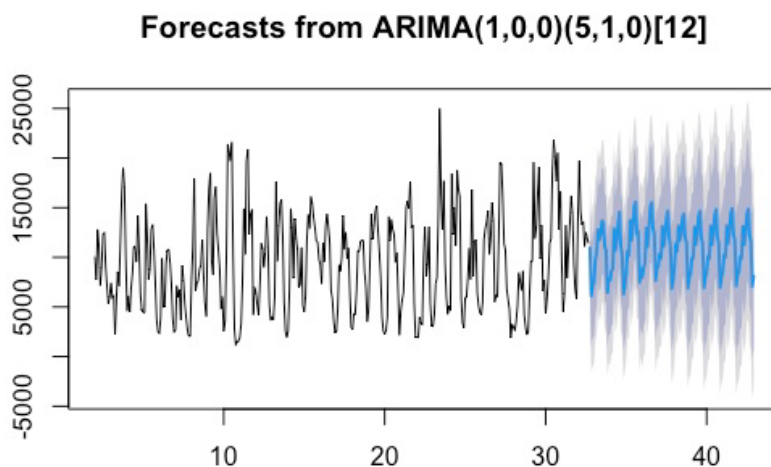


Figure 9. The residuals of ARIMA  $(1,0,0) \times (5,1,0)_{12}$  for TP loads



**Figure 10.** The forecasts of TP loads from ARIMA  $(1,0,0) \times (5,1,0)_{12}$ . The blue solid line represents the predictions of TP loads.

### 3.3.2 GPR prediction

The final optimal GPR model is selected by exploring all kernel functions available. The best combination of hyper parameters and kernel functions for the predictions of TN and TP loads are determined by the lowest RMSE score. To reach its best forecasting performance, the kernel chosen is a combination of linear kernel, radial basis function kernel, and exponential sine squared kernel. The optimal set of hyper parameters (the variance and length scale) is determined using cross-validation. The predictions of TN and TP loads from optimal GPR model are shown in Figures 11 and

12, respectively.

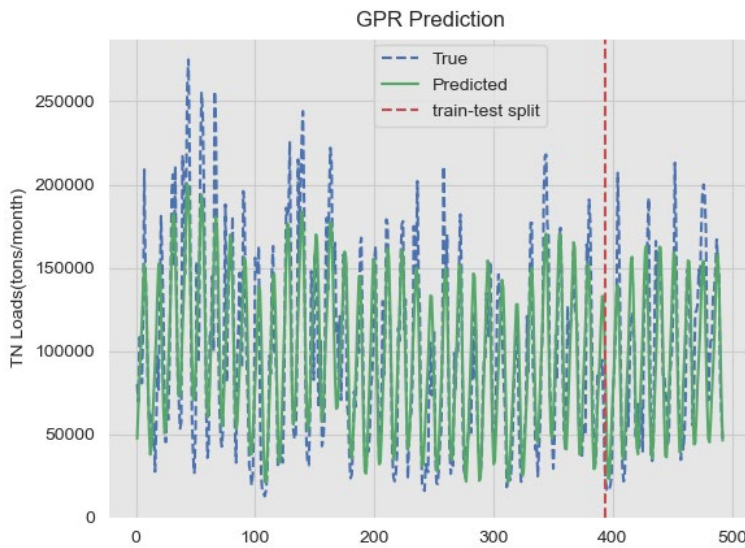


Figure 11. The predictions from best single layer GPR for TN loads

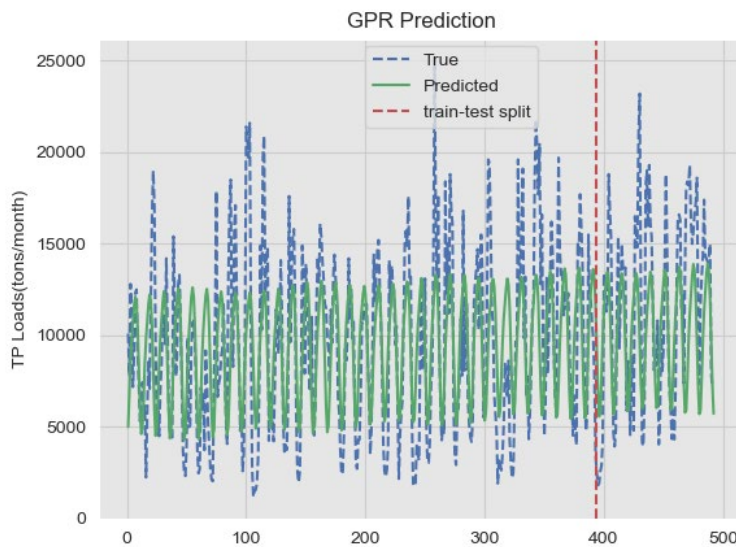


Figure 12. The predictions from best single layer GPR for TP loads

### 3.3.3 MLP prediction

Table 1 presented the best hyperparameters for all six single layer MLP models to predict TN loads. Table 1 suggests that the model with 50 neurons outperforms in RMSE of test data set while all six models present the similar R as shown in Table 1. Thus, a single layer MLP model with 50 neurons may be considered as the most likely best model to forecast TN loads. For the predictive model of TP loads, Table 2 suggests that the model with 200 neurons and the lowest RMSE can be considered as the winner among other candidates. From table 1, it can be seen that RMSE decreases on the range of neurons from 10 to 50, and then increases from 100 to 150, and then slightly decreases from 150 to 200. Table 2 shows the steadily decrease of RMSE from 10 to 200. All Durbin-Watson statistics ranging between 1.5 and 2.5 indicate that there are no significant presence of autocorrelations in the residuals of models. Fig. 13 and Fig. 14 represent the original TN and TP loads together with predictions obtained from the reproductions with the lowest RMSE score of the best single layer MLP model, respectively. In Fig. 13 and Fig. 14, the dotted curve represents the actual

values, whereas the solid curves represent the predictions in the test data. It can be seen that the prediction curves of over the test data set almost match with the curve of the actual observations of TN and TP loads. This implies that the optimal MLP model can learn nonlinear variations of the TN and TP loads quite well.

. Table 1. List of the best hyper parameters for single-layer MLP model for predictions of TN loads

| No. of Neurons | Optimizer | No. of Epoch | Batch Size | No. Of Inputs | Average Test RMSE | Average R score | The Durbin-Watson Statistic of Residuals |
|----------------|-----------|--------------|------------|---------------|-------------------|-----------------|--|
| 10             | Rmsprop   | 50           | 1          | 6             | 29779.76          | 0.656           | 1.7493                                   |
| 30             | Nadam     | 50           | 1          | 5             | 29690.46          | 0.644           | 1.6609                                   |
| 50             | Adam      | 50           | 1          | 5             | <b>29454.62</b>   | <b>0.661</b>    | <b>1.8440</b>                            |
| 100            | Nadam     | 50           | 1          | 5             | 29463.61          | 0.661           | 1.8209                                   |
| 150            | Adam      | 150          | 1          | 2             | 31195.69          | 0.619           | 1.7408                                   |
| 200            | Adam      | 50           | 1          | 2             | 29614.52          | 0.620           | 1.7352                                   |

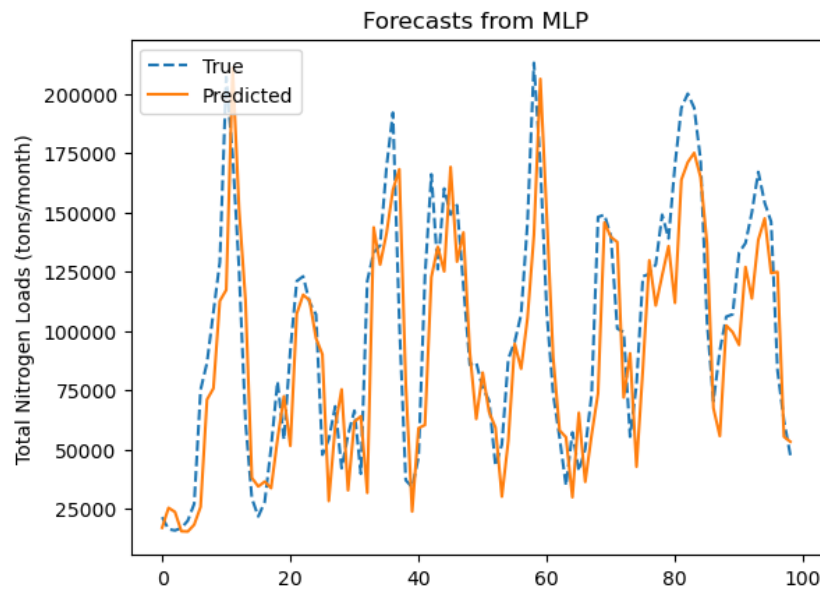


Figure 13. The predictions from best single layer MLP (optimizer: Adam; neurons: 50; learning rate: 0.01; inputs: 5; epoch: 50; batch size: 1) for TN loads

Table 2. List of the best hyper parameters for single layer MLP model for predictions of TP loads

| No. of Neurons | Optimizer    | No. of Epoch | Batch Size | No. Of Inputs | Average Test RMSE | Average R score | The Durbin-Watson Statistic of Residuals |
|----------------|--------------|--------------|------------|---------------|-------------------|-----------------|--|
| 10             | Adam         | 100          | 2          | 6             | 3692.70           | 0.485           | 1.9298                                   |
| 30             | Adam         | 100          | 1          | 5             | 3692.94           | 0.477           | 1.8508                                   |
| 50             | Adam         | 100          | 2          | 5             | 3682.14           | 0.488           | 1.8604                                   |
| 100            | Adam         | 50           | 4          | 5             | 3661.15           | 0.486           | 1.8574                                   |
| 150            | Adam         | 50           | 2          | 5             | 3659.85           | 0.492           | 1.8736                                   |
| <b>200</b>     | <b>Nadam</b> | <b>50</b>    | <b>8</b>   | <b>6</b>      | <b>3645.58</b>    | <b>0.492</b>    | <b>1.8964</b>                            |

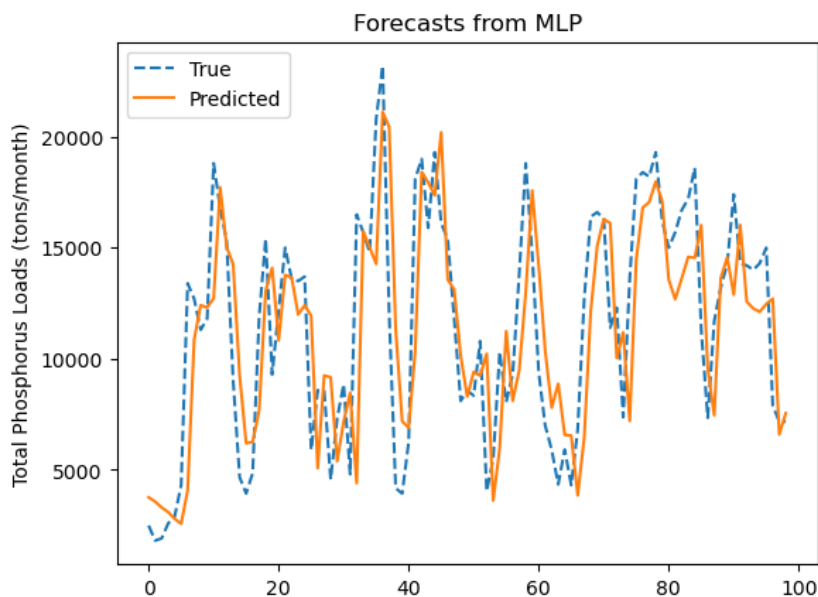


Figure 14. The predictions from the best single layer MLP (optimizer: Nadam; neurons: 200; learning rate: 0.01; inputs: 6; epoch: 50; batch size: 8) for TP loads

### 3.3.4 LSTM prediction

Table 3 presented the best hyperparameters for all six single layer LSTM models to predict TN loads. The results indicate that the model with 100 neurons outperforms in RMSE while all six models present the similar R score. Thus, a single layer LSTM model with 100 neurons may be considered as the most likely best model to forecast TN loads. For the predictive model of TP loads, Table 4 suggests that the model with 30 neurons and the lowest RMSE can be considered as the winner among other candidates. From table 3, it can be seen that RMSE decreases over neurons ranging from 10 to 100, and then rises from 100 to 200. Table 4 shows the decrease of RMSE from 10 to 30 and stay around 3730 from 30 to 200. The Durbin-Watson statistics of all models fall between 1.5 and 2.5 and indicate that there is no significant autocorrelation detected in the residuals. Fig. 15 and Fig. 16 represent the original TN and TP loads together with predicted values obtained from the reproductions of the optimal single-layered LSTM with the lowest test RMSE score, respectively. In Fig. 15 and Fig. 16, the dotted curve represents the actual observations, whereas the solid

curves represent the predictions in the test data. It can be seen that the predictions of the TN and TP loads in the test data closely follow the variations of the true observations. This indicates that the LSTM model is capable of capturing nonlinear patterns in the variations of the nutrient loads

Table 3. List of the best hyper parameters for single layer LSTM model for predictions of TN loads

| No. of Neurons | Optimizer   | No. of Epoch | Batch Size | No. Of Inputs | Average Test RMSE | Average R score | The Durbin-Watson Statistic of Residuals |
|----------------|-------------|--------------|------------|---------------|-------------------|-----------------|--|
| 10             | Nadam       | 100          | 1          | 4             | 28199.36          | 0.685           | 1.5490                                   |
| 30             | Nadam       | 100          | 1          | 4             | 27298.64          | 0.707           | 1.5024                                   |
| 50             | Adam        | 100          | 4          | 4             | 27560.25          | 0.710           | 1.7185                                   |
| <b>100</b>     | <b>Adam</b> | <b>100</b>   | <b>8</b>   | <b>4</b>      | <b>27251.68</b>   | <b>0.707</b>    | <b>1.6255</b>                            |
| 150            | Adam        | 100          | 1          | 4             | 28003.56          | 0.695           | 1.6555                                   |
| 200            | Nadam       | 100          | 1          | 4             | 28015.95          | 0.696           | 1.7363                                   |

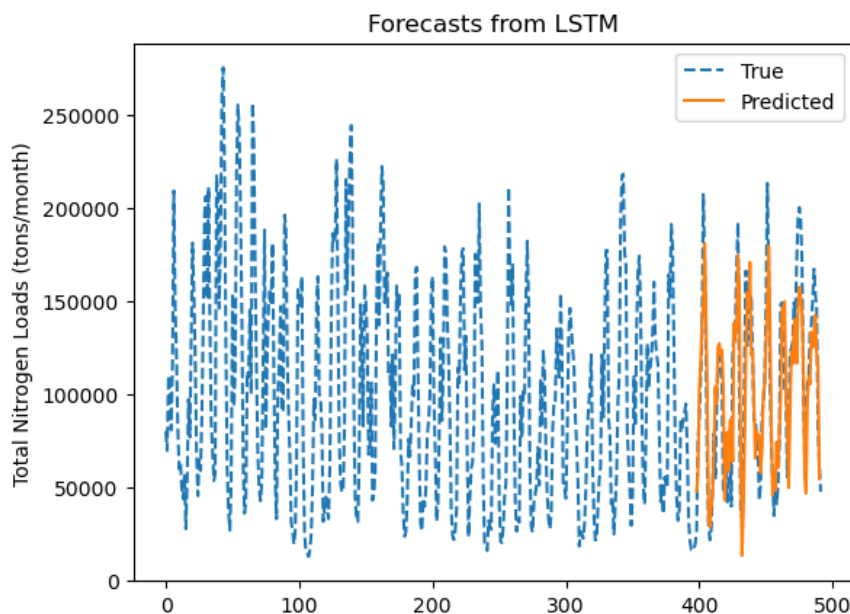


Figure 15. The predictions from best single layer LSTM (optimizer: Adam; neurons: 100; learning rate: 0.01; inputs: 4; epoch: 100; batch size: 8) for TN loads in the test dataset.

Table 4. List of the best hyper parameters for single layer LSTM model for predictions of TP loads

| No. of Neurons | Optimizer    | No. of Epoch | Batch Size | No. Of Inputs | Average Test RMSE | Average R score | The Durbin-Watson Statistic of Residuals |
|----------------|--------------|--------------|------------|---------------|-------------------|-----------------|--|
| 10             | Nadam        | 200          | 4          | 2             | 3748.94           | 0.484           | 1.7189                                   |
| <b>30</b>      | <b>Nadam</b> | <b>100</b>   | <b>1</b>   | <b>2</b>      | <b>3684.78</b>    | <b>0.482</b>    | <b>1.5469</b>                            |
| 50             | Rmsprop      | 100          | 1          | 5             | 3742.03           | 0.482           | 1.6803                                   |
| 100            | Rmsprop      | 150          | 1          | 2             | 3734.91           | 0.469           | 1.8065                                   |
| 150            | Adam         | 100          | 4          | 2             | 3704.34           | 0.486           | 1.5498                                   |
| 200            | Rmsprop      | 50           | 1          | 1             | 3737.01           | 0.479           | 1.5694                                   |

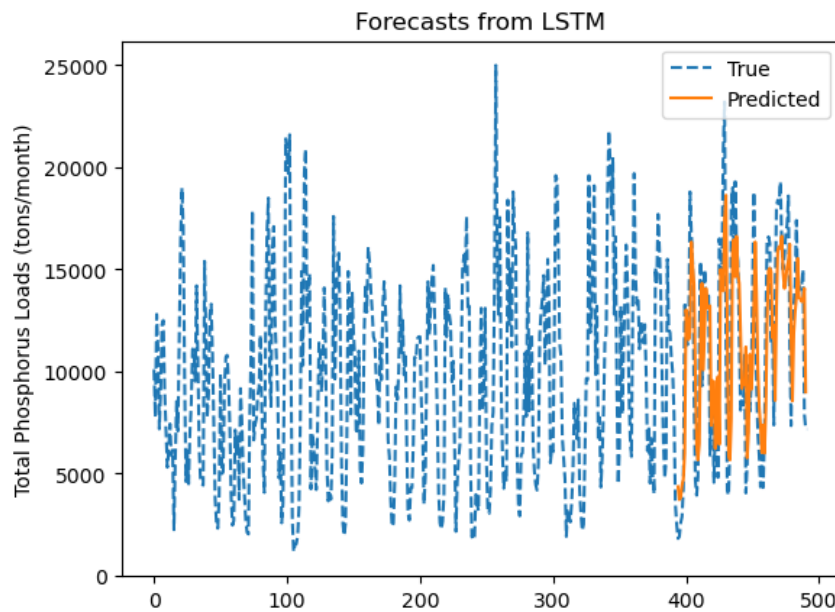


Figure 16. The predictions from best single layer LSTM (optimizer: Nadam; neurons: 30; learning rate: 0.01; inputs: 2; epoch: 100; batch size: 1) for TP loads in the test dataset.

### 3.3.5 Evaluation of model performance

The best model performance of multi-criteria based ARIMA, GPR, MLP, and LSTM models of TN and TP loads are summarized in Table 5 and Table 6, respectively. The results show that LSTM persistently outperformed in the predictions of TN loads while MLP slightly outperformed LSTM in the forecasts of TP loads. In addition, although GPR has better test RMSE score, its R score is much lower than ARIMA. Moreover, ARIMA achieved best R score compared to other three models. Overall, MLP and LSTM persistently outperform GPR and ARIMA due to their abilities of capturing nonlinearities and complexities in the variations of nutrient loads. Furthermore, the optimal number of input observations in the predictive MLP and LSTM model of TP loads are consistent with that of autoregressive order in ARIMA. It is worth to highlight that the multi-criteria-based ARIMA exhibits competitive performance.

Table 5. Summarizes the best performances of the three models.

| Model Type | Test RMSE score | R-squared score |
|------------|-----------------|-----------------|
| ARIMA      | 34710.54        | <b>0.760</b>    |
| MLP        | 29454.62        | 0.661           |
| LSTM       | <b>27251.68</b> | 0.707           |
| GPR        | 33035.13        | 0.551           |

Table 6. Summarizes the best performances of the predictive models for TP loads.

| Model Type | Test RMSE score | R-squared score |
|------------|-----------------|-----------------|
| ARIMA      | 4390.63         | <b>0.587</b>    |
| MLP        | <b>3645.58</b>  | 0.492           |
| LSTM       | 3684.78         | 0.482           |
| GPR        | 4367.47         | 0.136           |

#### 4. Discussion and Conclusions

The decreasing pattern of variations in TN from 1980 to 2020 implies that excessive nutrient fluxes are manageable through collective efforts [3]. The nutrient loads forecasting is one of research focuses in environmental science community. Precise nutrient loads prediction, however, is a challenging because of its complex origins and nonlinear behavior. There are several factors that can impact predictions of temporal and spatial properties such as land usage, industrial activities, precipitation, agricultural, and the distribution location of excess nutrient. In this study, ARIMA, GPR, MLP, and LSTM model were developed to predict the nutrient loads at USGS 07373420. MLP and LSTM persistently achieved the better performance compared to the ARIMA and GPR models because the neural network type models are capable of capturing complex underlying processes and nonlinear nature in the variations of nutrient loads. ARIMA, however, demonstrate the comparable forecasting ability if it is selected based on multi-criteria which indicated that AIC value alone is not adequate for selection of best ARIMA model while residuals of the fitted model and significances of coefficients of model are the most influential factors which affect the accuracy of predictions from ARIMA. In order to select ARIMA with best performance, multiple selecting criteria including results of the autocorrelation function, the partial autocorrelation function, significances of coefficients, Akaike information criterion values, residuals of the fitted model and test errors should be considered. The RMSE scores indicate that ARIMA (1,0,0) × (0,1,1)<sub>12</sub> and ARIMA (1,0,0) × (5,1,0)<sub>12</sub> are best models in predictions of TN loads and TP loads,

respectively. By experimenting with wide range combination of hyper parameters, the optimal architecture of the single-layer MLP for TN loads is identified as the model with neurons of 50 while for TP loads, number of neurons is 200. For single-layered LSTM, the model with best performance has number of neurons 100 for TN loads and 30 for TP loads. In addition, the optimal numbers of input observations needed in the predictive MLP and LSTM models of TP loads are consistent with those of autoregressive order in ARIMA. Moreover, although GPR achieved slightly better RMSE score than ARIMA which demonstrated the power of Bayesian inference in prediction, R-squared scores are much lower than ARIMA. Furthermore, it is surprisingly found that Rmsprop optimizer performed well in the LSTM models to forecast TP loads instead of robust Adam optimizer. This study demonstrates the combination of several criteria is an effective tool to select the ARIMA with the best performance and explores the ability of machine learning models to accurately reproduce nutrient loads delivered to Gulf of Mexico. Future plan is to improve the accuracy of predictions by hybridizing models with superior performances and to investigate the characteristics of the predictive models at upper, middle, and lower Mississippi/Atchafalaya River Basin.

**Acknowledgments:** We gratefully acknowledge the staff members in the Office of Education Programs and the Computational Science Initiative at Brookhaven National Laboratory for their support and assistance in this research. We gratefully acknowledge the three anonymous reviewers for their helpful comments and advice which have substantially improved the quality of this manuscript. This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Visiting Faculty Program (VFP).

**Conflicts of Interest:** We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

## References

1. Diaz, R. J.; Rosenberg, R. (2008). Spreading Dead Zones and Consequences for Marine Ecosystems. *Science*, 2008, 321,5891, 926. <https://doi.org/10.1126/science.1156401>
2. USEPA (U.S. Environmental Protection Agency). Nutrient Criteria Technical Guidance Manual - Lakes and Reservoirs. Washington, D.C. : U.S. Environmental Protection Agency, Office of Water, 2000, 232pp.
3. Robertson, D.M.; Schwarz, G.E.; Saad, D.A.; Alexander, R.A. Incorporating Uncertainty into the Ranking of SPARROW Model Nutrient Yields from Mississippi/Atchafalaya River Basin Watersheds. *Journal of the American Water Resources Association* 2009, 45, 534. <https://doi.org/10.1111/j.1752-1688.2009.00310.x>
4. Robertson, D.M.; Saad, D.A. Nitrogen and Phosphorus Sources and Delivery from the Mississippi/Atchafalaya River Basin: An Update Using 2012 SPARROW Models. *Journal of the American Water Resources Association* 2021, 57, 406. <https://doi.org/10.1111/1752-1688.12905>.
5. Nie, J.; Mirza, S.; Viteritto, M.; Li, Y.; Witherell, B. B.; Deng, Y.; Yoo, S.; Feng, H. Estimation of nutrient (N and P) fluxes into Newark Bay, USA. *Marine Pollution Bulletin* 2023, 190, 114832. <https://doi.org/10.1016/j.marpolbul.2023.114832>

6. He, S.; Chu, T.-J.; Lu, Z.; Li, D. Coupling Imports of Dissolved Inorganic Nitrogen and Particulate Organic Matter by Aquaculture Sewage to Zhangjiang Estuary, Southeastern China. *Water* **2024**, *16*, 2054. <https://doi.org/10.3390/w16142054>
7. Morales-Marín, L.A.; Chun, K.P.; Wheeler, H.S.; Lindenschmidt, K.E. Trend analysis of nutrient loadings in a large prairie catchment. *Hydrological Sciences Journal* **2017**, *62*, 657. <https://doi.org/10.1080/02626667.2016.1255747>
8. Feng, H.; Qian, Y.; Cochran, J. K.; Zhu, Q.; Hu, W.; Yan, H.; Li, L.; Huang, X.; Chu, Y. S.; Liu, H.; Yoo, S.; Liu, C. J. Nanoscale measurement of trace element distributions in *Spartina alterniflora* root tissue during dormancy. *Scientific Reports*, **2017**, *7*, 40420. <https://doi.org/10.1038/srep40420>
9. Antonopoulos V. Z.; Papamichail, D. M.; Mitsiou, K. A. Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece. *Hydrol. Earth Syst. Sci.* **2001**, *5*, 679. <https://doi.org/10.5194/hess-5-679-2001>
10. Alexander, R.B.; Smith, R.A. Trends in the nutrient enrichment of U.S. rivers during the late 20th century and their relation to changes in probable stream trophic conditions. *Limnol. Oceanogr.* **2006**, *51*, 639. [https://doi.org/10.4319/lo.2006.51.1\\_part\\_2.0639](https://doi.org/10.4319/lo.2006.51.1_part_2.0639)
11. Fernández del Castillo, A.; Yebra-Montes, C.; Verduzco Garibay, M.; de Anda, J.; Garcia-Gonzalez, A.; Gradilla-Hernández, M.S. Simple Prediction of an Ecosystem-Specific Water Quality Index and the Water Quality Classification of a Highly Polluted River through Supervised Machine Learning. *Water* **2022**, *14*, 1235. <https://doi.org/10.3390/w14081235>
12. Du, J.L.; Feng H.; Nie, J.; Li, Y.; Witherell, B.B. Characterisation and assessment of spatiotemporal variations in nutrient concentrations and fluxes in an urban watershed: Passaic River Basin, New Jersey, USA". *Int.J.Environment and Pollution* **2018**, *63*, 154. <https://doi.org/10.1504/IJEP.2018.097310>
13. Lloyd, C.E.M.; Freer, J.E.; Johnes, P. J.; Collins, A.L. Using hysteresis analysis of high-resolution water quality monitoring data, including uncertainty, to infer controls on nutrient and sediment transfer in catchments. *Science of The Total Environment* **2016**, *543*, 388. <https://doi.org/10.1016/j.scitotenv.2015.11.028>
14. Goolsby, D. A.; Battaglin, W.A.; Lawrence, G. B.; Artz, R. S.; Aulenbach, B. T.; Hooper, R. P.; Keeney, D.R.; Stensland, G. J. Flux and Sources of Nutrients in the Mississippi-Atchafalaya River Basin: Topic 3 Report for the Integrated Assessment on Hypoxia in the Gulf of Mexico. *NOAA Coastal Ocean Program Decision Analysis Series No. 17. NOAA Coastal Ocean Program, Silver Spring, MD.* **1999**, 130 pp.
15. David, M.B.; Drinkwater, L.E.; McIssac, G.F. Sources of Nitrate Yields in the Mississippi River Basin. *Journal of Environmental Quality* **2010**, *39*, 1657. <https://doi.org/10.2134/jeq2010.0115>
16. Jacobson, L.M.; David, M.B.; Drinkwater, L.E. A Spatial Analysis of Phosphorus in the Mississippi River Basin. *Journal of Environmental Quality* **2011**, *40*, 931. <https://doi.org/10.2134/jeq2010.0386>
17. Feng, H.; Qian, Y.; Cochran, J. K.; Zhu, Q.; Heilbrun, C.; Li, L.; Hu, W.; Yan, H.; Huang, X.; Ge, M.; Nazareski, E.; Chu, Y. S.; Yoo, S.; Zhang, X.; Liu, C. J. Seasonal differences in trace element concentrations and distribution in

- Spartina alterniflora root tissue. *Chemosphere* **2018**, 204, 359. <https://doi.org/10.1016/j.chemosphere.2018.04.058>
18. Schreiber, S. G.; Schreiber, S.; Tanna, R. N.; Roberts, D. R. Arciszewski, T. J. Statistical tools for water quality assessment and monitoring in river ecosystems – a scoping review and recommendations for data analysis. *Water Quality Research Journal* **2022**, 57, 40. <https://doi.org/10.2166/wqrj.2022.028>
  19. de Andrade Costa, D.; Soares de Azevedo, J.P.; dos Santos, M.A.; dos Santos R. Water quality assessment based on multivariate statistics and water quality index of a strategic river in the Brazilian Atlantic Forest. *Scientific Reports* **2020**, 10, 22038. <https://doi.org/10.1038/s41598-020-78563-0>
  20. Yang, W.; Zhao Y.; Wang D.; Wu, H.; Lin, A; He L. Using Principal Components Analysis and IDW Interpolation to Determine Spatial and Temporal Changes of Surface Water Quality of Xin'anjiang River in Huangshan, China. *Int J Environ Res Public Health* **2020**, 17, 2942. <https://doi.org/10.3390/ijerph17082942>. PMID: 32344554; PMCID: PMC7215294.
  21. Singh, K.P.; Malik, A.; Mohan, D.; Sinha, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Res.* **2004**, 38, 3980. <https://doi.org/10.1016/j.watres.2004.06.011>
  22. Dutta, S.; Dwivedi, A.; Suresh Kumar, M. Use of water quality index and multivariate statistical techniques for the assessment of spatial variations in water quality of a small river. *Environ. Monit. Assess.* **2018**, 190, 718. <https://doi.org/10.1007/s10661-018-7100-x>
  23. Zhen, Y.; Feng, H. ; Yoo, S. Structuring Nutrient Yields throughout Mississippi/Atchafalaya River Basin Using Machine Learning Approaches, *Environments* **2023**, 10, 162. <https://doi.org/10.3390/environments10090162>
  24. Neitsch, S. L.; Arnold, J. G.; Kiniry, J. R.; Williams, J. R. Soil and Water Assessment Tool Theoretical Documentation Version 2009. **2011**, Texas Water Resources Institute.
  25. Worku, T.; Khare, D.; Tripathi, S. Modeling runoff–sediment response to land use/land cover changes using integrated GIS and SWAT model in the Beressa watershed. *Environmental Earth Sciences* **2017** 76, 550.
  26. Robertson, D. M.; Saad, D.A. SPARROW Models Used to Understand Nutrient Sources in the Mississippi/Atchafalaya River Basin. *Journal of Environmental Quality* **2013**, 42, 1422. <https://doi.org/10.2134/jeq2013.02.0066>
  27. Robertson, D.M.; Saad, D.A.; Schwarz, G.E. Spatial Variability in Nutrient Transport by HUC8, State, and Subbasin based on Mississippi/Atchafalaya River Basin SPARROW models. *Journal of the American Water Resources Association* **2014**, 50, 988. <https://doi.org/10.1111/jawr.12153>
  28. Adebisi, A.A.; Adewumi, A.O.; Ayo, C. K. Stock Price Prediction Using the ARIMA Model. *UKSim-AMSS 16th International Conference on Computer Modeling and Simulation*. **2014**.
  29. Alonso, A. M.; Garcia-Martos, C. Time Series Analysis - Forecasting with ARIMA models. *Universidad Carlos III de Madrid, Universidad Politecnica de Madrid*. **2012**.

30. Brownlee, J. Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras. <https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/>. **2016**.
31. Box, G.; Jenkins, G. Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day, **1970**.
32. Earnest, A.; Chen, M.I.; Ng, D.; Sin, L.Y. Using Autoregressive Integrated Moving Average (ARIMA) Models to Predict and Monitor the Number of Beds Occupied During a SARS Outbreak in a Tertiary Hospital in Singapore. *BMC Health Service Research*, **2005**, *5*, 36. <https://doi.org/10.1186/1472-6963-5-36>
33. Krauss, C.; Do, X.A.; Huck, N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* **2017**, *259*, 689. <https://doi.org/10.1016/j.ejor.2016.10.031>
34. Patterson, J. Deep Learning: A Practitioner's Approach, O'Reilly Media, **2017**.
35. Hum Nath Bhandari, Binod Rimal, Nawa Raj Pokhrel, Ramchandra Rimal, Keshab R. Dahal, Rajendra K.C. Khatri. Predicting stock market index using LSTM. *Machine Learning with Applications* **2022**, *9*, 100320. <https://doi.org/10.1016/j.mlwa.2022.100320>.
36. Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* **1998**, *6*, 107. <https://doi.org/10.1142/S0218488598000094>
37. Rasmussen, C.E.; Williams, C.K.I. (2006). Gaussian Processes for Machine Learning. MIT Press. ISBN 978-0-262-18253-9.
38. Haykin, S. Neural networks: a comprehensive foundation. Prentice Hall PTR, **1994**.
39. Elman, J.L. Finding structure in time. *Cognitive Science* **1990**, *14*, 179. ISSN 0364-0213. [https://doi.org/10.1016/0364-0213\(90\)90002-E](https://doi.org/10.1016/0364-0213(90)90002-E)
40. Patterson, J. Deep Learning: A Practitioner's Approach, O'Reilly Media, **2017**.
41. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Computation*, **1997**, *9*, 1735. <https://doi.org/10.1162/neco.1997.9>.