CHAPTER 7
RESULTS AND DISCUSSION ON MODEL PERFORMANCE

Investigated water quality parameter data of artificial lakes from Trial 1 and Trial 2 were used for training of predictive data-driven eutrophication models. The input parameters were initially finalized out of the different experimentally investigated water quality parameters for prediction of each of the desired outputs using a parameter trimming method under neural network topology. Based on optimum architecture of considered data-driven approach, models were trained for prediction of eutrophication indicators. To check the feasibility of the developed models to be used as eutrophication management tool for water bodies in Assam, the well trained models were then tested against some natural water body data. Sensitivity analysis was conducted on the final optimum models to check the impact of input variables on the target prediction. Different eutrophication models were developed from the dataset generated from experimental Trial 1 and Trial 2, results of which are presented separately in the succeeding sections.

7.1 MODEL PERFORMANCE FROM 1ST TRIAL

From the water quality parameters investigated during 1st trial, two major eutrophication indicators viz. DO and SD were chosen as target variables for the models and other parameters being probable input variables. For prediction of DO and SD, data-driven modelling approach in the form of ANN, SVR, and GPR were chosen. In the neural network architecture, two types namely MLP and TDNN were used. A simple multiple linear regression (MLR) model was also developed for each of the target variables initially to compare the performance of linear model with sophisticated machine learning regression methods.

7.1.1 Input Selection

For prediction of DO and SD six random input combinations were considered for determining the best parameter combination from investigated water quality indices

as shown in Table 7.1. Coefficient of correlation (R) and mean squared error (MSE) were considered as the assessment criteria for choice of optimum parameter combination. Every parameter combination represents importance of excluding parameter in the model training. Based on highest R and lowest MSE values it can be concluded that out of different scenarios considered under MLP, combinations 3 and 6 were the most significant ones for prediction of DO, SD, TN, and TP models respectively. So, for the DO prediction, parameters pH, EC, BOD, TN, TP, Turbidity, Temp were chosen as inputs. Parameters pH, EC, Turbidity and Temp were found as optimum inputs for SD prediction.

Table 7.1: Selection of input variables for the investigated models (Trial 1)

Sl. No.	Input Variables	R	MSE
	DO model		
1	pH, EC, TDS, TN, TP, BOD, Turbidity, SD, Temp	0.897	0.0121
2	pH, EC, TDS, TN, TP, Turbidity, Temp	0.901	0.0101
3	pH, EC, BOD, TN, TP, Turbidity, Temp	0.936	0.0094
4	pH, EC, TDS, TN, TP, Temp	0.788	0.0203
5	pH, EC, TN, TP, Temp	0.804	0.0204
6	pH, EC, TN, TP	0.755	0.0214
	SD model		
1	pH, EC, TDS, DO, TN, TP, Turbidity, Temp	0.883	0.0110
2	pH, EC, TDS, DO, TN, TP, Temp	0.771	0.0175
3	pH, EC, Turbidity, TN, TP, Temp	0.845	0.0120
4	pH, EC, TN, TP, Temp	0.783	0.0198
5	TN, TP, Turbidity, Temp	0.748	0.0118
6	pH, EC, Turbidity, Temp	0.921	0.0033

7.1.2 MLR Models

A multiple linear regression (MLR) model was trained initially for DO and SD prediction. The results of experimentally observed vs model predicted values are presented with Figure 7.1 and Figure 7.2 for DO and SD model respectively. It was observed that the MLR models' training efficiency was poor in general for prediction of SD. R^2, E, RMSE, and MAE values of 0.89, 0.89, 4.93 cm, and 3.94 cm were observed

for SD prediction through MLR model. In case of the DO model slightly better results were observed and R^2, E, RMSE, and MAE values of 0.94, 0.94, 0.22 mg/L, and 0.17 mg/L was achieved. As the relationships between ecological parameters are generally complex and are not linearly corelated, as found in the present study (Table 6), hence the MLR models training were not found satisfactory. The results were found consistent with previous works [4, 29, 30] where MLR models were not adequate enough for prediction of all eutrophication indicators that prompt use of other sophisticated non-linear machine learning tools.

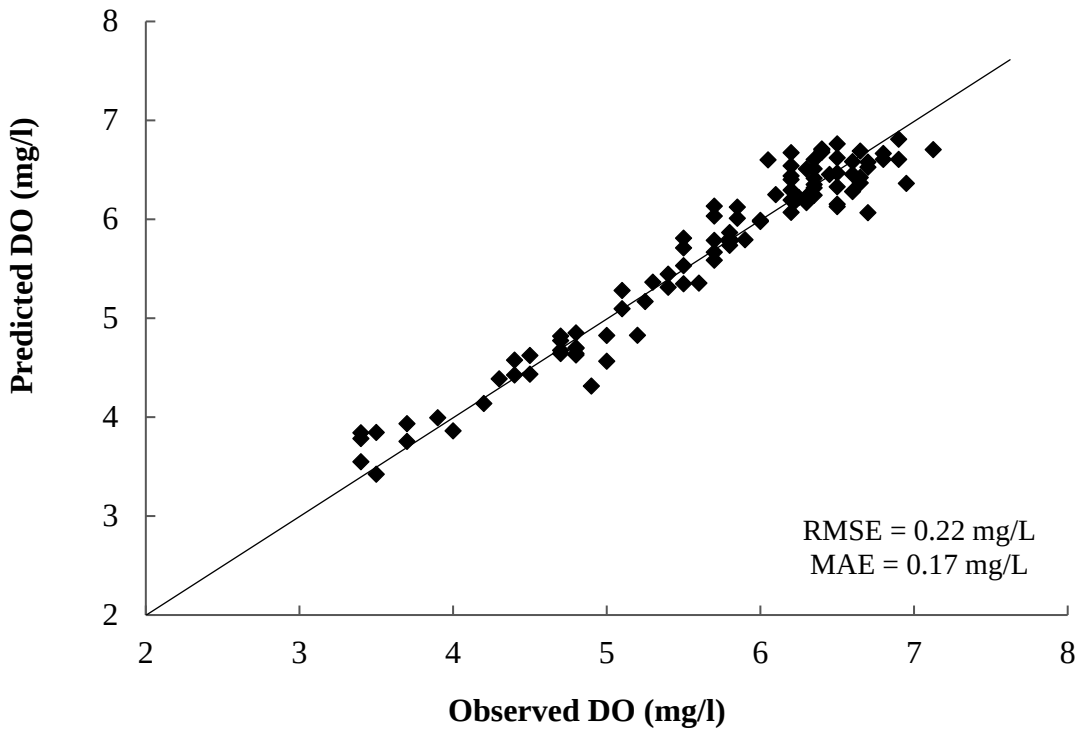

Figure 7.1: Result of MLR model training for prediction of DO (Trial 1)

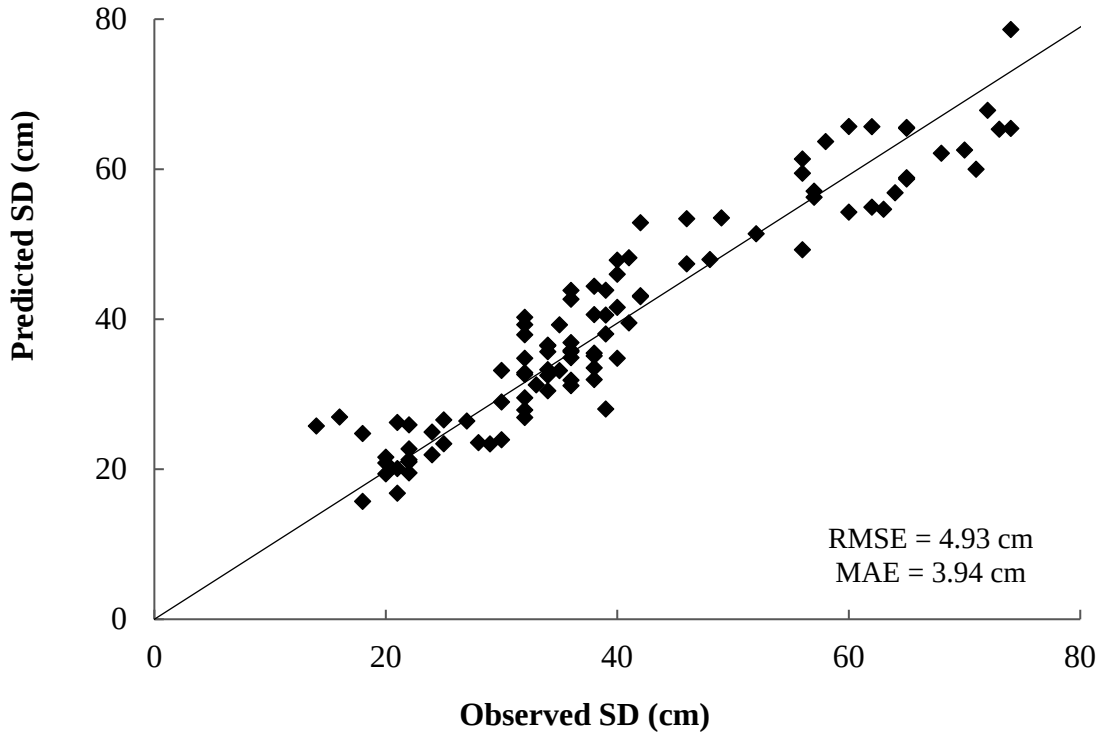

Figure 7.2: Result of MLR model training for prediction of SD (Trial 1)

7.1.3 ANN Models

Two types of neural network models, i.e., MLP and TDNN was employed in this present work for prediction of eutrophication indicators with the experimental data carried out on artificial lakes. To find optimum number of neurons in the hidden layer of the ANN models, five empirical methods from previous research were used and calculated values are presented in Table 7.2. A trial and error method was used thereafter in between minimum and maximum numbers of neurons calculated using the empirical methods to obtain optimum values for the developed ANN models. Neuron numbers in hidden layers were altered from 3 to 20 for DO model and 2 to 14 for SD model respectively to develop several neural networks. Each network was evaluated based on its R and MSE values and from the results presented in Table 7.3 it can be seen that 20 and 12 number of neurons in the hidden layer were optimum for DO and SD prediction respectively in ANN topology.

Table 7.2: Estimation of number of neurons in hidden layer of ANN models (Trial 1)

References	Method	DO Model (n=7)	SD Model (n=4)
[167]	$N = \frac{4n^2 + 3}{n^2 - 8}$	5	9
[99]	$N = \frac{\sqrt{1 + 8n} - 1}{2}$	3	3
[207]	$N = \frac{2^n}{n} + 1$	20	5
[168]	$N = \sqrt{N_i N_o}$	3	2
[80]	$N = \frac{N_i + \sqrt{N_p}}{L}$	17	14

Where, **N** = No of neurons in hidden layer; $\mathbf{N_i}$ & **n** = No of input neurons
$\mathbf{N_o}$ = No of output neurons; $\mathbf{N_p}$= No of input sample; **L** = No of hidden layers

Table 7.3: Fixing number of neurons in hidden layer of ANN models (Trial 1)

Model	No of Hidden Neuron	R	MSE
DO Model	3	0.902	0.0078
	5	0.915	0.0064
	7	0.894	0.0084
	10	0.904	0.0079
	12	0.889	0.0095
	15	0.917	0.0089
	17	0.919	0.0047
	20	**0.931**	**0.0038**
SD Model	2	0.841	0.0061
	3	0.890	0.0076
	5	0.907	0.0039
	8	0.896	0.0041
	10	0.894	0.0045
	12	**0.918**	**0.0031**
	14	0.824	0.0082

A neural network structure (inputs-hidden neuron-output) of 7-20-1, 4-12-1, were finalized as optimum in the MLP topology for prediction of DO and SD respectively. Thereafter keeping the same number of neurons in the hidden layer and with the same input parameters, TDNN models were developed with two steps ahead prediction of eutrophication indicators. The MLP and TDNN model structure in MATLAB is shown in Figure 7.3. Using Levenberg-Marquardt backpropagation training algorithm the MLP and TDNN models were trained between 0 to 1000 epochs until best validation results were obtained. Model training results revealed satisfactory performance of both the MLP and TDNN approach. It was observed that strong correlation was there between the observed and model predicted values and the results are presented from Figure 7.4 to Figure 7.7. Goodness of fit parameters R^2 and E were obtained greater than 0.97 for DO and SD prediction in MLP and TDNN training and error estimation parameters were also reasonable as presented in Table 7.4. For DO prediction RMSE of 0.16 mg/L and 0.13 mg/L were obtained with MLP and TDNN models respectively. RMSE value of 2.44 cm and 2.13 cm were found through MLP and TDNN models for SD prediction. Compared with previous research works by Huo et. al. [70], Akkoyunlu and Akiner [4] and Kuo et. al. [92] to predict DO and SD in eutrophic lakes using MLP models, present study produced a higher coefficient of determination (R^2) and lower RMSE values during model training. Out of the two adopted neural network approaches, it was seen that TDNN models were slightly superior to MLP and the results are consistent with the work carried by Aria et al. [11].

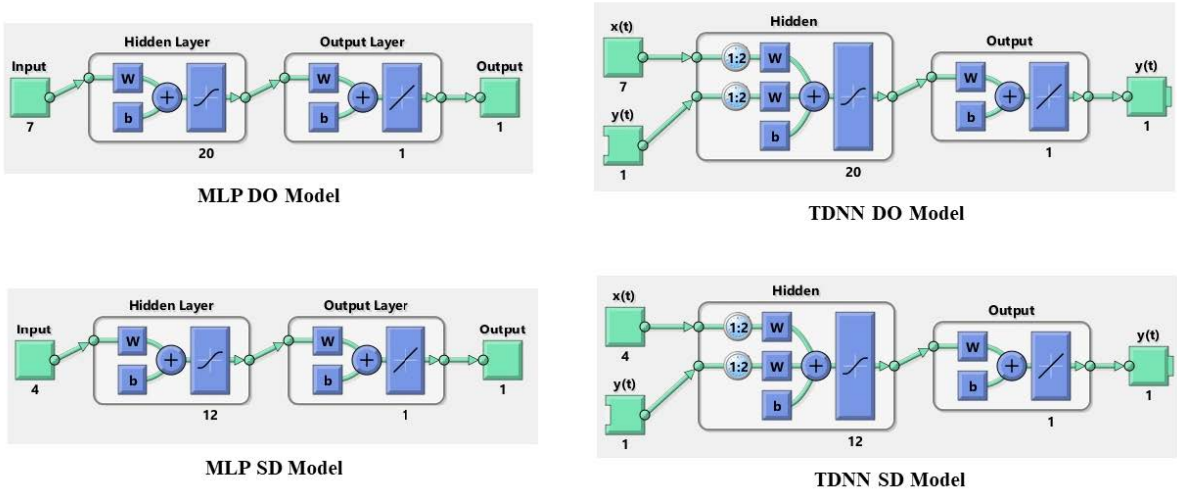

Figure 7.3: ANN architecture of proposed DO and SD model

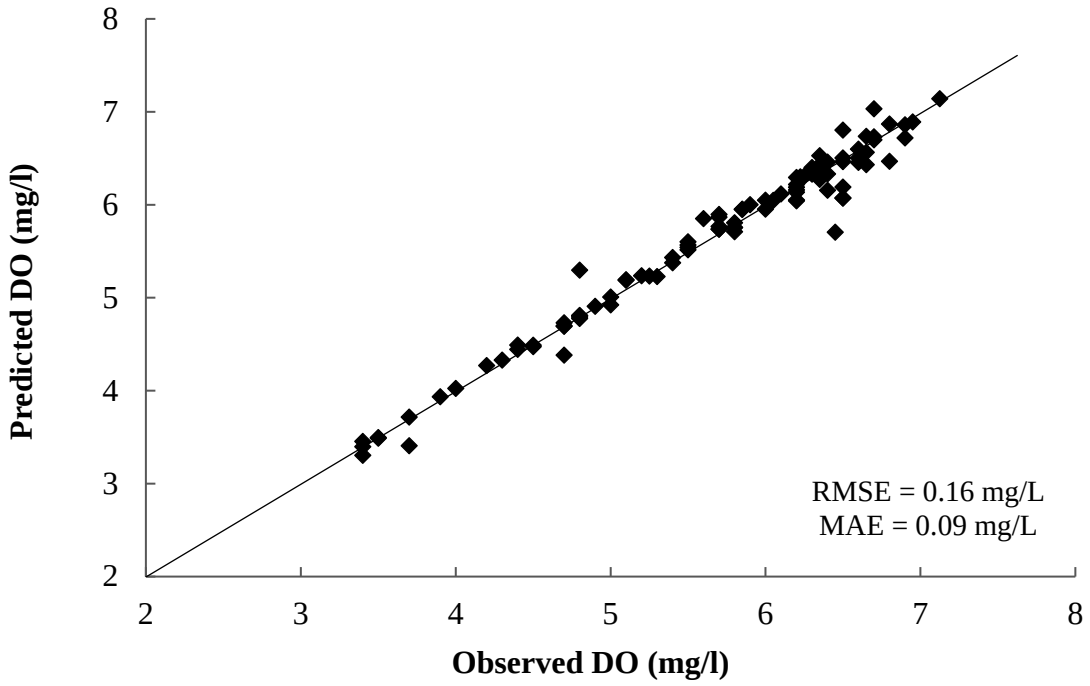

Figure 7.4: Observed vs predicted plot of DO model training under MLP ANN (Trial 1)

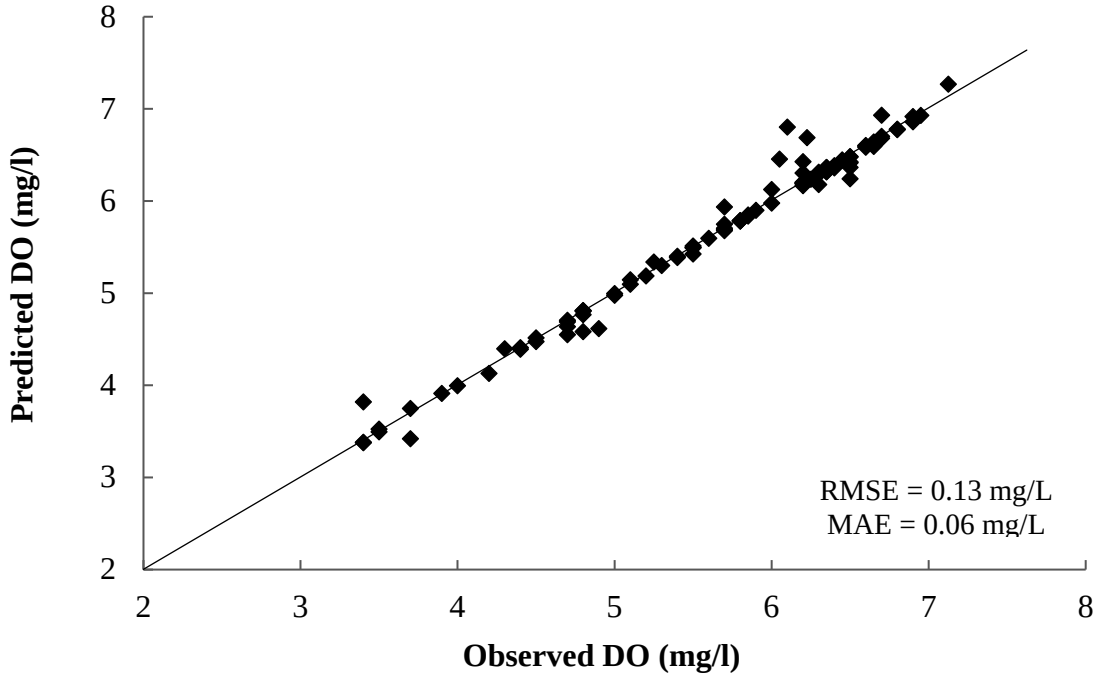

Figure 7.5: Observed vs predicted plot of DO model training under TDNN ANN (Trial 1)

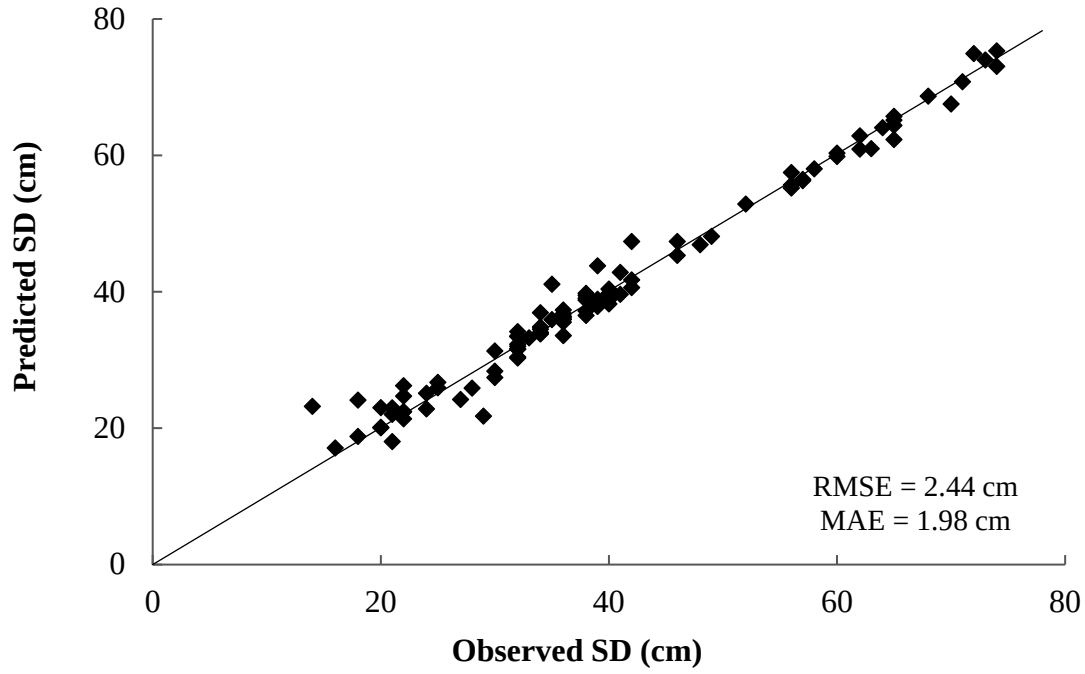

Figure 7.6: Observed vs predicted plot of SD model training under MLP ANN (Trial 1)

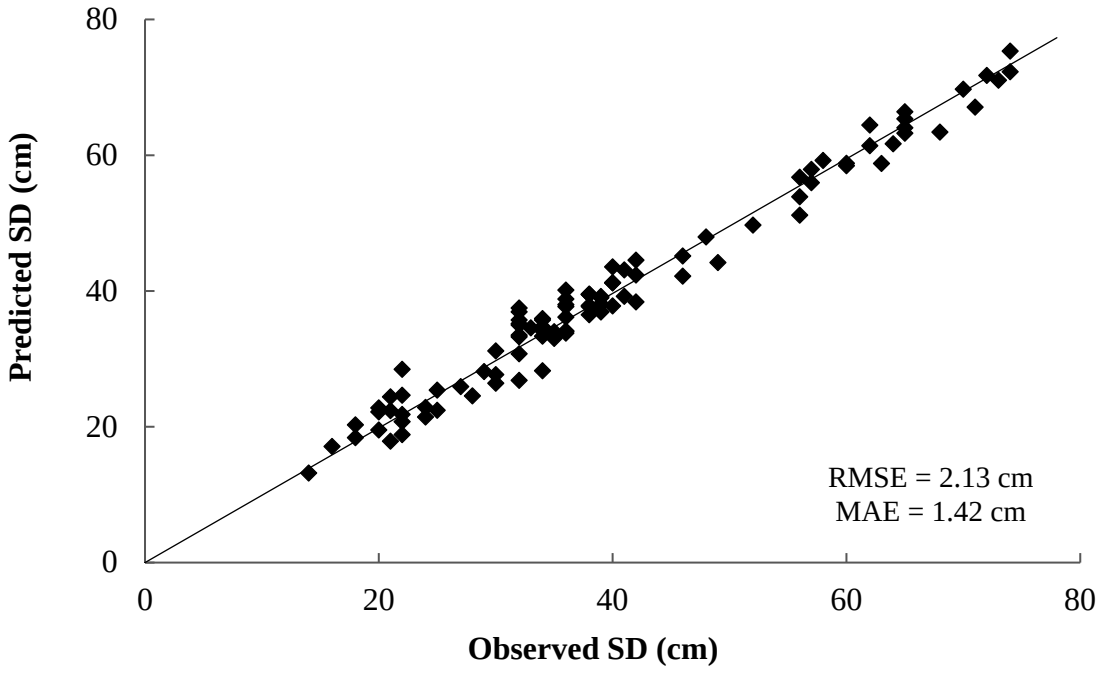

Figure 7.7: Observed vs predicted plot of SD model training under TDNN ANN (Trial 1)

7.1.4 SVR Models

A 5-fold cross-validation technique was employed throughout the SVR model training process and different kernel functions were used initially for each predictor variable. Using default values of C and ε under MATLAB environment linear, quadratic, cubic and Gaussian kernels were compared. Based on highest R^2 and lowest RMSE values, it was found that Gaussian kernel based SVR models yield better prediction accuracy compared to the other kernels as reported in previous works of Xu et al. [194]; García Nieto et al. [54]. Further different values of C (changed from 0 to 100 in increment of 10 for DO and SD model) and ε (from 0 to 0.1 by increment 0.01) were considered to find the best combination under Gaussian kernel. It was found that C and ε combinations of 10 and 0.01 for DO, 20 and 0.02 for SD respectively were the optimal. After finalizing the structure of the SVR, model training was done for each targeted output and the results are shown in Table 7.4. The performance of the models in terms of the observed and predicted values are presented with Figure 7.8 and Figure 7.9 for DO and SD model respectively. Good correlation between model predicted and observed values were found for DO model and R^2, E, RMSE, and MAE were reported as 0.95, 0.97, 0.17 mg/L and 0.13 mg/L. Compared to DO model, SD model training results under SVR were slightly inferior where R^2, E, RMSE, and MAE values were 0.93, 0.95, 3.16 cm, and 2.22 cm. Similar SVR models were used to predict the TP concentration in eutrophic lakes by García Nieto et al. [54] where R^2 value of 0.90 had been achieved. So, the presented SVR models for eutrophication indicators DO, SD, TN, and TP holds significant prediction accuracy. However, from Table 16 it can be observed that SVR model training efficiency is inferior compared to MLP, TDNN, and GPR models for DO and SD prediction.

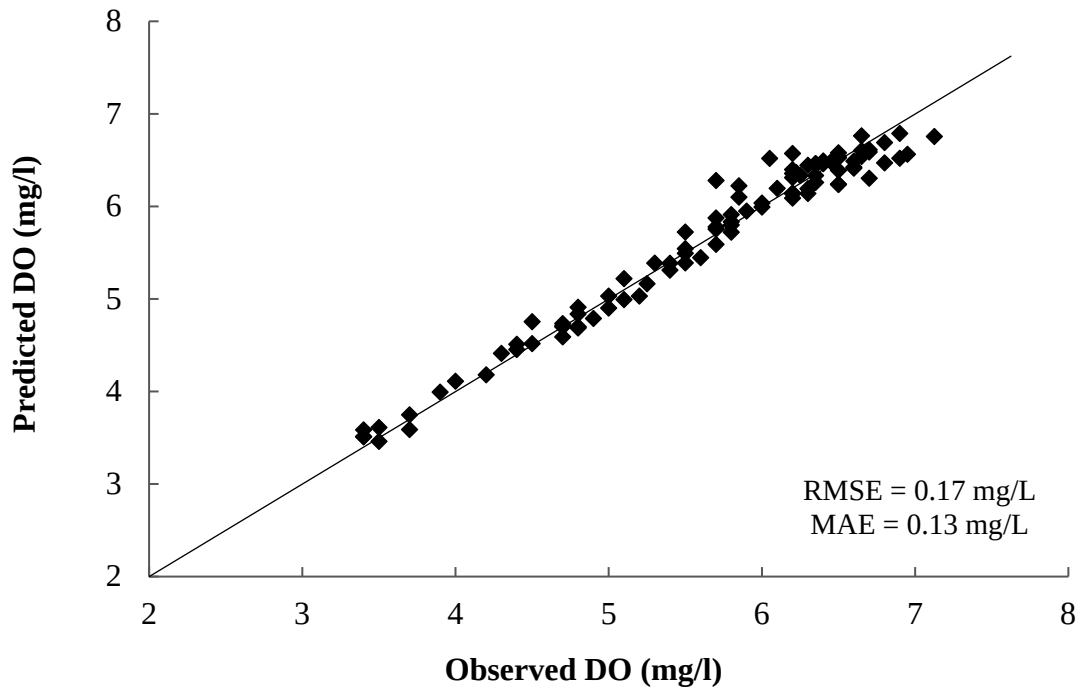

Figure 7.8: Observed vs predicted plot of DO model training under SVR (Trial 1)

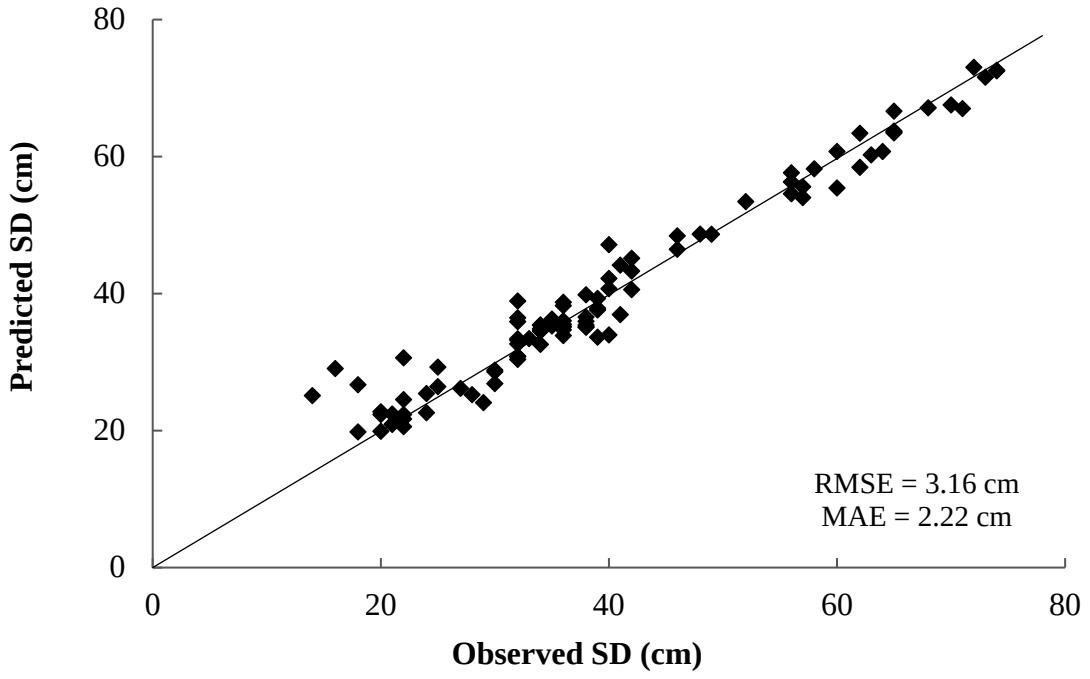

Figure 7.9: Observed vs predicted plot of SD model training under SVR (Trial 1)

7.1.5 GPR Models

Under GPR, squared exponential, rational quadratic, matern 5/2 and exponential kernels were tried to model DO and SD considering 5-fold cross-validation technique and a constant basis function in MATLAB. Considering R^2 and RMSE values squared exponential kernel was found to produce better training efficiency among the others which was also reported in earlier research works of García-Nieto et al. [52]. Table 7.4 presents the results of GPR model training with squared exponential kernel and the correlation between the actual and model predicted values are shown in Figure 7.10 and Figure 7.11 for DO and SD model respectively. The GPR models were found to be very superior in terms of prediction accuracy during training and R^2 and E values of 0.96 and 0.98 was observed for DO and SD respectively. The error parameters RMSE and MAE were also quite low during the GPR model training. As GPR models have the advantage of handling uncertainty as in the case of lake ecosystems and can perform well on small datasets [52], the performance of the presented models are quite promising. Moreover, as GPR is explored very less in literature to predict lake eutrophication indicators, the presented DO and SD models holds major significance.

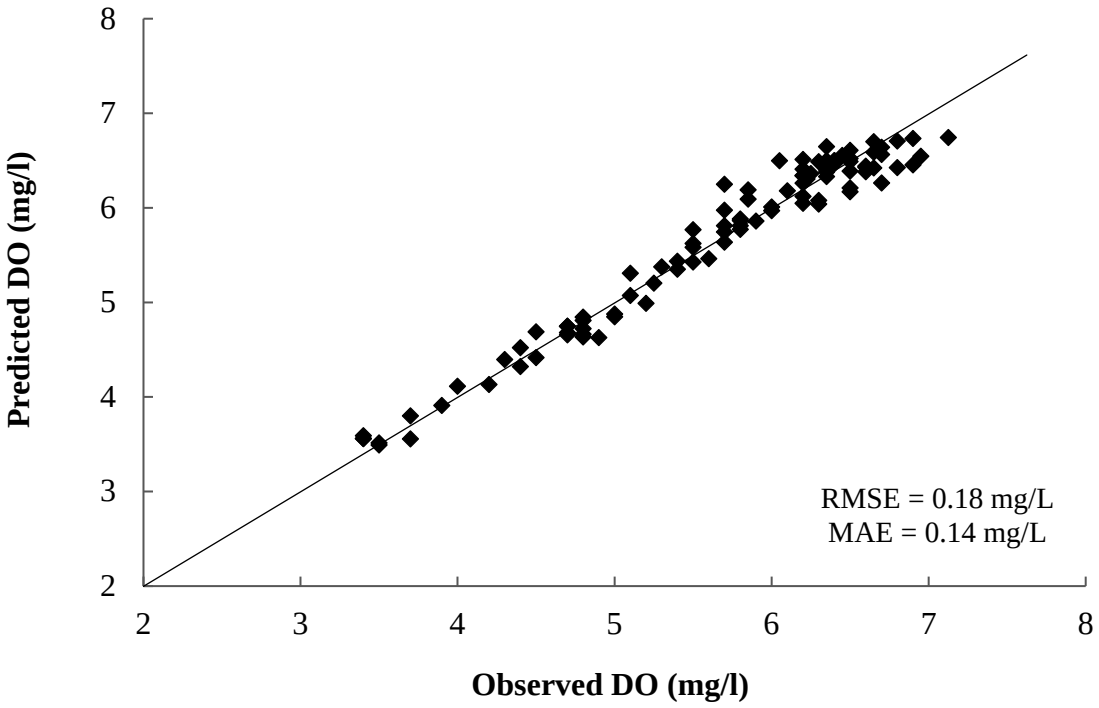

Figure 7.10: Observed vs predicted plot of DO model training under GPR (Trial 1)

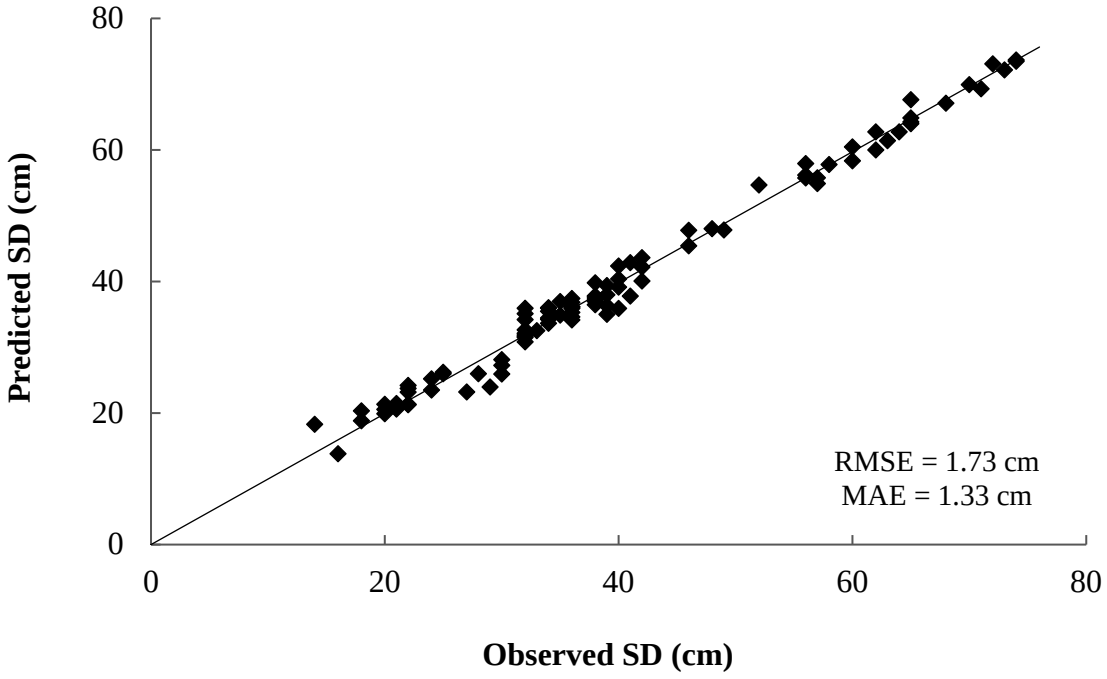

Figure 7.11: Observed vs predicted plot of SD model training under GPR (Trial 1)

Table 7.4: Performance results of trained models (Trial 1)

		MLP	TDNN	SVR	GPR
DO Model	**R^2**	0.97	0.98	0.95	0.96
	E	0.97	0.98	0.97	0.96
	RMSE	0.16 mg/L	0.13 mg/L	0.17 mg/L	0.18 mg/L
	MAE	0.09 mg/L	0.06 mg/L	0.13 mg/L	0.14 mg/L
SD Model	**R^2**	0.98	0.98	0.93	0.98
	E	0.98	0.98	0.95	0.98
	RMSE	2.44 cm	2.13 cm	3.16 cm	1.73 cm
	MAE	1.98 cm	1.42 cm	2.22 cm	1.33 cm

7.1.6 Model validation

To check the feasibility of the well-trained models to be used as a management tool for waterbodies in Assam, the MLR, MLP, TDNN, SVR, and GPR based optimum

models were used for forecasting the pre-observed values of DO and SD in some natural waterbodies as mentioned in Table 6.2. The sampling locations of water bodies i.e., Deepor Bil, the marsh, artificial lake, and the village pond were investigated for TSI and the results are presented in Table 7.5. It was seen that the water bodies were in eutrophic condition as the calculated average TSI values were found to be greater than 70.

Table 7.5: Average trophic status index of investigated natural waterbodies

	Deepor Bil	**Artificial Pond**	**Marsh**	**Village Pond**
TSI_{MEAN}	78.57	75.50	76.16	74.92

The natural water body data was initially checked against the linear regression model and the model validation performance have been presented with Figure 7.12 and Figure 7.13 for DO and SD model respectively. It can be observed from the figures that the performance of the MLR models were very poor to predict the desired eutrophication indicators in natural water bodies. In case of the DO model, R^2 and E values of 0.71 and 0.66 have been attained during model testing. The error estimation parameters were also high and RMSE and MAE were reported as 0.77 mg/L and 0.67 mg/L. The correlation between observed and model predicted values were also poor for SD model and R^2, E, RMSE and MAE values were reported as 0.65, 0.50, 8.33 cm, and 7.63 cm. So, from the training and testing results of MLR models, it can be concluded that the performance of these models were not sufficient for prediction of eutrophication indicators in natural water bodies in Assam.

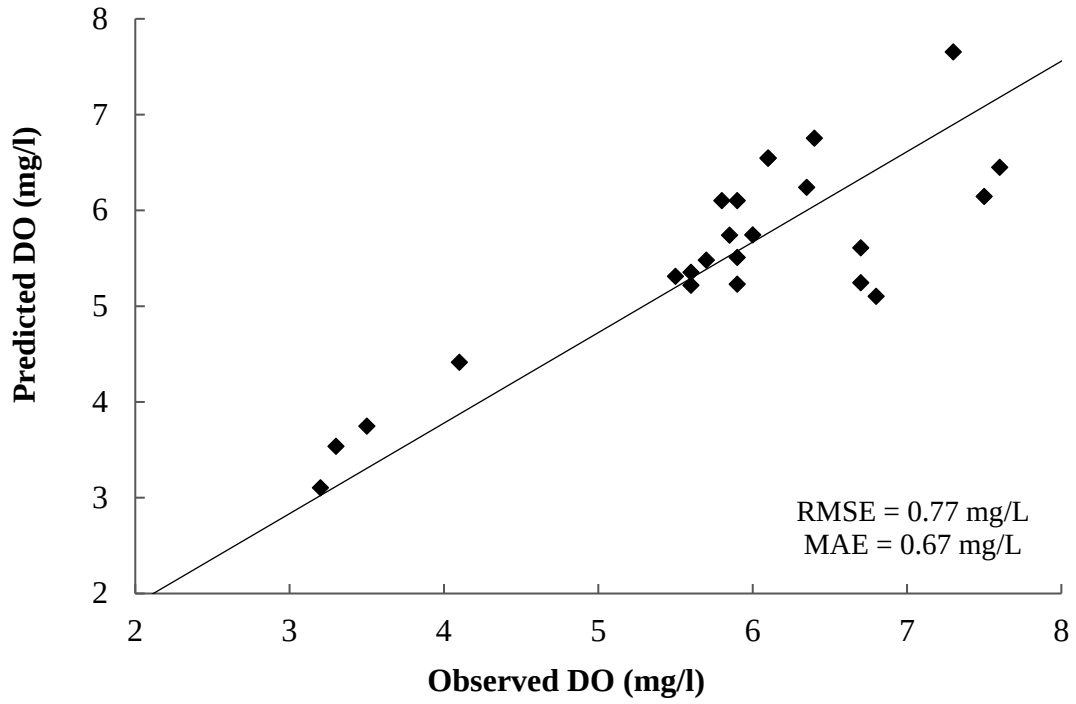

Figure 7.12: Performance of MLR for DO prediction against natural water body (Trial 1)

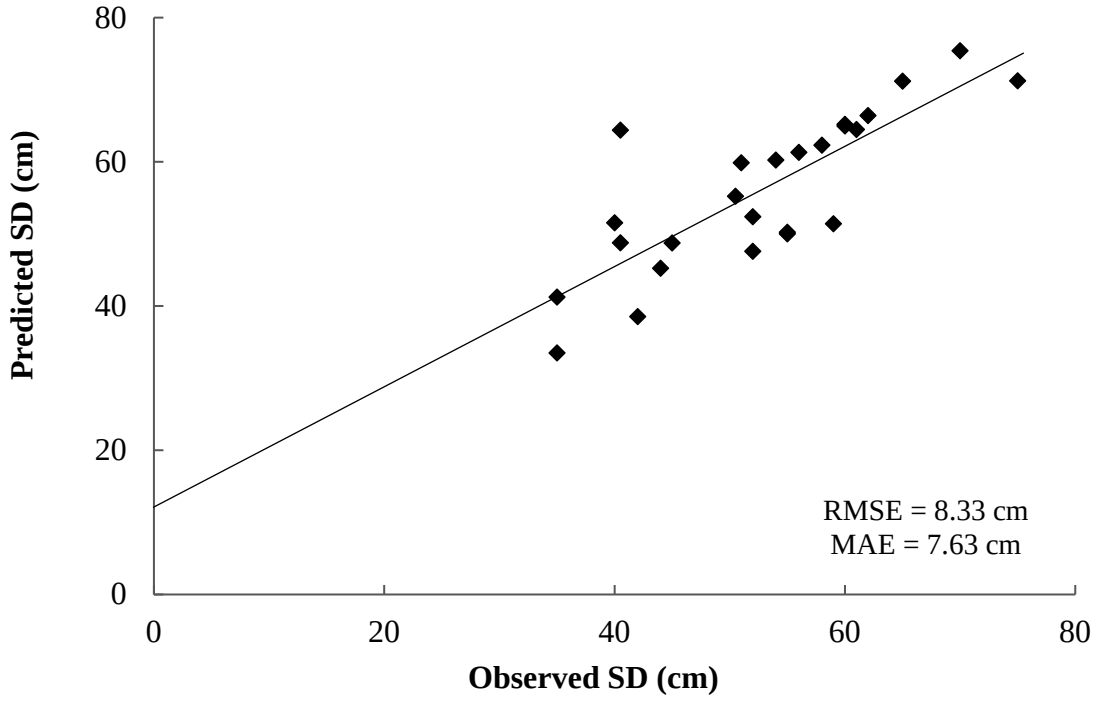

Figure 7.13: Performance of MLR for SD prediction against natural water body (Trial 1)

The results of correlation achieved between observed values of DO in natural waterbodies with its model predicted values are presented from Figure 7.14 to Figure 7.17 for MLP, TDNN, SVR, and GPR models respectively. Figure 7.18 to Figure 7.21 shows the observed vs predicted plot of SD model testing under MLP, TDNN, SVR, and GPR architecture respectively. The statistical evaluation of model testing results are presented in Table 7.6 below. It can be observed that for prediction of DO and SD the trained MLP, TDNN and GPR models showed better consistency between field data and predicted value as R^2 value greater than 0.90 was observed for all the models. The error parameters for these models were also found as relatively smaller indicating suitability of the modelling approach for eutrophication prediction. Compared to these models, the SVR model performance was poor during testing phase. The observed R^2 and E values were slightly lower and error parameters were higher for SVR models compared to its counterparts. Overall, the errors in target prediction from all the models were found slightly higher during testing phase compared to model training phase. This may be due to the fact that the models were trained based on data of artificially eutrophied lakes under controlled environment which is not an ideal condition in natural water bodies.

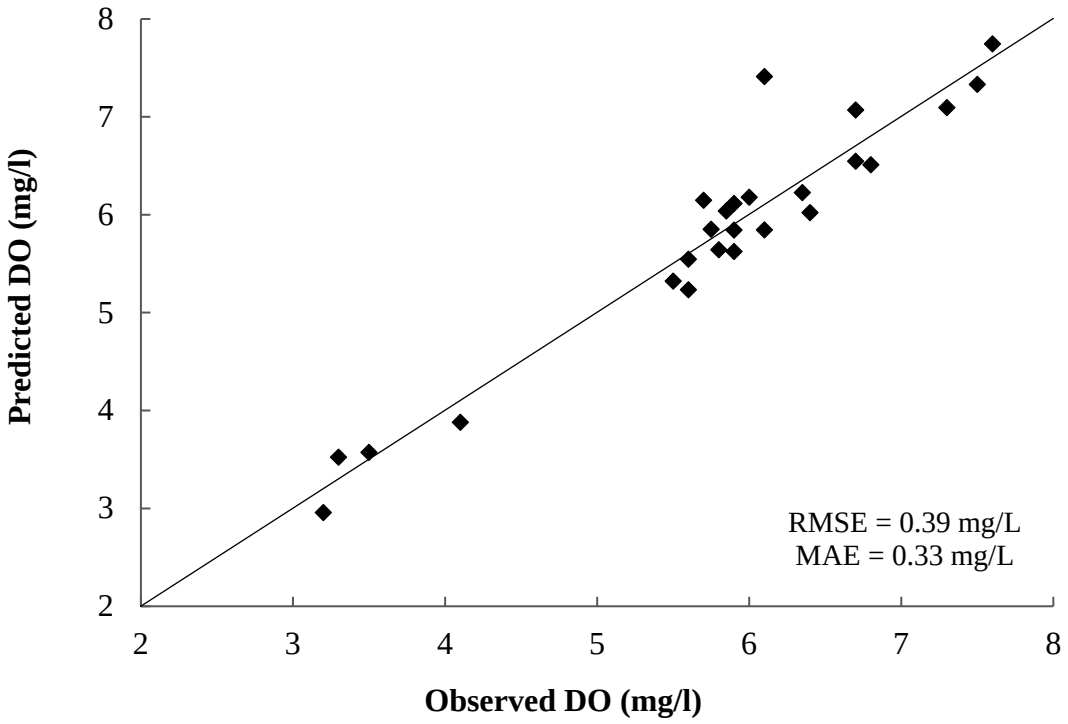

Figure 7.14: Performance of MLP for DO prediction against natural water body (Trial 1)

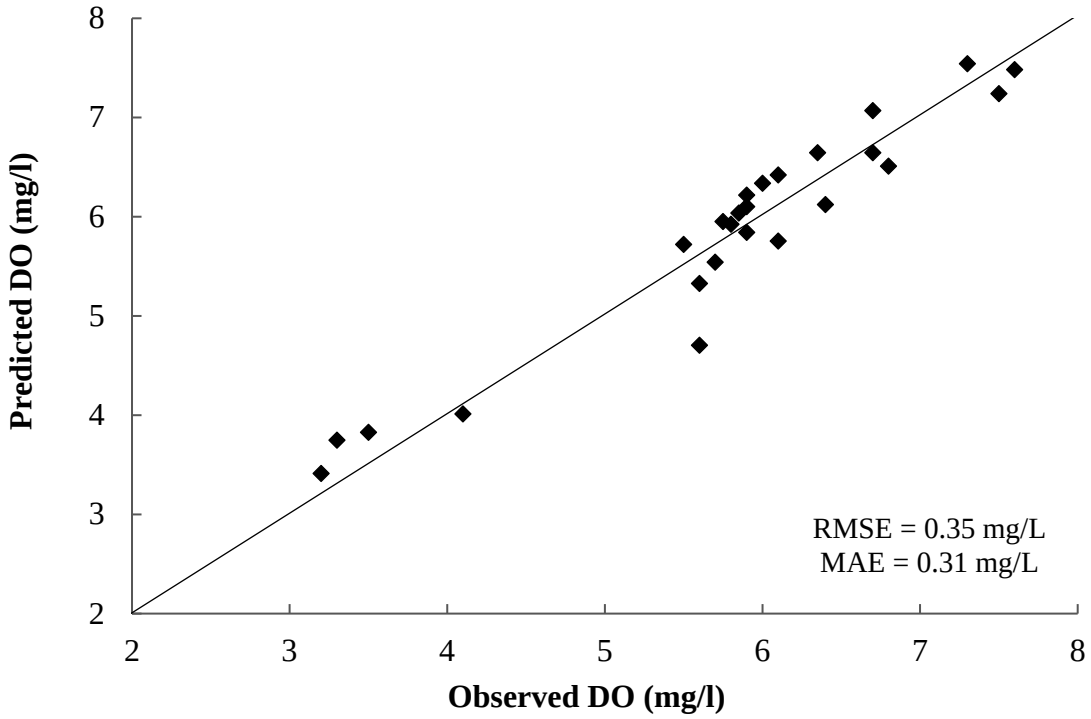

Figure 7.15: Performance of TDNN for DO prediction against natural water body (Trial 1)

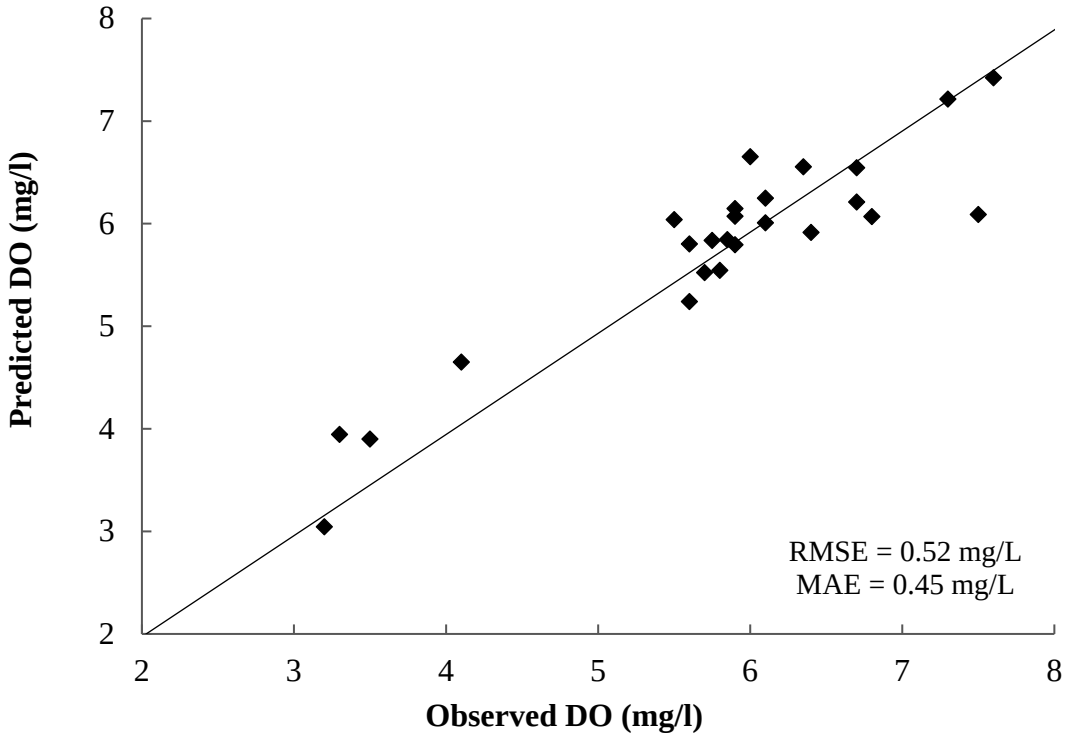

Figure 7.16: Performance of SVR for DO prediction against natural water body (Trial 1)

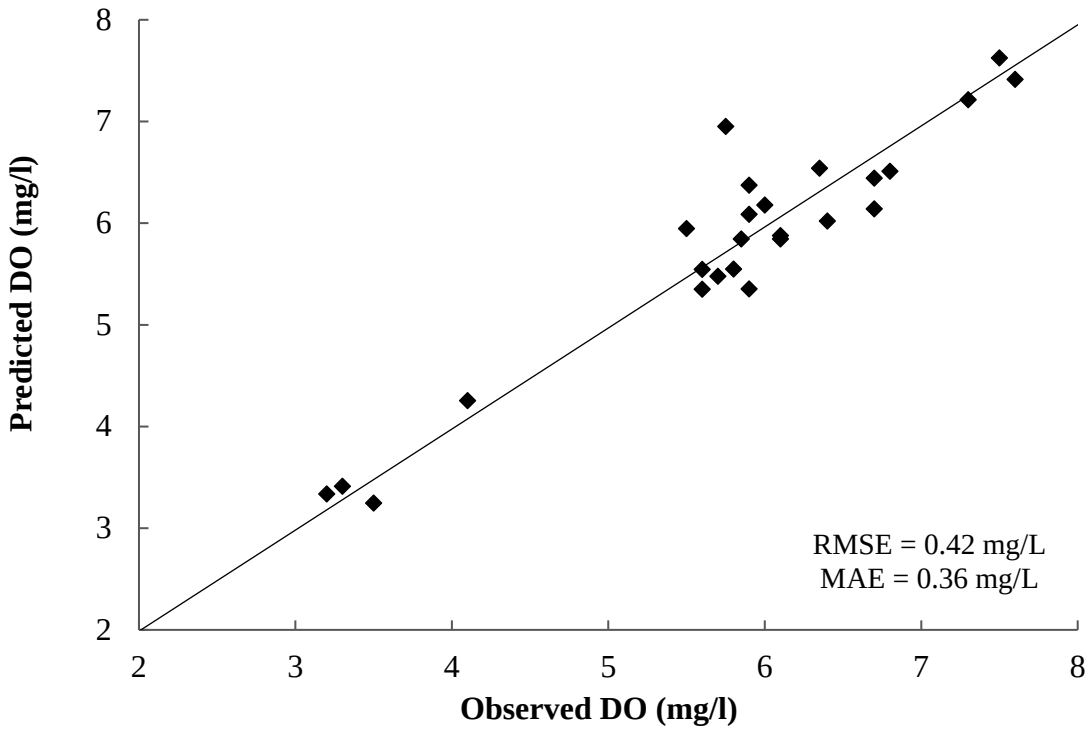

Figure 7.17: Performance of GPR for DO prediction against natural water body (Trial 1)

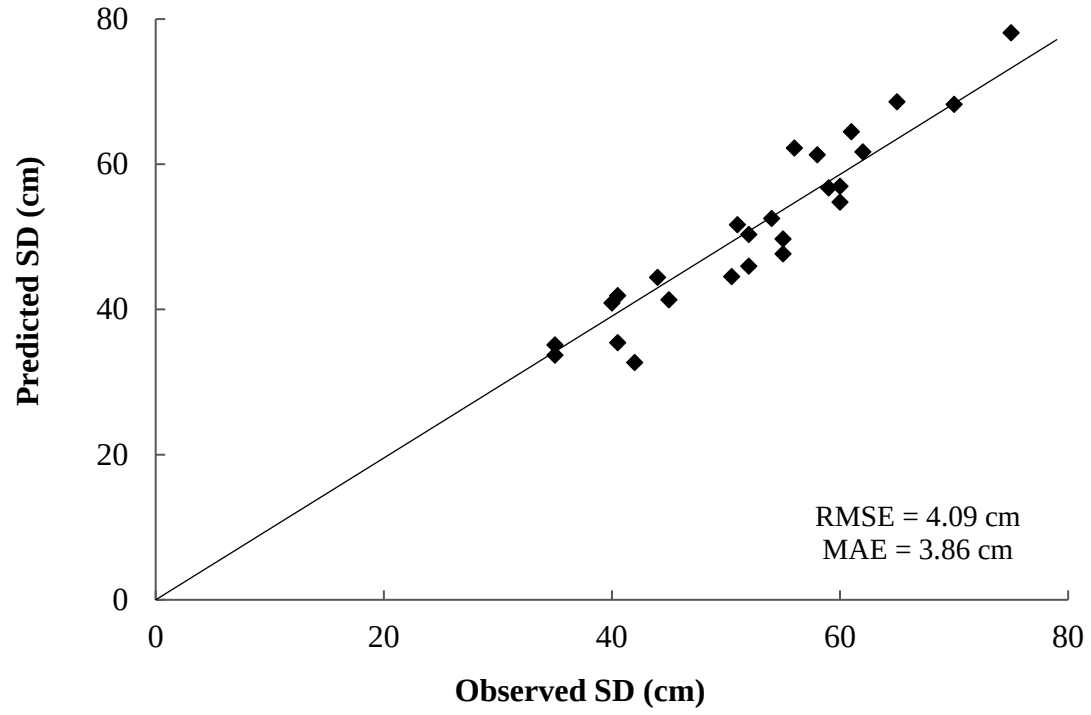

Figure 7.18: Performance of MLP for SD prediction against natural water body (Trial 1)

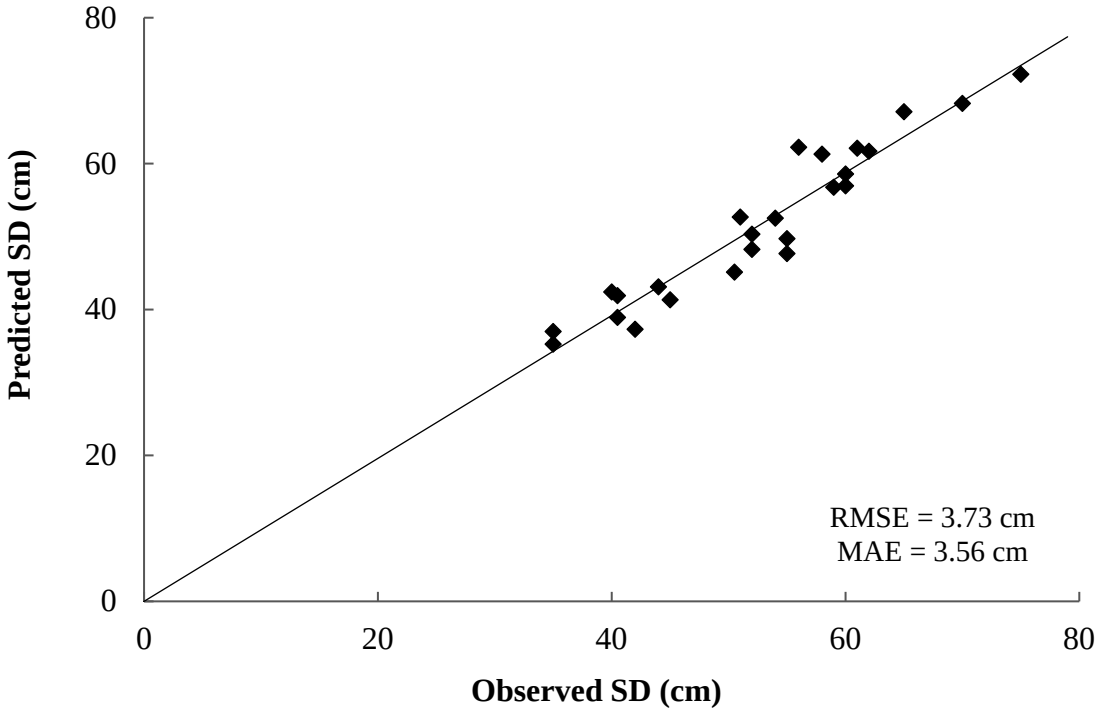

Figure 7.19: Performance of TDNN for SD prediction against natural water body (Trial 1)

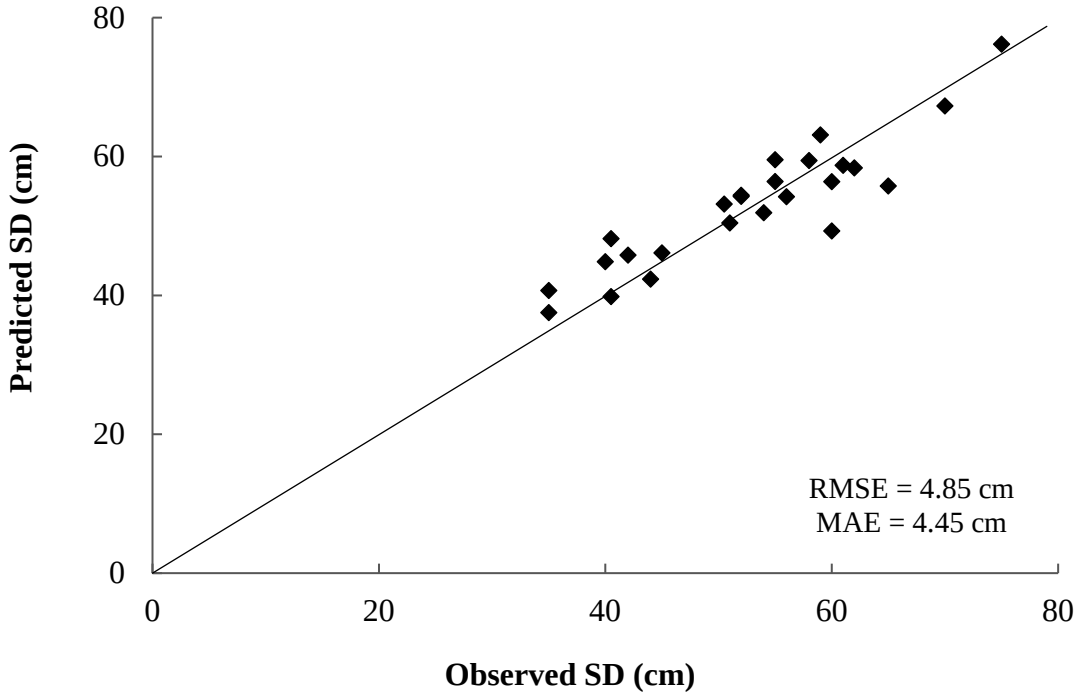

Figure 7.20: Performance of SVR for SD prediction against natural water body (Trial 1)

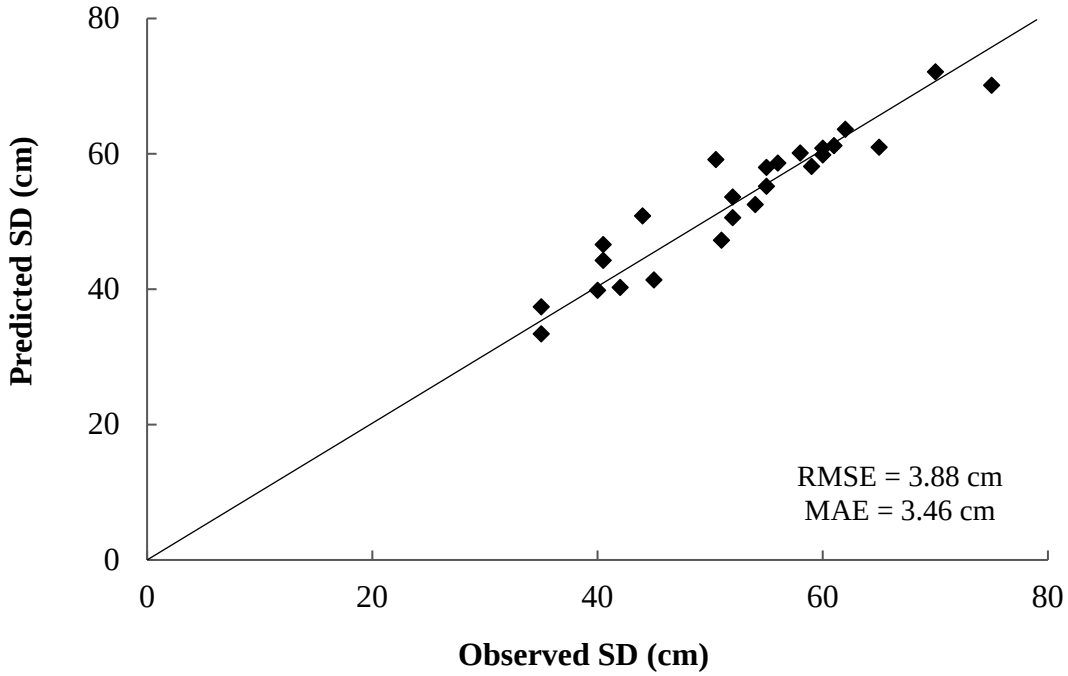

Figure 7.21: Performance of GPR for SD prediction against natural water body (Trial 1)

Table 7.6: Performance result of the model testing against natural waterbody (Trial 1)

		MLP	TDNN	SVR	GPR
DO Model	**R^2**	0.92	0.93	0.85	0.90
	E	0.91	0.93	0.85	0.89
	RMSE	0.39 mg/L	0.35 mg/L	0.52 mg/L	0.42 mg/L
	MAE	0.33 mg/L	0.31 mg/L	0.45 mg/L	0.36 mg/L
SD Model	**R^2**	0.90	0.91	0.84	0.90
	E	0.88	0.90	0.83	0.89
	RMSE	4.09 cm	3.73 cm	4.85 cm	3.88 cm
	MAE	3.86 cm	3.56 cm	4.45 cm	3.46 cm

7.1.7 Sensitivity Analysis

Data perturbation method was used to check impact of input parameters on the prediction of DO and SD. Input parameters were increased and decreased by 20% one at

a time keeping other parameters unchanged. The percentage change in output due to this data perturbation is reported as sensitivity of the changed parameter on desired output. During model training and validation stage MLP, TDNN and GPR based models were found to produce better prediction accuracy, so sensitivity analysis is conducted on these models and the results are presented with Table 7.7. Similar sensitivity trends of inputs were observed for DO and SD prediction respectively under MLP, TDNN, and GPR models. For the DO model, increase in BOD and TN values were observed as relatively sensitive input parameters compared to others. The higher sensitivity of TN can be correlated with experimental findings also where it was seen that TN was mostly utilized during algal growth. So, variations in TN values can bring significant changes in DO prediction. The sensitivity of BOD in DO prediction may be attributed to the fact that higher BOD values can lead to reduction of DO in the water body. Compared to the DO model, input variables pH, EC, temperature, and turbidity used for SD prediction were found more or less consistent and sensitivity values were less than 20% in MLP, TDNN, and GPR models. Out of the four inputs, increase in turbidity was found to be relatively sensitive for SD prediction. which may be due to the fact that these two parameters are closely related with water transparency.

Table 7.7: Result of sensitivity analysis with data perturbation (Trial 1)

Parameters	**DO Model**						**SD Model**					
	MLP		TDNN		GPR		MLP		TDNN		GPR	
	+20%	-20%	+20%	-20%	+20%	-20%	+20%	-20%	+20%	-20%	+20%	-20%
pH	16.17	13.46	15.42	12.41	17.52	15.34	14.17	17.84	15.34	18.34	14.87	16.34
EC	16.45	17.74	14.32	15.23	14.83	13.56	14.21	11.24	13.55	12.45	12.08	11.73
TN	32.83	16.46	29.65	14.87	27.64	17.65	-	-	-	-	-	-
TP	14.69	15.29	16.54	17.08	18.38	17.33	-	-	-	-	-	-
Temp	17.43	15.33	14.51	12.94	13.82	12.43	12.35	11.08	10.65	9.42	11.47	10.09
BOD	27.26	18.18	28.95	17.44	25.64	14.55	-	-	-	-	-	-
Turb	15.74	18.93	16.42	17.61	14.28	16.47	19.36	15.69	18.65	17.45	18.07	16.42

7.1.8 Major observations on model performance

From the results and discussion of the model performance the following major conclusions can be drawn:

- Using a model-based data-pruning method 7 and 4 input parameters were optimized for prediction of DO and SD under MLR, MLP, TDNN, SVR, and GPR methods.
- For both DO and SD models, R^2, E, RMSE and MAE values were found to be acceptable for MLP, TDNN, SVR, and GPR models during training.
- MLP, TDNN, SVR, and GPR models were able to forecast the eutrophication indices DO and SD in the considered Assam's natural water bodies with an acceptable level of accuracy.
- Linear Regression models showed poor prediction accuracy during both model training and validation stage.
- Performance of MLP, TDNN, and GPR models were found to be superior compared to SVR models for DO and SD prediction.
- Increase in the values of BOD and TN were reported as major sensitive parameters for DO prediction but the input parameters of SD model were consistent in general.

7.2 MODEL PERFORMANCE FROM 2ND TRIAL

From the water quality parameters investigated during 2nd trial, more exhaustive data-driven eutrophication models were proposed. In addition to DO and SD, another important eutrophication indicator Chl-a was also chosen as model output parameter. The data of concrete Tank 1 and artificial pond Tank 3 were used separately for model training models of DO, SD, and Chl-a. As ANN and GPR based models were found to produce better prediction accuracy compared to SVR method during 1st trial, ANN and GPR methods were considered again in the 2nd trial. As MLP neural networks are the most commonly employed ANN architecture in ecological modelling, for brevity of work only this type is considered during 2nd trial. Moreover, sophisticated machine learning algorithm ANFIS was also employed for DO, SD, and Chl-a prediction. For each experimental prototype lake, for every indicator parameter three models were

trained using ANN, GPR, and ANFIS methods. As illustrated with Figure 7.8, a total of eighteen models were trained and tested during 2nd trial. The models trained with Tank 1 data are denoted as DO-I, SD-I and Chl-a-I respectively and that trained with Tank 3 data are presented as DO-II, SD-II and Chl-a-II respectively in the succeeding sections. The input selection procedure and modelling methodology for ANN and GPR were similar to 1st trial.

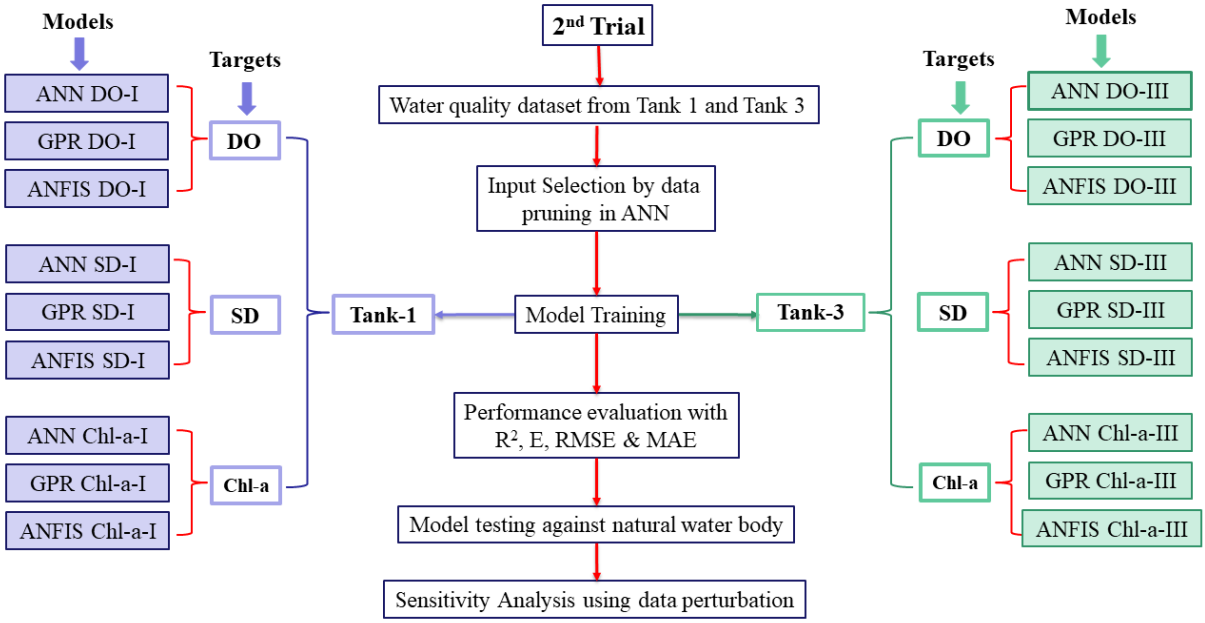

Figure 7.22: Details of models developed from 2nd Trial

7.2.1 Input Selection

For selection of optimum input parameters for DO, SD and Chl-a forecasting in eutrophic lakes, a model based data pruning method has been employed similar to 1st trial. For brevity of work, ten input scenarios in ad-hoc basis were compared in ANN methodology for choosing optimum input parameters. Each combination represents significance of each omitted parameter in the model training performance. Results of each combination was evaluated based on coefficient of correlation (R) and mean squared error (MSE) and are presented in Table 7.8. It has been observed from Table 7.8 that combination 8 for DO model, combination 5 for SD and combination 9 for Chl-a

Table 7.8: Combinations for input parameter selection (Trial 2)

Sl. No.	Input Parameters	R	MSE
	DO model		
1	pH, EC, TDS, TN, TP, BOD, Turbidity, Chl-a, Temp	0.9266	0.0106
2	pH, EC, TDS, TN, TP, Turbidity, Temp^{r}, Chl-a	0.9227	0.0079
3	pH, EC, BOD, TN, TP, Turbidity, Chl-a, Temp	0.9198	0.0062
4	pH, EC, BOD, TN, TP, Chl-a**,** Temp	0.9073	0.0104
5	pH, EC, TN, TP, Chl-a, Temp	0.9351	0.0036
6	pH, EC, TN, TP, Chl-a	0.9340	0.0166
7	pH, EC, TN, TP, Temp	0.8899	0.0061
8	**pH, EC, TP, Chl-a, Temp**	**0.9600**	**0.0029**
9	pH, TDS, TP, Chl-a, Temp	0.9132	0.0085
10	pH, TDS, Chl-a, Temp	0.8998	0.0100
	SD model		
1	pH, EC, TDS, TN, TP, BOD, Turbidity, Chl-a, Temp	0.891	0.0036
2	pH, EC, TDS, TN, TP, Turbidity, Chl-a, Temp	0.9053	0.0050
3	pH, EC, Turbidity, TN, TP, Chl-a, Temp	0.8829	0.0236
4	pH, EC, TN, TP, Chl-a, Temp	0.7554	0.0337
5	**pH, EC, Turbidity, Chl-a, Temp**	**0.9135**	**0.0035**
6	TN, TP**,** Turbidity, Chl-a, Temp	0.8824	0.0079
7	pH, EC, Chl-a, Temp	0.8267	0.0279
8	pH, EC, Temp^{r}, Turbidity	0.8591	0.0085
9	pH, TDS, Turbidity, Chl-a, Temp	0.8555	0.0173
10	EC, TDS, Turbidity, Chl-a, Temp	0.7801	0.0036
	Chl-a model		
1	pH, EC, TDS, TN, TP, BOD, Turbidity, DO, Temp	0.8827	0.0160
2	pH, EC, TN, TP, BOD, Turbidity, DO, Temp	0.8428	0.0093
3	pH, EC, TN, TP, Turbidity, DO, Temp	0.8668	0.0097
4	pH, TN, TP, Turbidity, DO, Temp	0.8469	0.0088
5	pH, TN, TP, Turbidity, DO, BOD, Temp	0.8978	0.0084
6	pH, TN, TP, DO, BOD, Temp	0.8374	0.0237
7	pH, TN, TP, BOD, Turbidity, DO, SD, Temp	0.8841	0.0108
8	pH, EC, TDS, TN, TP, BOD, Turbidity, DO, SD, Temp	0.9024	0.0067
9	**pH, EC, TN, TP, BOD, Turbidity, DO, SD, Temp**	**0.9161**	**0.0022**
10	EC, TN, TP, BOD, Turbidity, DO, SD, Temp	0.8389	0.0180

model were having highest correlation and minimum MSE values and hence were finalised as model input parameters for presented models during the 2nd trial. As such pH, EC, TP, Chl-a, and temperature were chosen as input parameters for DO prediction and pH, EC, Turbidity, Chl-a, and temperature were inputs for SD model. For Chl-a model optimized input parameters were pH, EC, TN, TP, BOD, Turbidity, DO, SD, and temperature.

7.2.2 ANN Models

A feed-forward neural networks with back propagation algorithm i.e., MLP was employed to predict DO, SD and Chl-a in eutrophic lakes. To find optimum number of neurons in the hidden layer of the ANN models, five empirical methods were used and calculated values are presented in Table 7.9. A trial and error method was used thereafter in between minimum and maximum numbers of neurons calculated using the empirical methods to obtain optimum values for the developed ANN models. Neuron numbers in hidden layers were altered from 2 to 15 for DO and SD models and 3 to 58 for Chl-a model to develop several neural networks. Each network was evaluated based on its R and MSE values and the results are presented in Table 7.10. It was found that 8, 7 and 18 number of neurons in the hidden layer produced highest R and lowest MSE values and so finalized as optimum for DO, SD and Chl-a prediction respectively in ANN topology.

Table 7.9: Estimation of number of neurons in hidden layer of ANN models (Trial 2)

References	Method	DO Model (n=5)	SD Model (n=5)	Chl-a Model (n=9)
[167]	$N = \frac{4n^2 + 3}{n^2 - 8}$	6	6	5
[99]	$N = \frac{\sqrt{1 + 8n} - 1}{2}$	3	3	4
[207]	$N = \frac{2^n}{n} + 1$	8	8	58
[168]	$N = \sqrt{N_i N_o}$	2	2	3
[80]	$N = \frac{N_i + \sqrt{N_p}}{L}$	15	15	19

Table 7.10: Estimation of number of neurons in hidden layer of ANN models (Trial 2)

Model	No of Hidden Neuron	R	MSE
DO Model	2	0.922	0.0070
	3	0.935	0.0044
	6	0.954	0.0044
	8	**0.954**	**0.0029**
	9	0.951	0.0030
	10	0.917	0.0049
	12	0.909	0.0047
	15	0.895	0.0158
SD Model	2	0.830	0.0063
	6	0.910	0.0176
	7	**0.937**	**0.0036**
	8	0.856	0.0034
	9	0.930	0.0046
	10	0.885	0.0047
	12	0.922	0.0055
	15	0.724	0.0112
Chl-a Model	3	0.815	0.0087
	5	0.822	0.0085
	12	0.825	0.0068
	15	0.873	0.0057
	18	**0.938**	**0.0048**
	20	0.903	0.0089
	22	0.887	0.0042
	25	0.702	0.0198
	30	0.743	0.0228
	35	0.906	0.0292
	40	0.884	0.0208
	45	0.887	0.0148
	50	0.706	0.0406
	55	0.791	0.0076
	58	0.723	0.1176

After finalization of input parameters for each model, all the input and output data were normalized in the range 0.15 to 0.85 to increase the training efficiency. For

eutrophication forecasting, optimized networks of structure having neurons in input, hidden and output layer as 5-8-1, 5-7-1 and 9-18-1 were finalized for DO, SD and Chl-a models respectively. The results of ANN model training are illustrated as plot between the observed vs predicted values of for DO-I and DO-III model in Figure 7.23 and Figure 7.24. SD-I and SD-III model training results are shown in Figure 7.25 and Figure 7.26. Figure 7.27 and Figure 7.28 illustrates the performance of trained Chl-a-I and Chl-a-III model respectively. The model performance was evaluated with correlation parameters R^2 and E and error estimation parameters RMSE and MAE; results for which were shown in Table 23. It is evident from the figures that, very good correlation exists between observed values and model predicted values for the investigated parameters. It was observed that the performance of all the neural network DO, SD and Chl-a models were quite satisfactory as reflected by higher R^2 and E values. R^2 value of 0.99 was obtained for all the trained models. The RMSE and MAE between observed and predicted values were also low considering range of DO, SD, and Chl-a dataset. Compared to DO and Chl-a models, the goodness of fit for SD-I and SD-III ANN models were inferior as indicated by slightly lower E values. Performance of models developed through dataset of Tank 1 and Tank 3 were quite comparable in terms of R^2 and E values. However slightly lesser error values were observed for DO-I and Chl-a-I models compared to DO-III and Chl-a-III models.

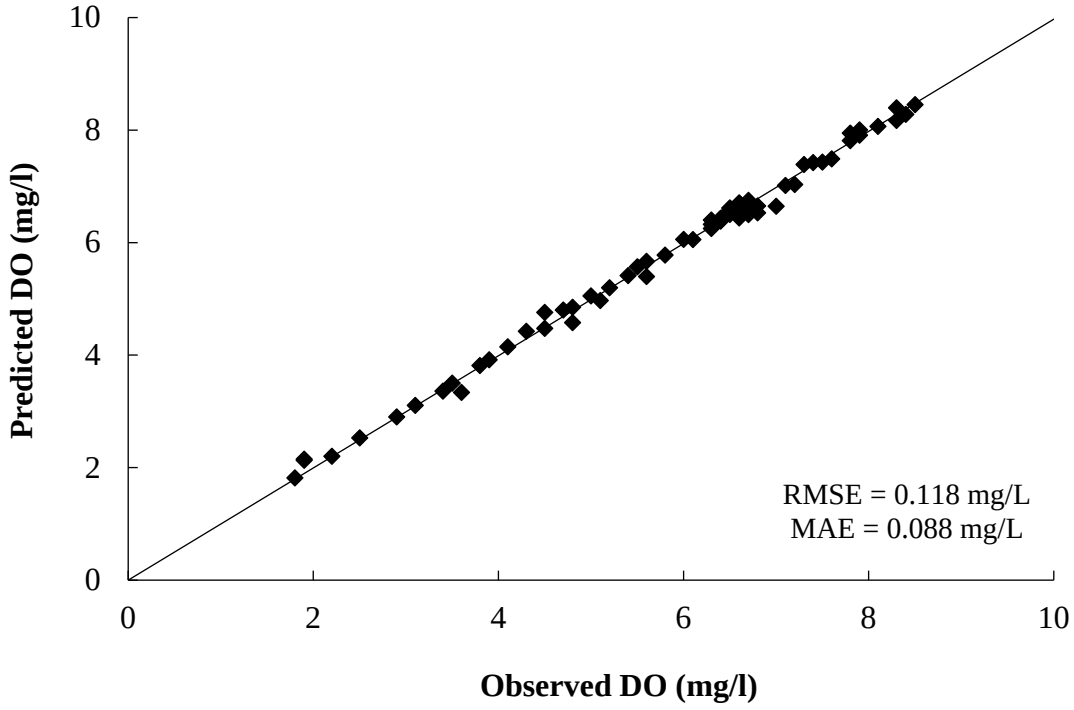

Figure 7.23: Observed vs Predicted plot of DO-I model under ANN (Trial 2)

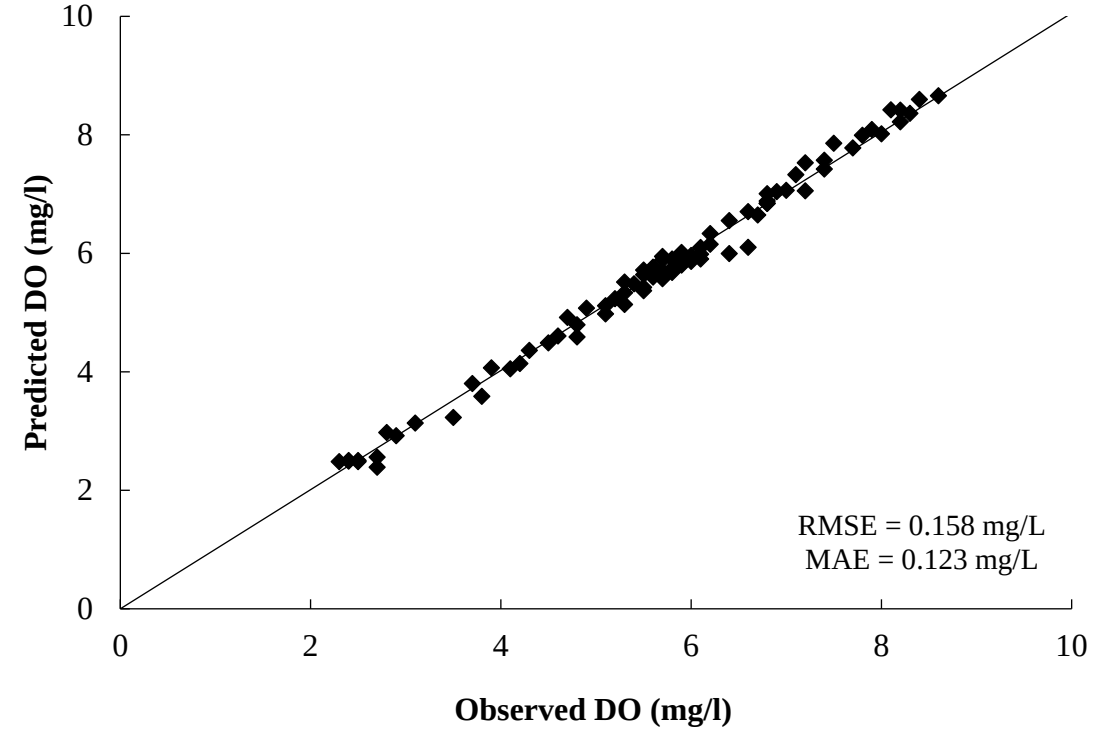

Figure 7.24: Observed vs Predicted plot of DO-III model under ANN (Trial 2)

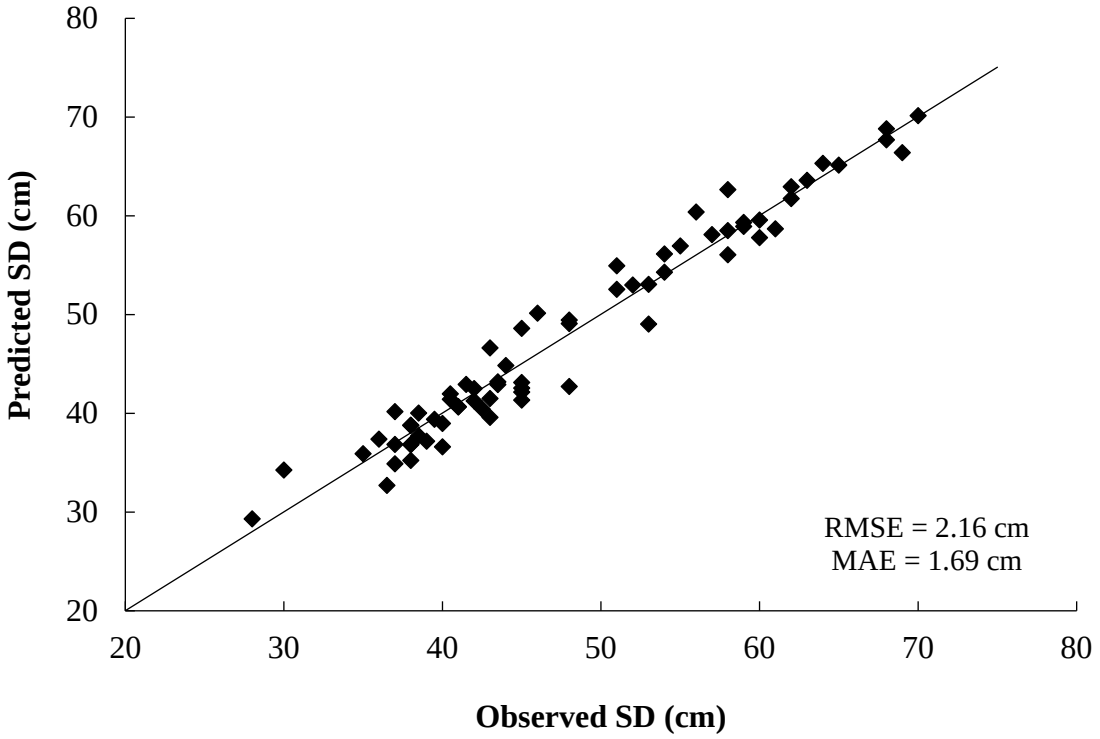

Figure 7.25: Observed vs Predicted plot of SD-I model under ANN (Trial 2)

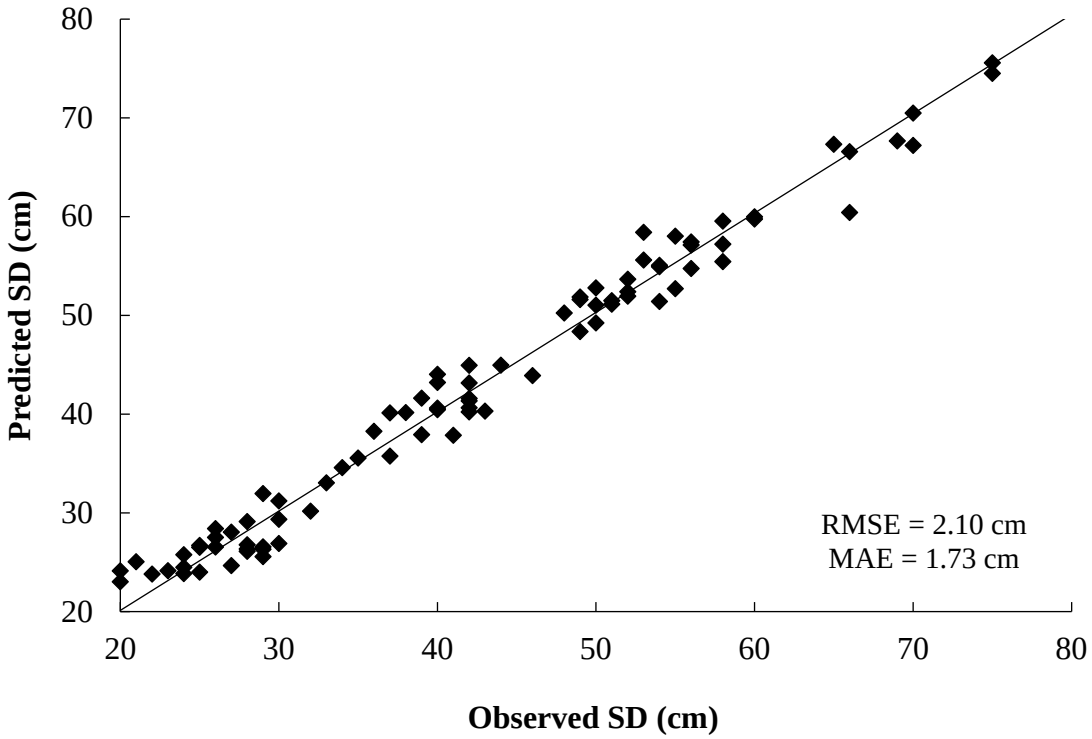

Figure 7.26: Observed vs Predicted plot of SD-III model under ANN (Trial 2)

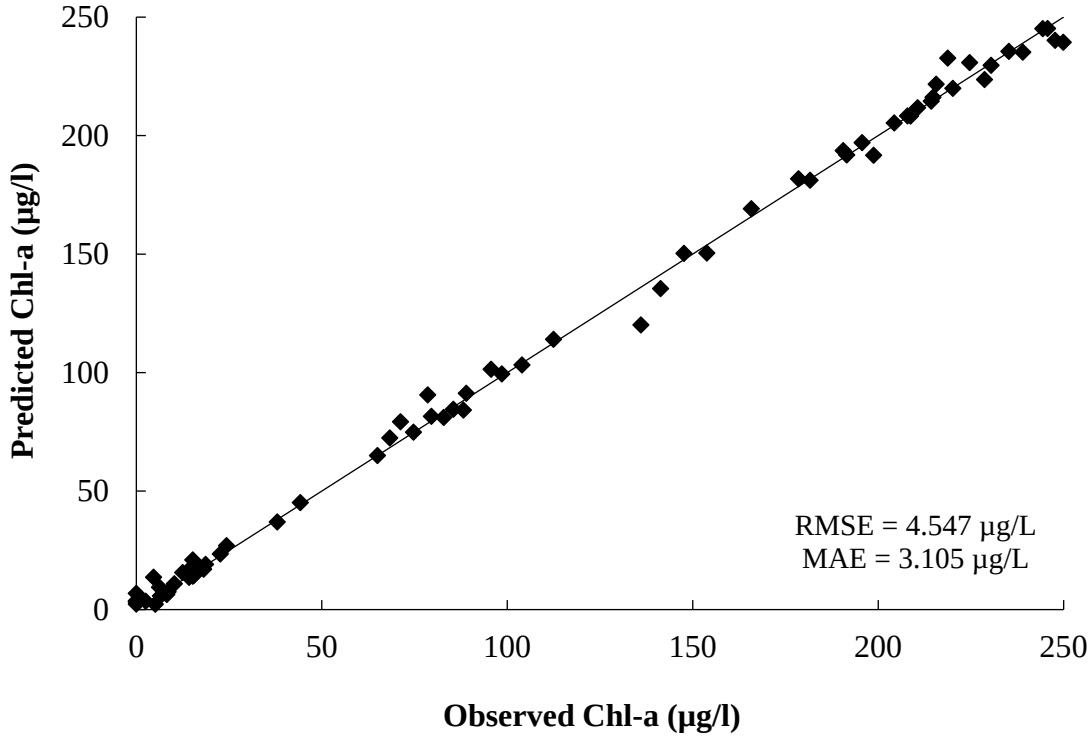

Figure 7.27: Observed vs Predicted plot of Chl-a-I model under ANN (Trial 2)

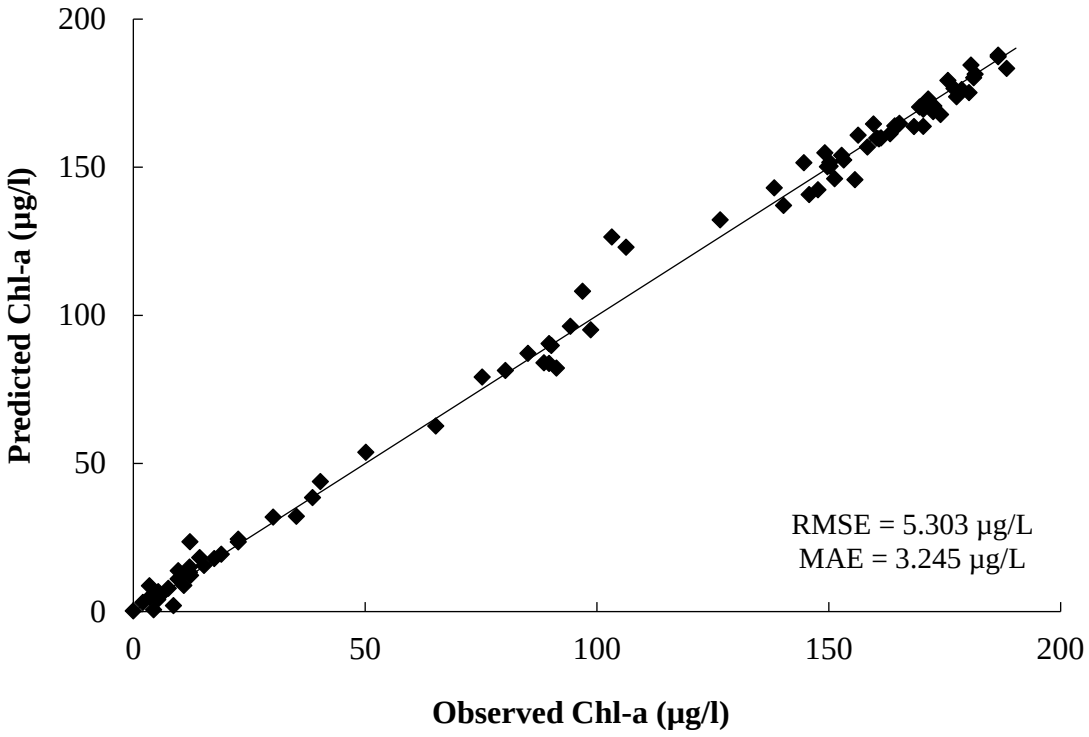

Figure 7.28: Observed vs Predicted plot of Chl-a-III model under ANN (Trial 2)

7.2.3 GPR Models

GPR with squared exponential kernel and a 5-fold cross-validation technique was used to model DO, SD, and Chl-a similar to 1st trial. Models trained under GPR revealed very good correlation between model predicted and observed values of DO, SD, and Chl-a. Figure 7.29 and Figure 7.30 shows the performance of DO-I and DO-III models. SD-I and SD-III model results are presented with Figure 7.31 and Figure 7.32. The training performance of Chl-a-I and Chl-a-III models are shown in Figure 7.33 and Figure 7.34 respectively. The statistical evaluation of model training performance has been summarised in Table 7.11. The training efficiency of the models were quite high as reflected by higher R^2 and E values and lower RMSE and MAE values. R^2 and E value greater than 0.99 and 0.98 respectively were obtained for all the GPR models. RMSE was found as 0.09 mg/L and 0.12 mg/L for DO-I and DO-III; 1.20 cm and 1.79 cm for SD-I and SD-III; 3.79 µg/L and 4.16 µg/L for the Chl-a-I and Chl-a-III models respectively. It was found that the error estimation parameters were lower for Tank 1 based DO, SD, and Chl-a models compared to Tank 3 counterparts.

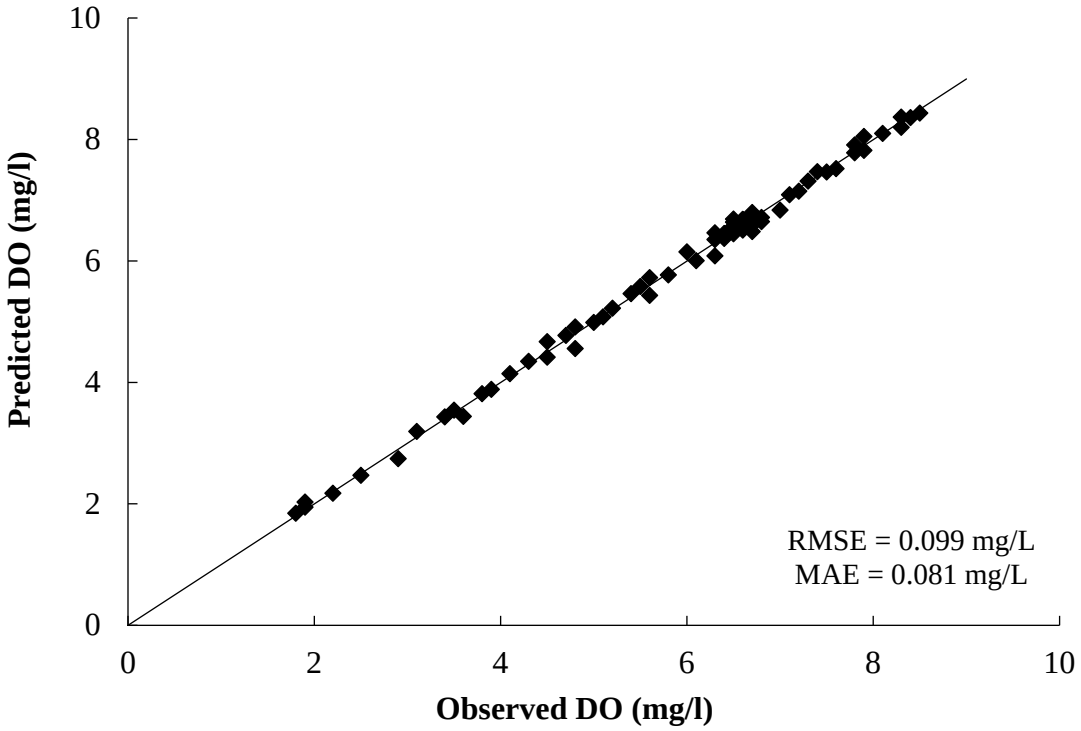

Figure 7.29: Observed vs Predicted plot of DO-I model under GPR (Trial 2)

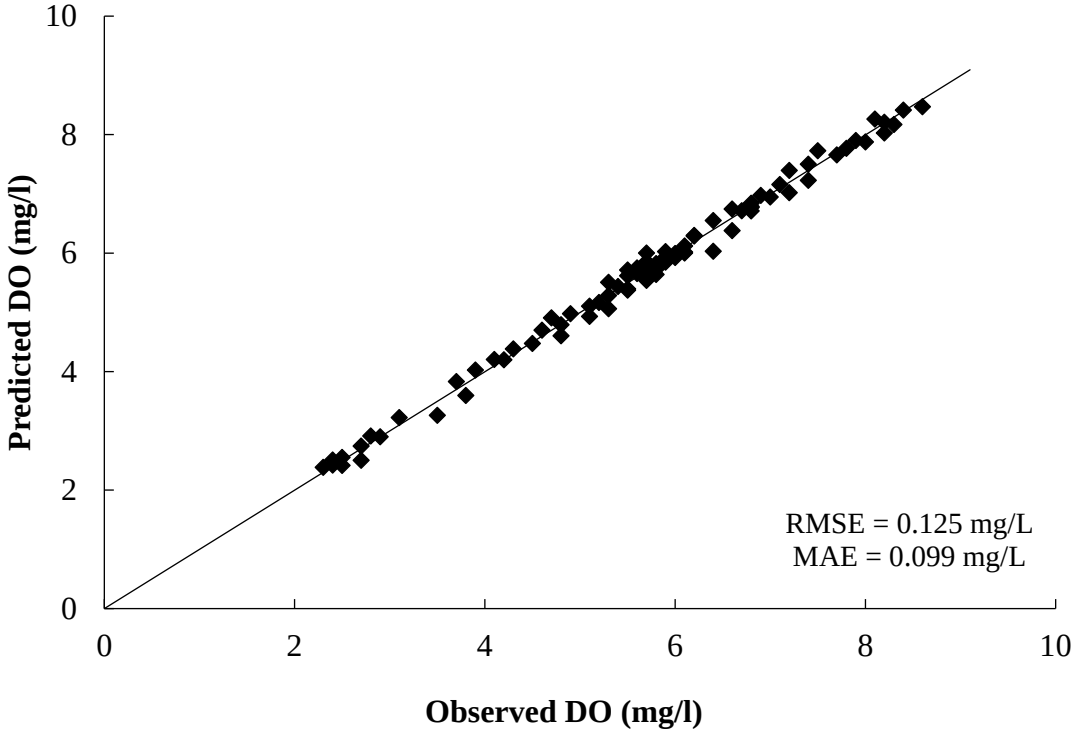

Figure 7.30: Observed vs Predicted plot of DO-III model under GPR (Trial 2)

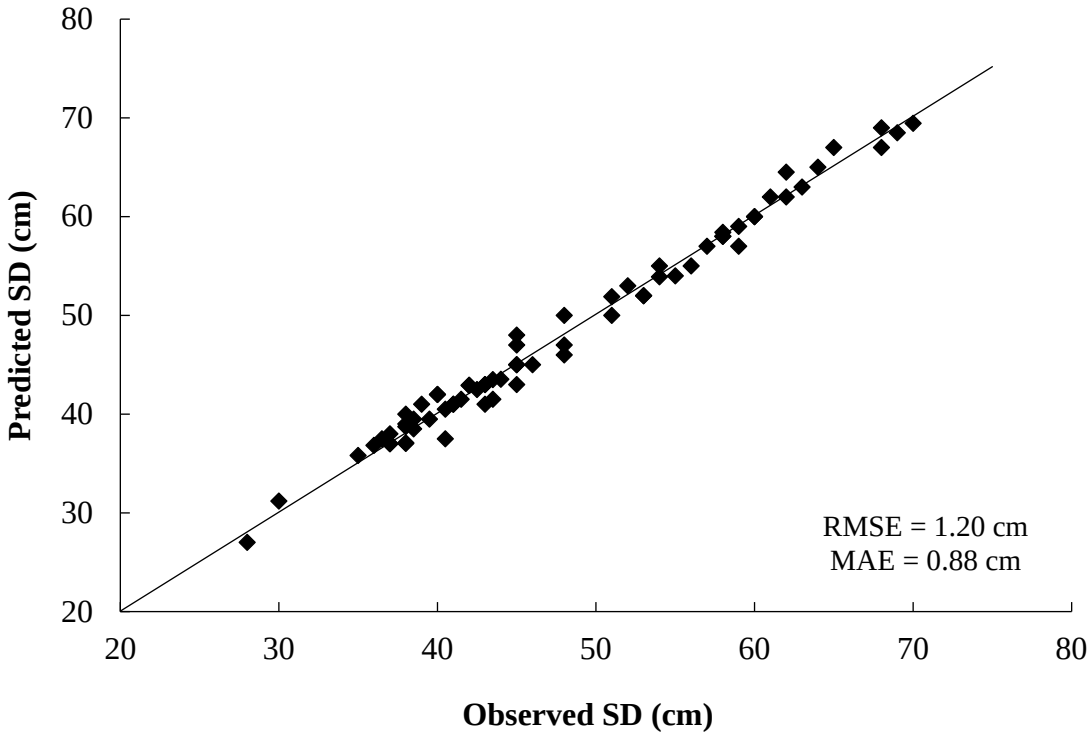

Figure 7.31: Observed vs Predicted plot of SD-I model under GPR (Trial 2)

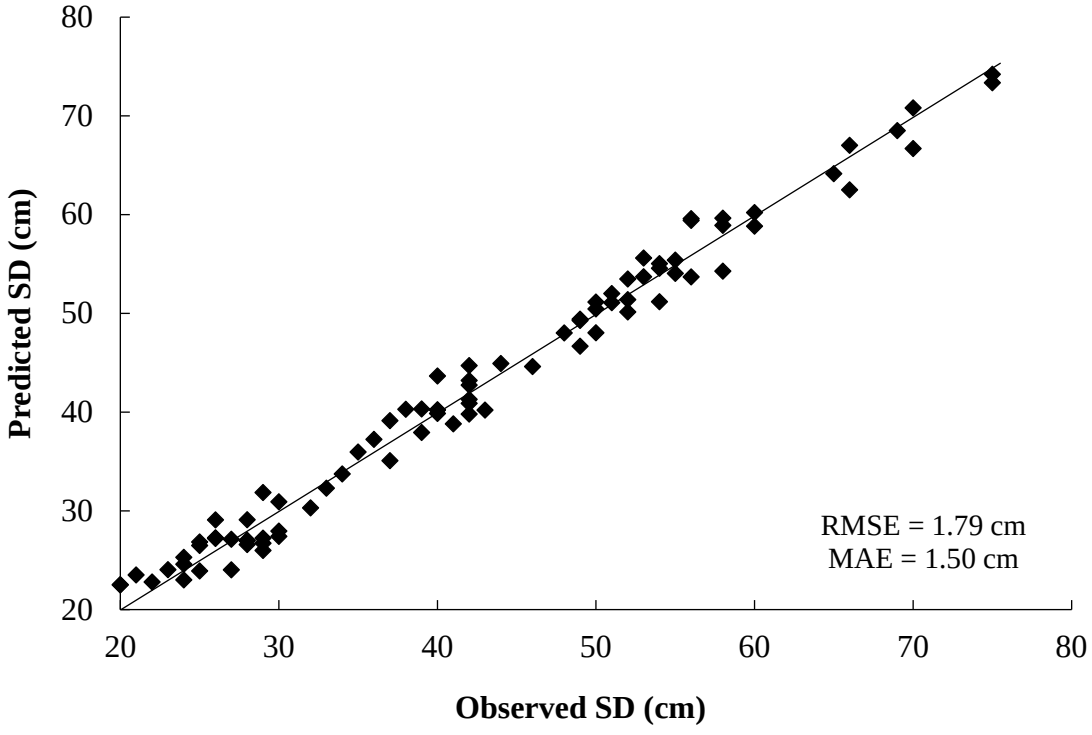

Figure 7.32: Observed vs Predicted plot of SD-III model under GPR (Trial 2)

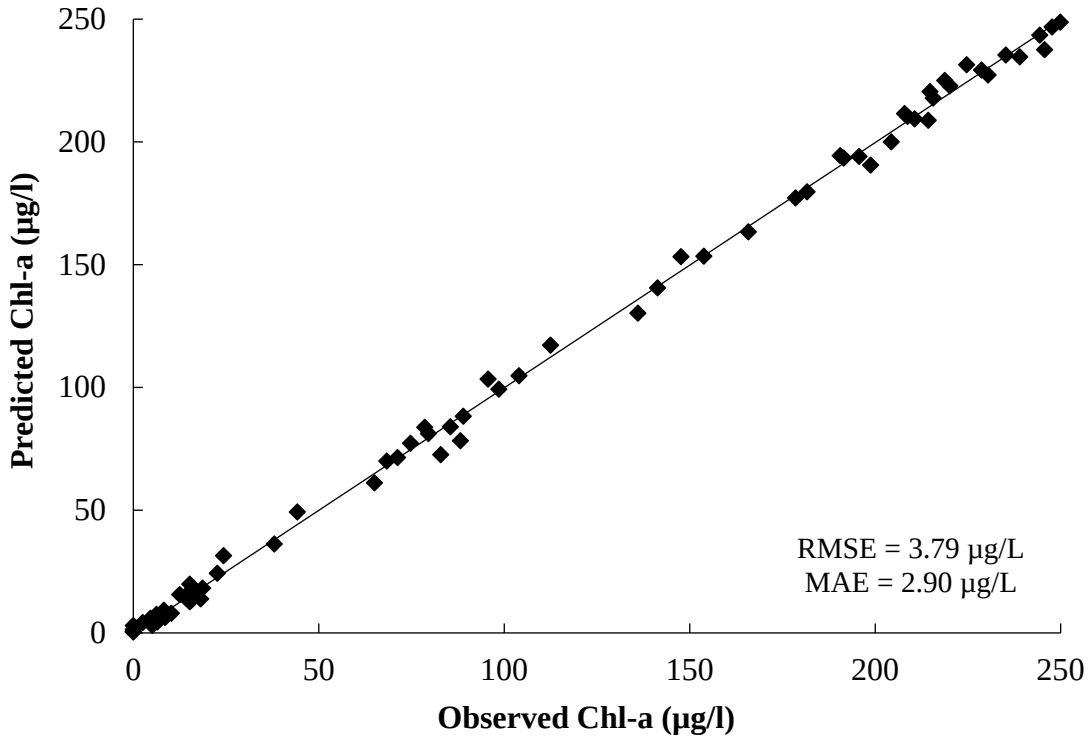

Figure 7.33: Observed vs Predicted plot of Chl-a-I model under GPR (Trial 2)

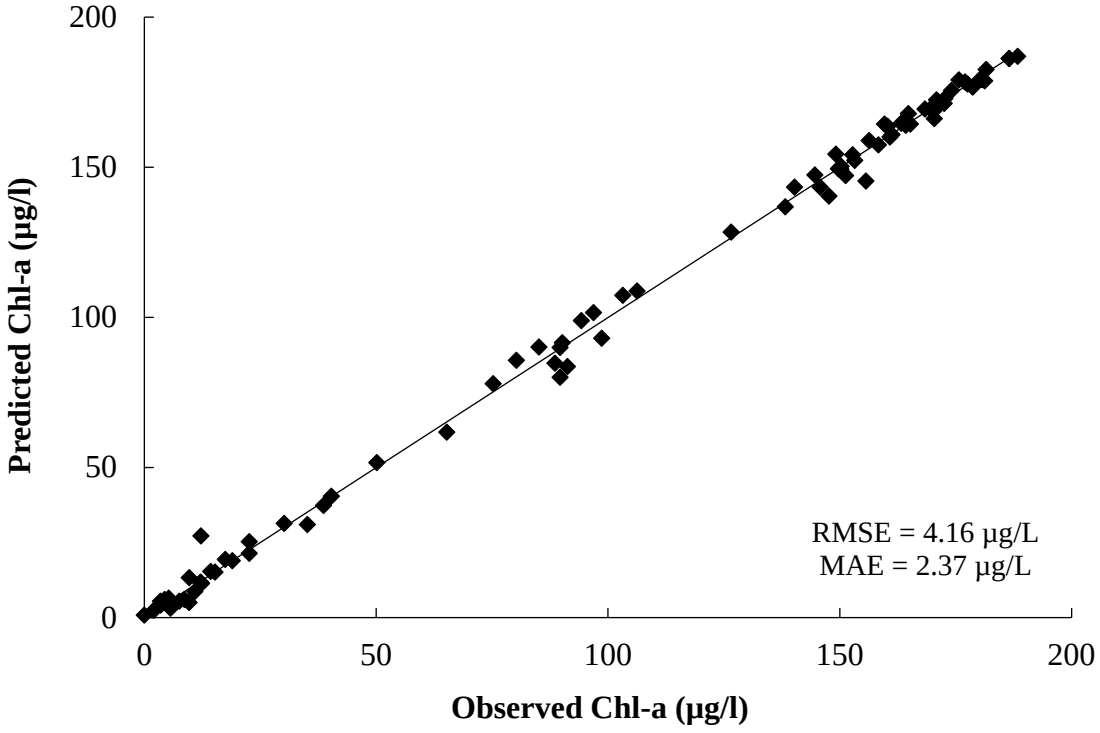

Figure 7.34: Observed vs Predicted plot of Chl-a-III model under GPR (Trial 2)

7.2.4 ANFIS Models

As ANFIS approach has good accuracy in decision making and prediction problems, it was used to model DO, SD, and Chl-a in this study. The required outputs were predicted using the same input parameters as in the case of corresponding ANN and GPR models. The entire normalised input dataset was randomly divided into training and testing datasets in a 4:1 ratio. This study used a subtractive clustering method combined with a hybrid learning algorithm to develop the best fitting ANFIS model for the specified output parameters. The 5-layer ANFIS structure of DO, SD, and Chl-a model in MATLAB environment is illustrated with Figure 7.35. Under subtractive clustering, different combinations of range of influence and squash factor values were considered, and each FIS systems were trained using a hybrid optimization

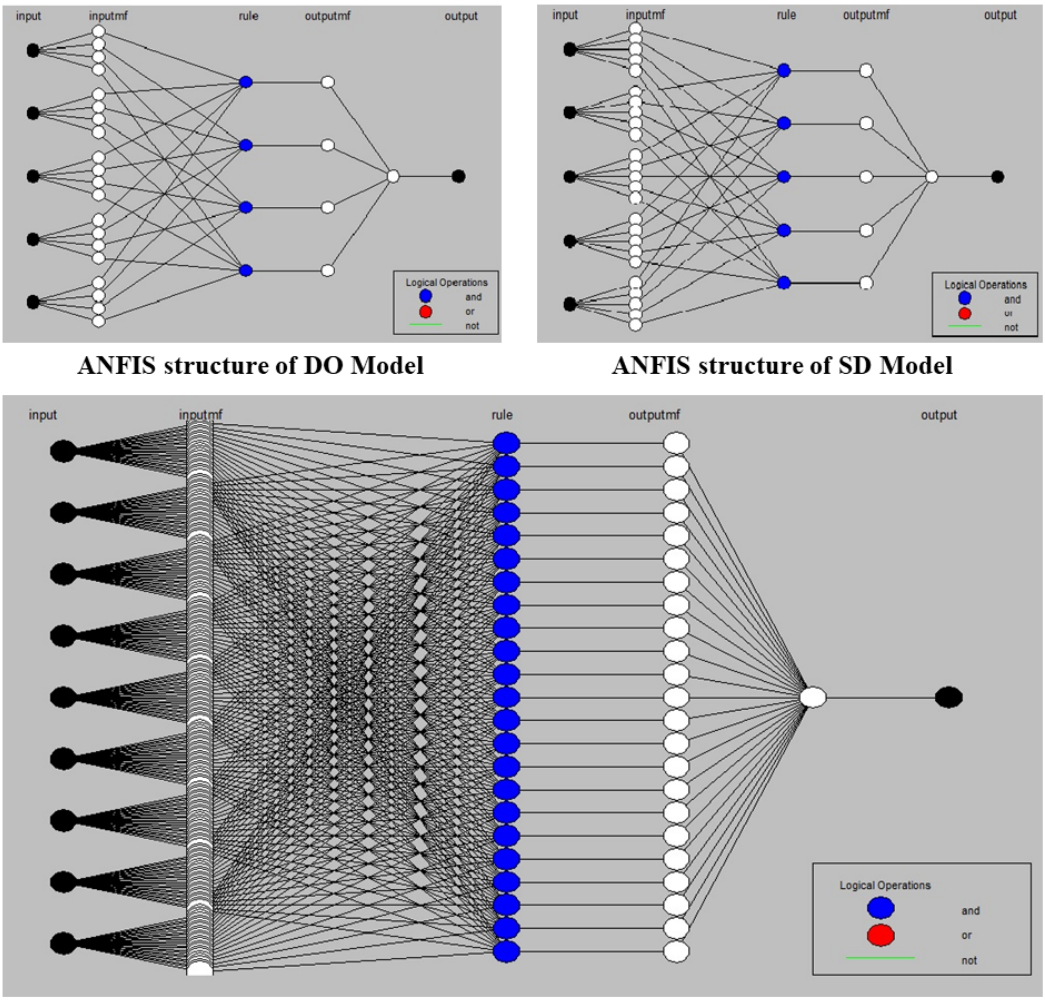

Figure 7.35: 5-layer ANFIS structure of developed models in MATLAB

method. Using a comprehensive methodology, testing different parameter combinations of subtractive clustering, the best structure of ANFIS was finalized with parameters 0.5 (range of influence), 1.5 (Squash factor), 0.5 (Accept ratio), 0.15 (Reject ratio) and 100 (no. of epoch).

The optimized network of ANFIS was used to train the DO, SD and Chl-a models and Table 7.11 shows the statistical performance results of the trained models. Figure 7.36 and Figure 7.37 illustrates the correlation achieved through ANFIS training between observed and predicted values for DO-I and DO-III models. The performance results of SD-I and SD-III model training is presented with Figure 7.38 and Figure 7.39 respectively. Training performance of Chl-a-I and Chl-a-III models have been shown with Figure 7.40 and Figure 7.41 respectively. From these figures it can be observed that very good correlation was there between actual and ANFIS forecasted values for all the model types. The goodness of fit parameters R^2 and E which were of the order 0.99 for all the models trained under ANFIS topology. The errors in target prediction was also smaller for each model types trained under ANFIS. The error in target prediction as indicated by RMSE and MAE values for DO-I model was lesser than the DO-III model. For the SD-I and Chl-a-I model the error parameters were slightly higher than the SD-III and Chl-a-III models respectively

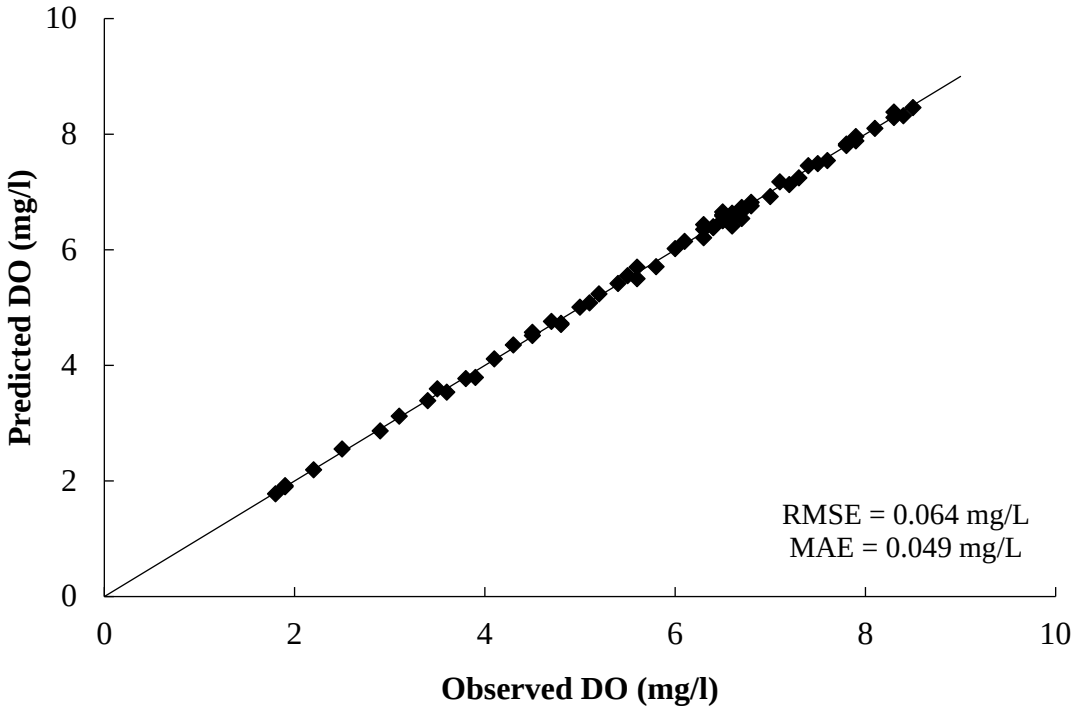

Figure 7.36: Observed vs Predicted plot of DO-I model under ANFIS (Trial 2)

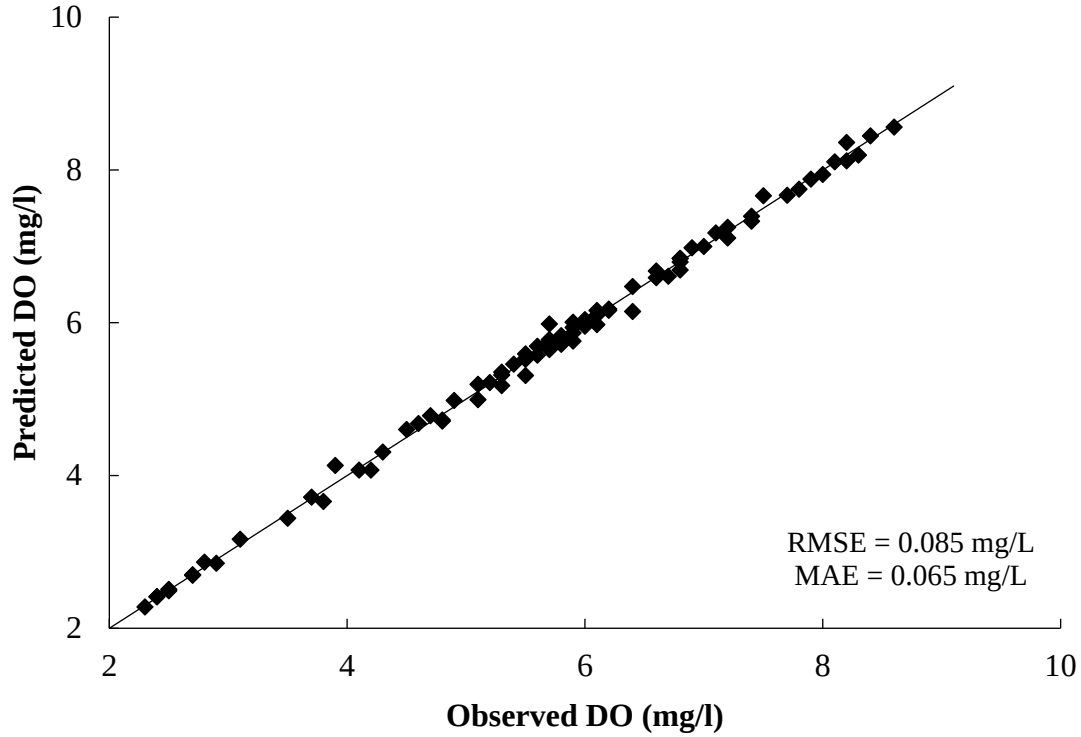

Figure 7.37: Observed vs Predicted plot of DO-III model under ANFIS (Trial 2)

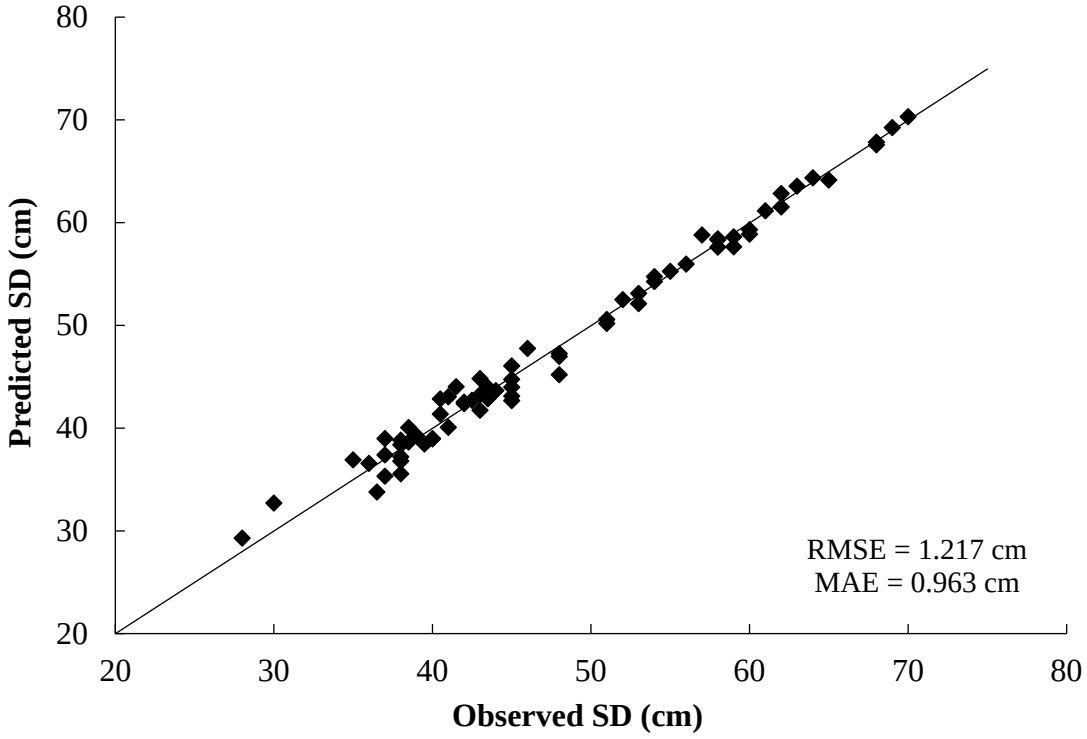

Figure 7.38: Observed vs Predicted plot of SD-I model under ANFIS (Trial 2)

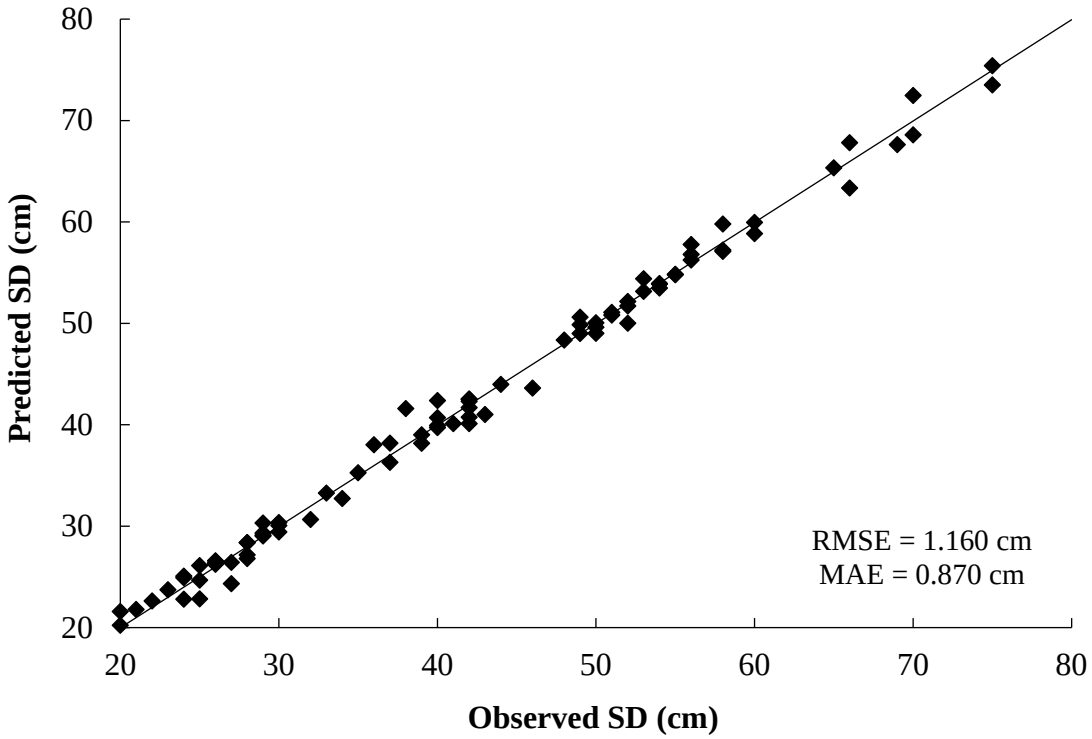

Figure 7.39: Observed vs Predicted plot of SD-III model under ANFIS (Trial 2)

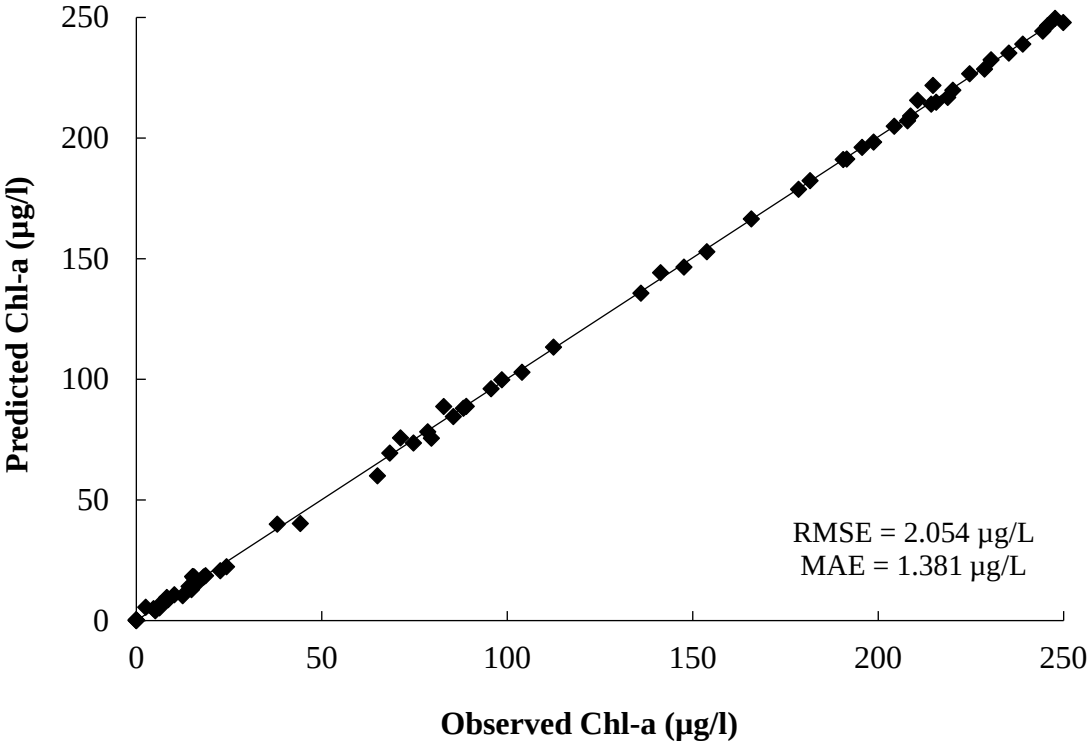

Figure 7.40: Observed vs Predicted plot of Chl-a-I model under ANFIS (Trial 2)

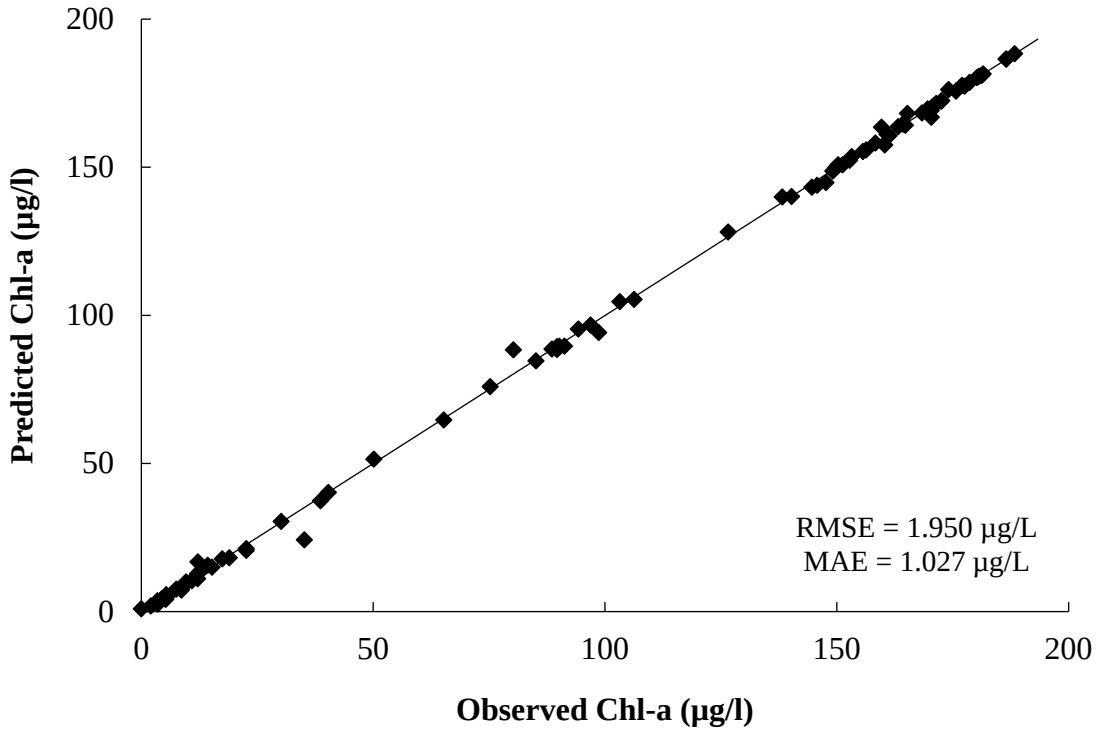

Figure 7.41: Observed vs Predicted plot of Chl-a-III model under ANFIS (Trial 2)

Comparing the training results of ANN, GPR, and ANFIS models from Table 7.11, it was observed that for all the developed models good corelation existed between actual and model predicted values. The E values of DO-III, SD-I and SD-III models improved in GPR and ANFIS training as compared to ANN. The error in model prediction as given by RMSE and MAE values were also lesser in case of GPR models compared to ANN models. The errors in target prediction further reduced during ANFIS training compared to GPR models. Overall, the training performance of ANFIS models were found to be more robust than the ANN and GPR models. Another important observation can be made from the performance of the trained models that both Tank 1 and Tank 3 data based models produced satisfactory results. However, slightly lesser error in target prediction has been observed in general for Tank 1 models under ANN, GPR, and ANFIS methods compared to its Tank 3 counterparts.

7.2.5 Model Validation

The well trained DO, SD and Chl-a models from Tank 1 and Tank 3 data were tested with natural waterbody data as mentioned in Table 6.2 to check the prediction performance of the models under natural condition. Twenty five samples were collected

from the six sampling points and their DO, SD and Chl-a values were compared with model predicted values. Model simulation results of the DO, SD and Chl-a models revealed that the developed models were able to predict the output values within considerable accuracy. The testing performance of models under ANN are presented with Figure 7.42 and Figure 7.43 for DO-I and DO-III, Figure 7.44 and Figure 7.45 for SD-I and SD-III, and Figure 7.46 and Figure 7.47 for Chl-a-I and Chl-a-III respectively. R^2 and E values greater than 0.90 was obtained between observed and predicted values and RMSE and MAE were also small for the DO models trained under ANN environment.

Table 7.11: Performance result of ANN, GPR and ANFIS model training (Trial 2)

Model	R^2	E	RMSE	MAE
ANN Models				
DO-I	0.999	0.995	0.118 mg/L	0.088 mg/L
DO-III	0.999	0.983	0.158 mg/L	0.123 mg/L
SD-I	0.998	0.956	2.160 cm	1.690 cm
SD-III	0.998	0.978	2.100 cm	1.730 cm
Chl-a-I	0.999	0.997	4.547 µg/L	3.105 µg/L
Chl-a-III	0.998	0.994	5.303 µg/L	3.245 µg/L
GPR Models				
DO-I	0.999	0.996	0.099 mg/L	0.081 mg/L
DO-III	0.999	0.983	0.125 mg/L	0.099 mg/L
SD-I	0.999	0.986	1.20 cm	0.88 cm
SD-III	0.998	0.984	1.79 cm	1.50 cm
Chl-a-I	0.999	0.998	3.79 µg/L	2.90 µg/L
Chl-a-III	0.998	0.996	4.16 µg/L	2.37 µg/L
ANFIS Models				
DO-I	0.999	0.998	0.064 mg/L	0.049 mg/L
DO-III	0.999	0.997	0.085 mg/L	0.065 mg/L
SD-I	0.999	0.986	1.217 cm	0.963 cm
SD-III	0.999	0.993	1.160 cm	0.870 cm
Chl-a-I	0.999	0.999	2.054 µg/L	1.381 µg/L
Chl-a-III	0.999	0.999	1.950 µg/L	1.027 µg/L

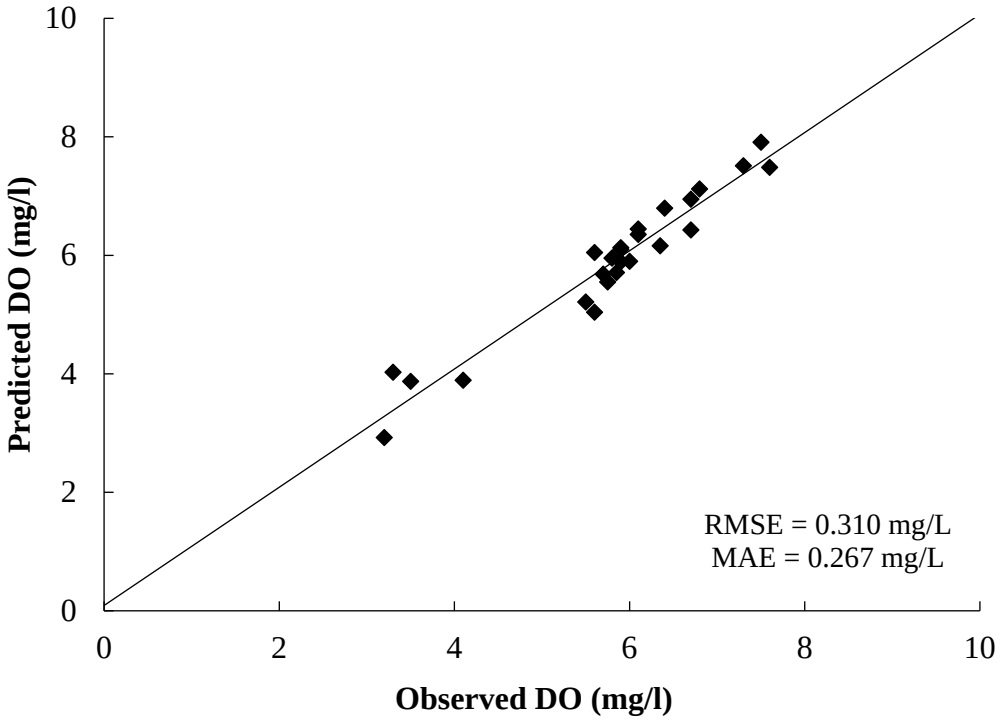

Figure 7.42: Prediction performance of DO-I ANN model testing against natural water body data (Trial 2)

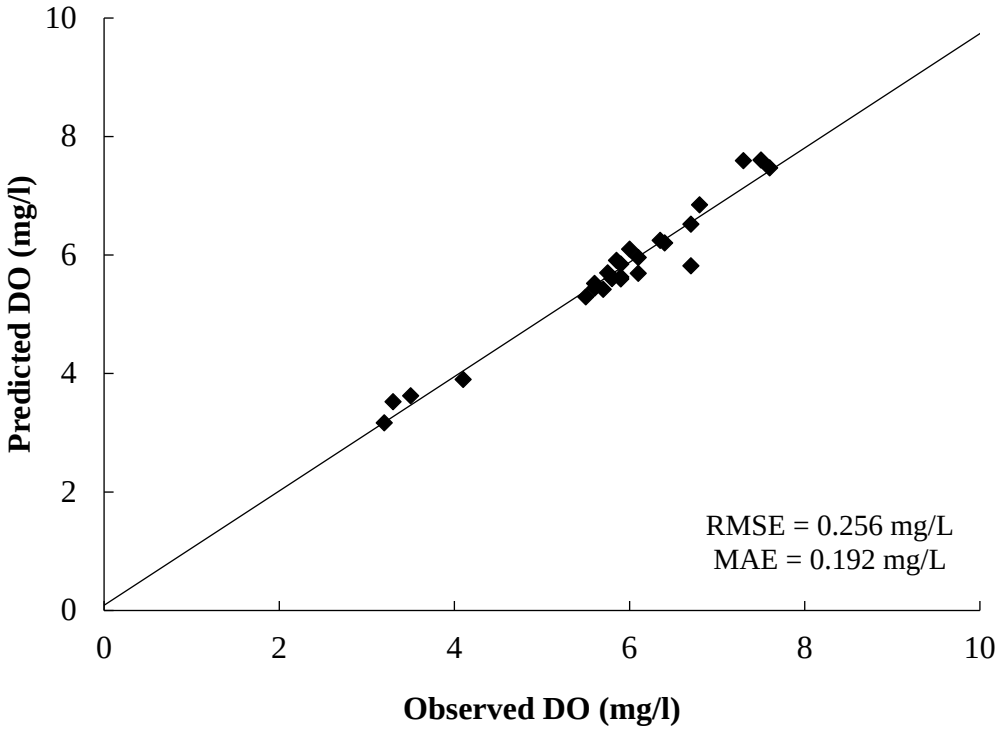

Figure 7.43: Prediction performance of DO-III ANN model testing against natural water body data (Trial 2)

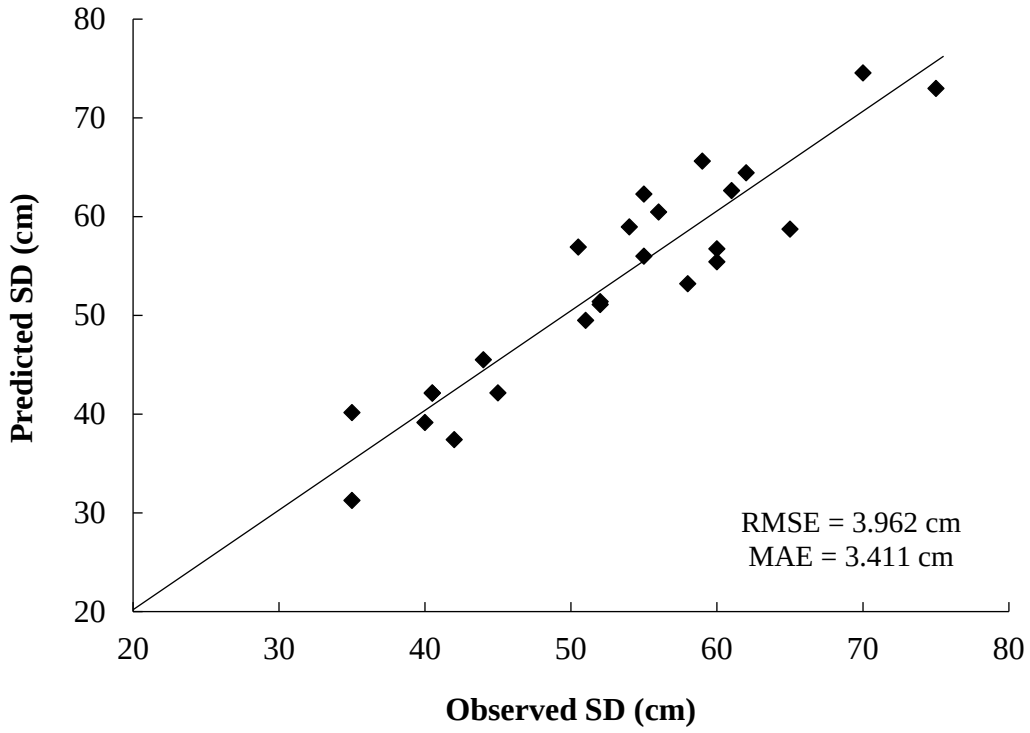

Figure 7.44: Prediction performance of SD-I ANN model testing against natural water body data (Trial 2)

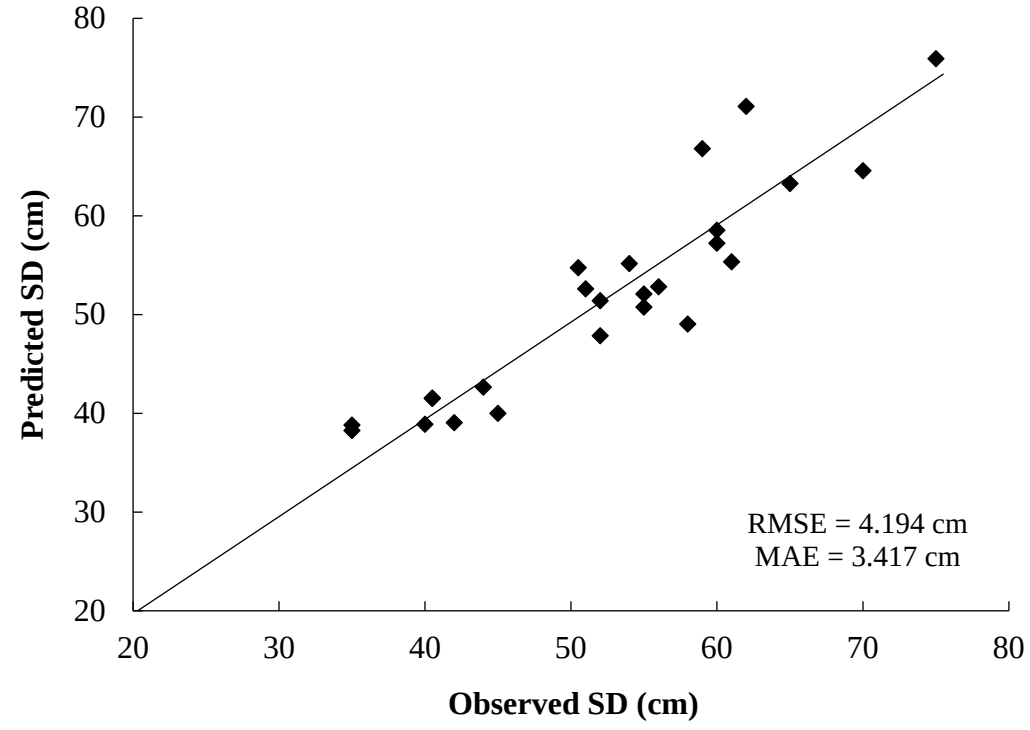

Figure 7.45: Prediction performance of SD-III ANN model testing against natural water body data (Trial 2)

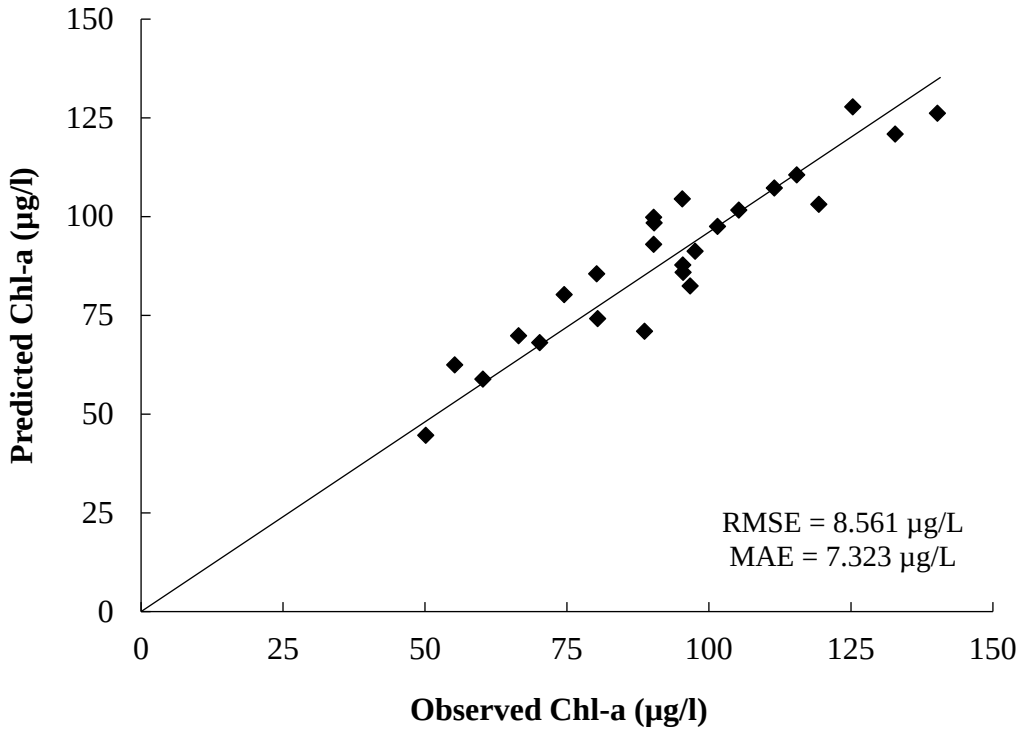

Figure 7.46: Prediction performance of Chl-a-I ANN model testing against natural water body data (Trial 2)

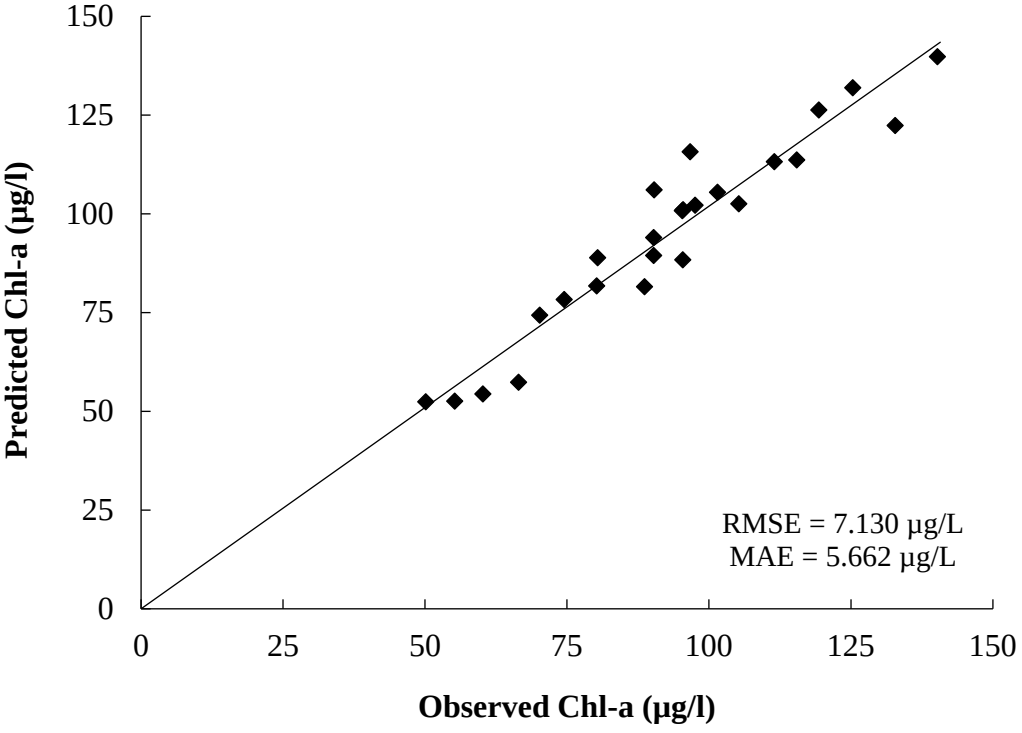

Figure 7.47: Prediction performance of Chl-a-III ANN model testing against natural water body data (Trial 2)

For SD-I and SD-III ANN models R^2 of 0.94 and 0.93 and E values of 0.85 and 0.83 respectively has been achieved during model testing. Very good correlation between observed and model predicted values for Chl-a-I and Chl-a-III models. R^2 values of 0.99 and E values of 0.86 and 0.90 was observed for Chl-a-I and Chl-a-III ANN models respectively during model testing against natural water body data.

The testing performance of GPR based models are presented from Figure 7.48 to Figure 7.53 for DO-I, DO-III, SD-I, SD-III, Chl-a-I, and Chl-a-III respectively. Very good correlation has been achieved during model testing and R^2 of 0.99 was obtained for all the GPR based models. The observed E values were also high and reported as 0.85, 0.89, 0.87, 0.90, 0.83, and 0.84 respectively for DO-I, DO-III, SD-I, SD-III, Chl-a-I, and Chl-a-III models.

Figure 7.54 to Figure 7.59 presents the results of model testing under ANFIS methodology for DO-I, DO-III, SD-I, SD-III, Chl-a-I, and Chl-a-III respectively. It is evident from the figures that the well trained ANFIS models were able to accurately predict the target eutrophication indicators in natural water bodies. R^2 value of 0.99 and E values greater than 0.93 was obtained during testing of all the ANFIS models.

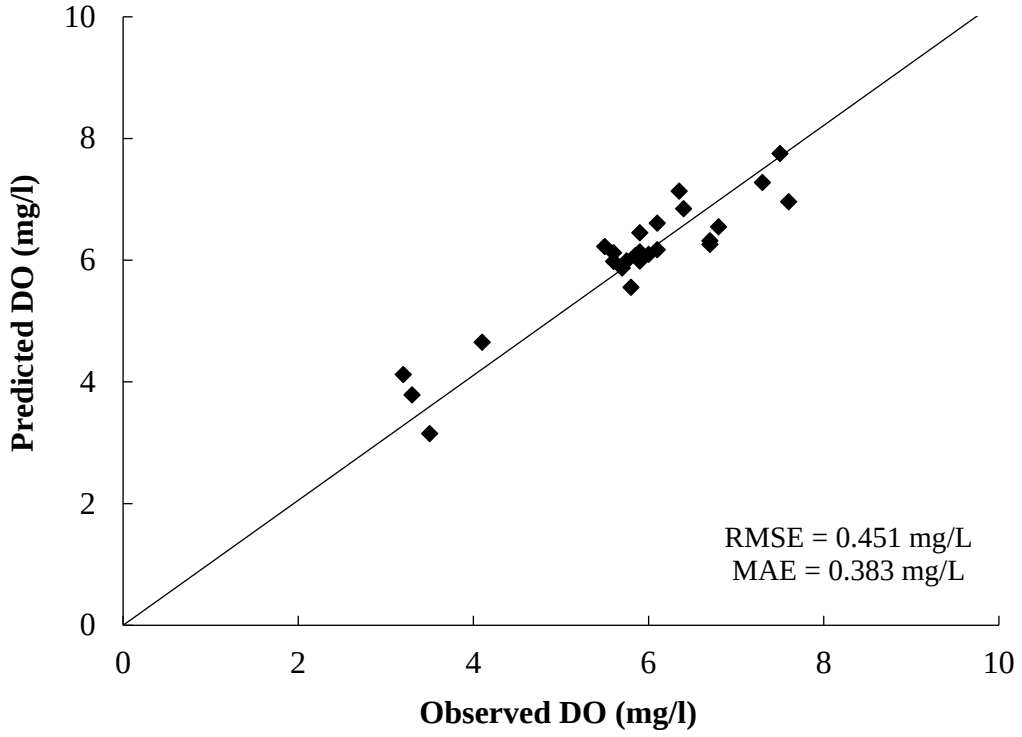

Figure 7.48: Prediction performance of DO-I GPR model testing against natural water body data (Trial 2)

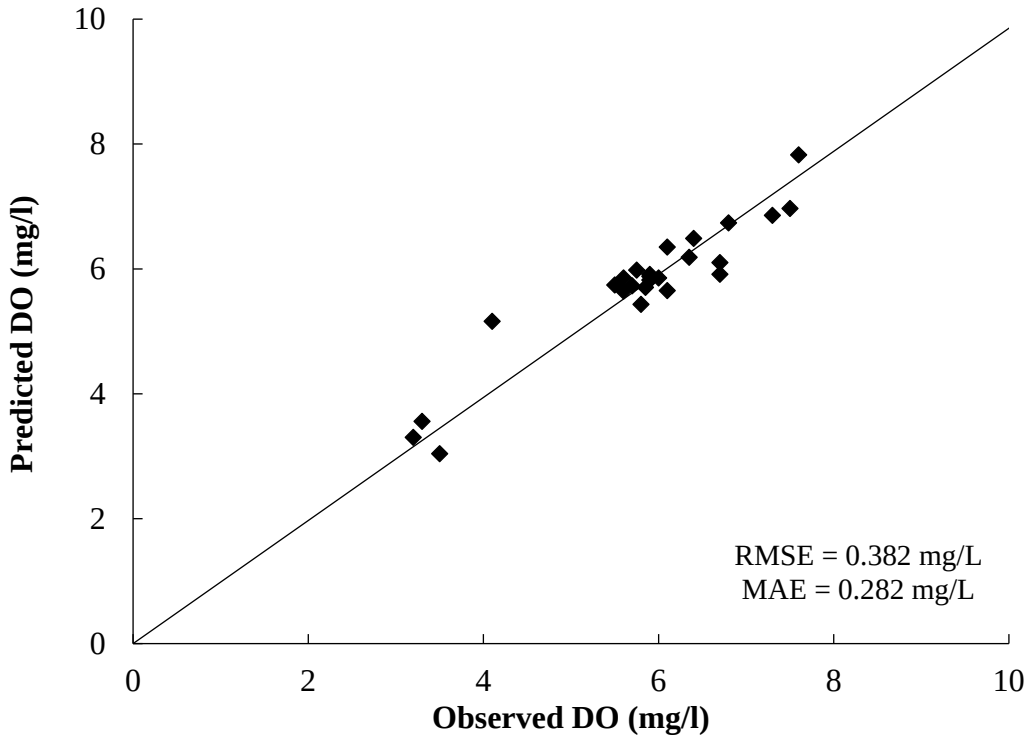

Figure 7.49: Prediction performance of DO-III GPR model testing against natural water body data (Trial 2)

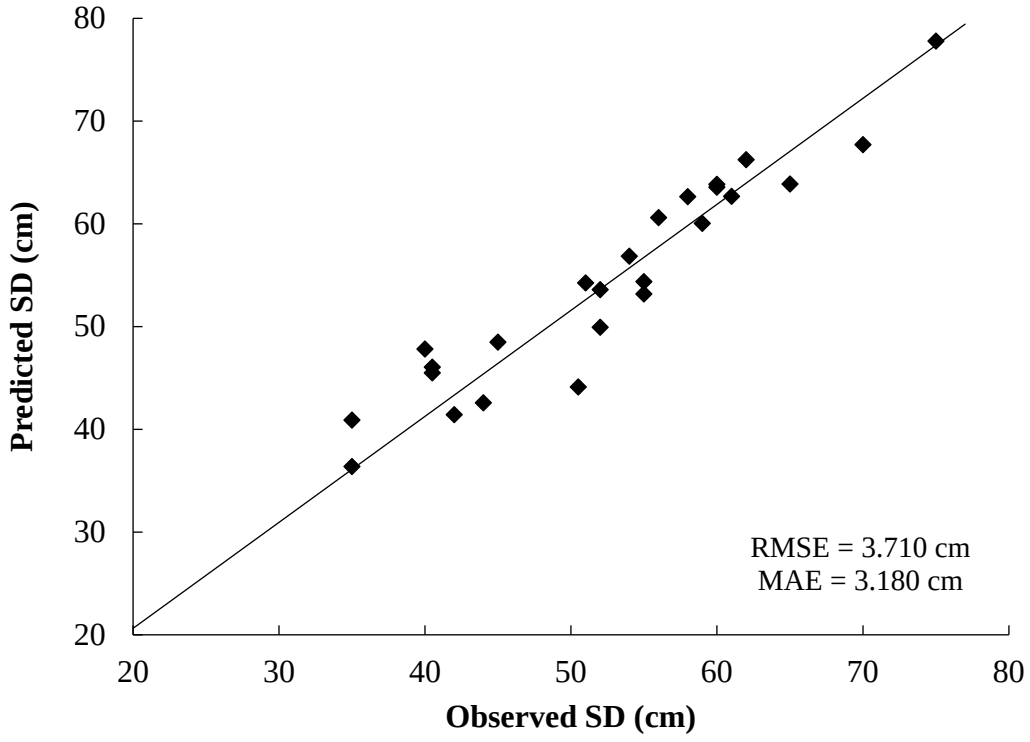

Figure 7.50: Prediction performance of SD-I GPR model testing against natural water body data (Trial 2)

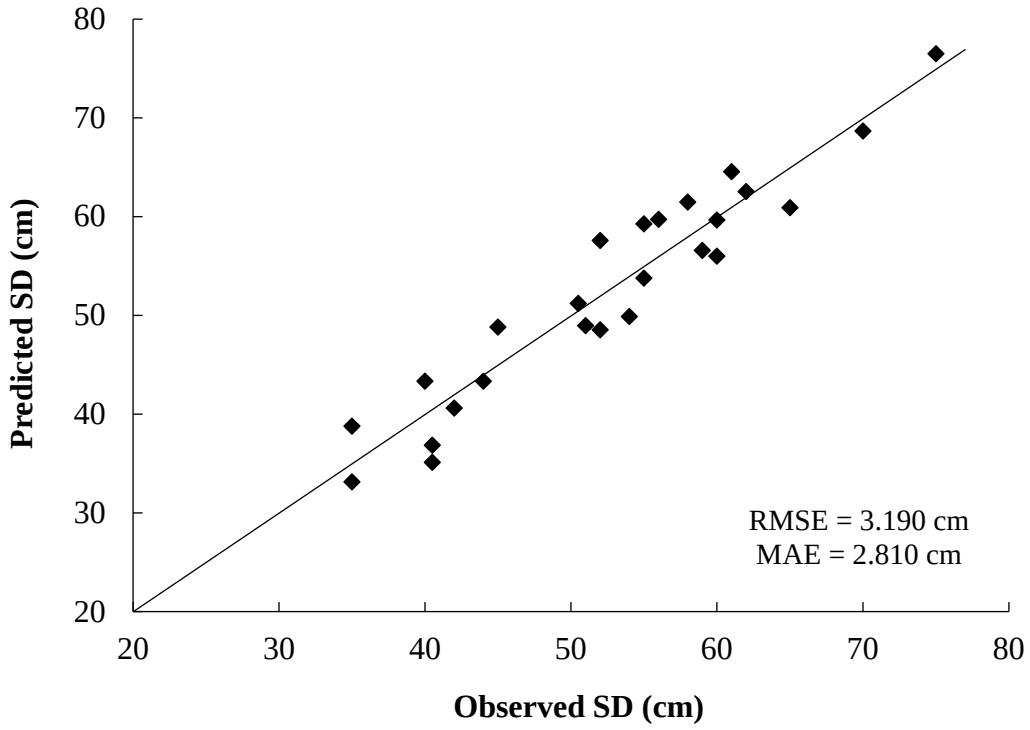

Figure 7.51: Prediction performance of SD-III GPR model testing against natural water body data (Trial 2)

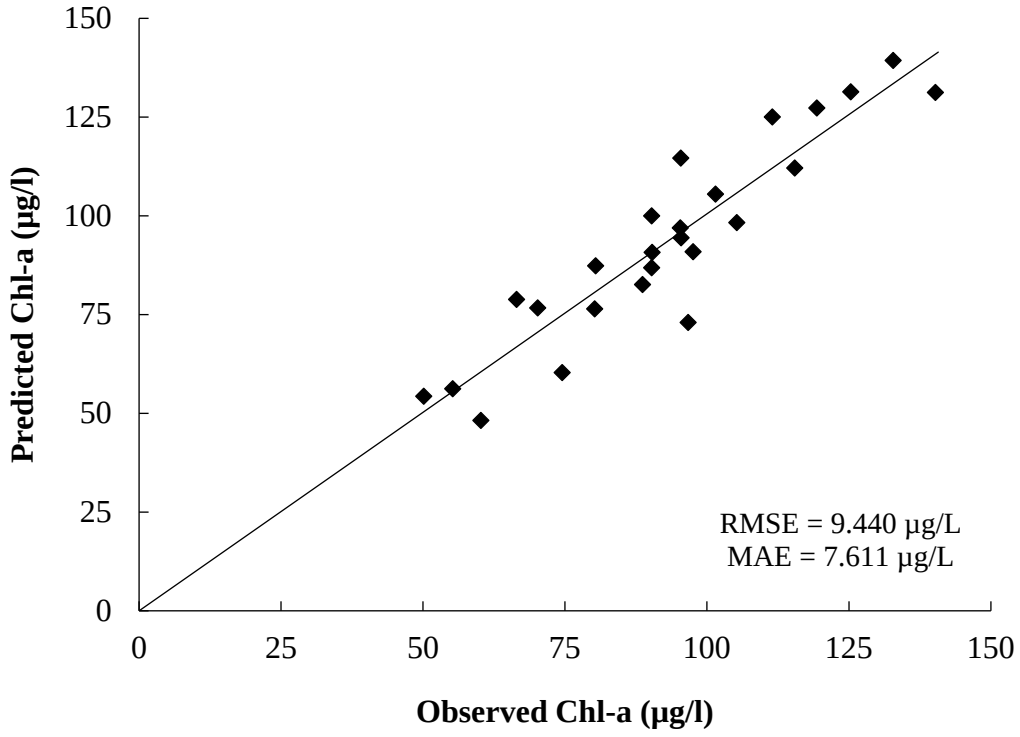

Figure 7.52: Prediction performance of Chl-a-I GPR model testing against natural water body data (Trial 2)

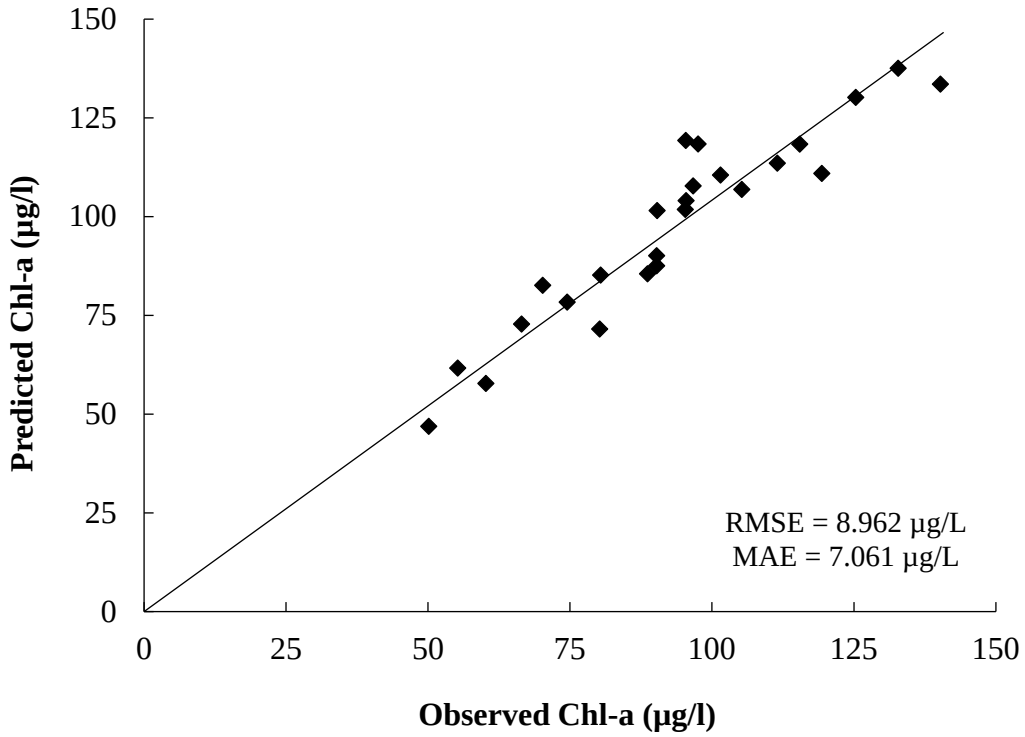

Figure 7.53: Prediction performance of Chl-a-III GPR model testing against natural water body data (Trial 2)

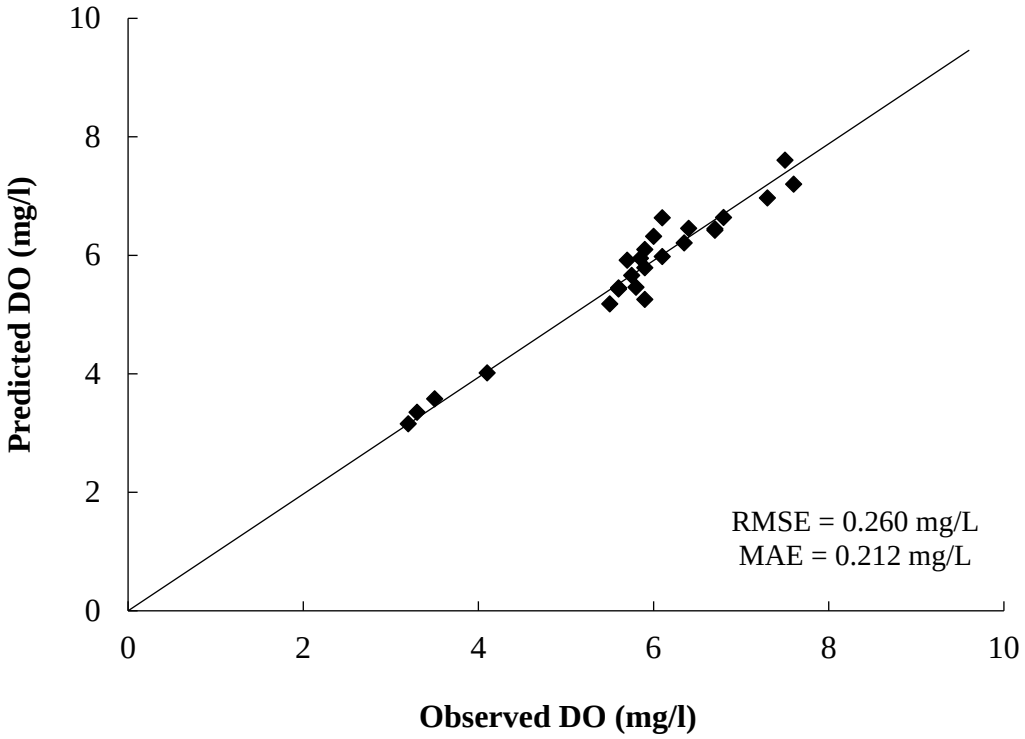

Figure 7.54: Prediction performance of DO-I ANFIS model testing against natural water body data (Trial 2)

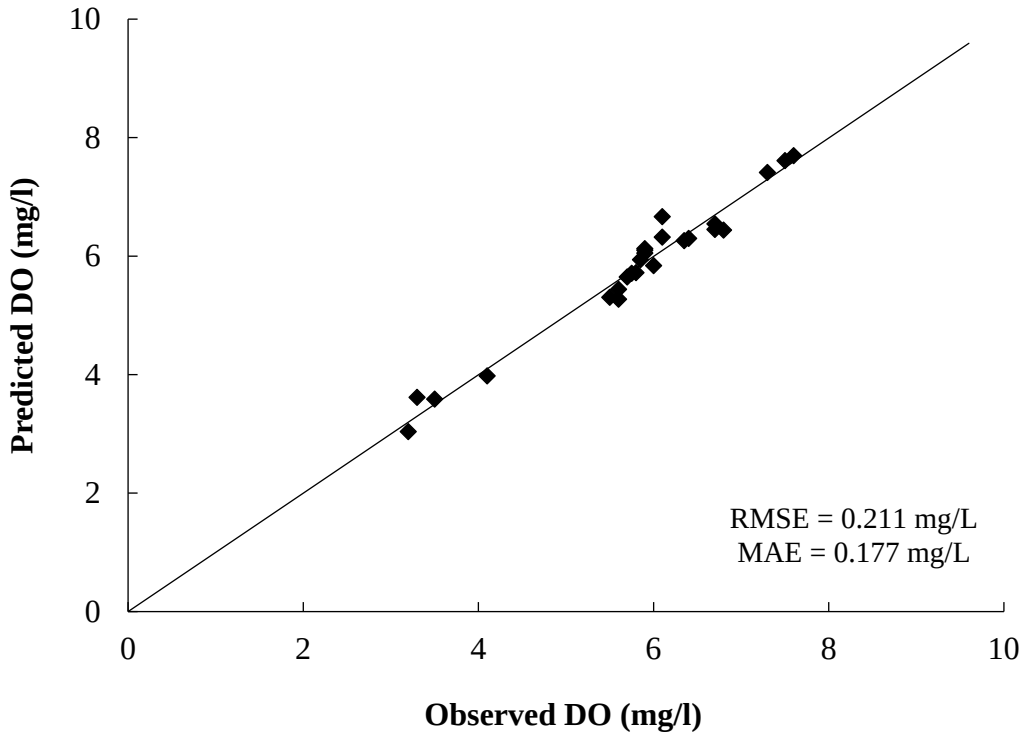

Figure 7.55: Prediction performance of DO-III ANFIS model testing against natural water body data (Trial 2)

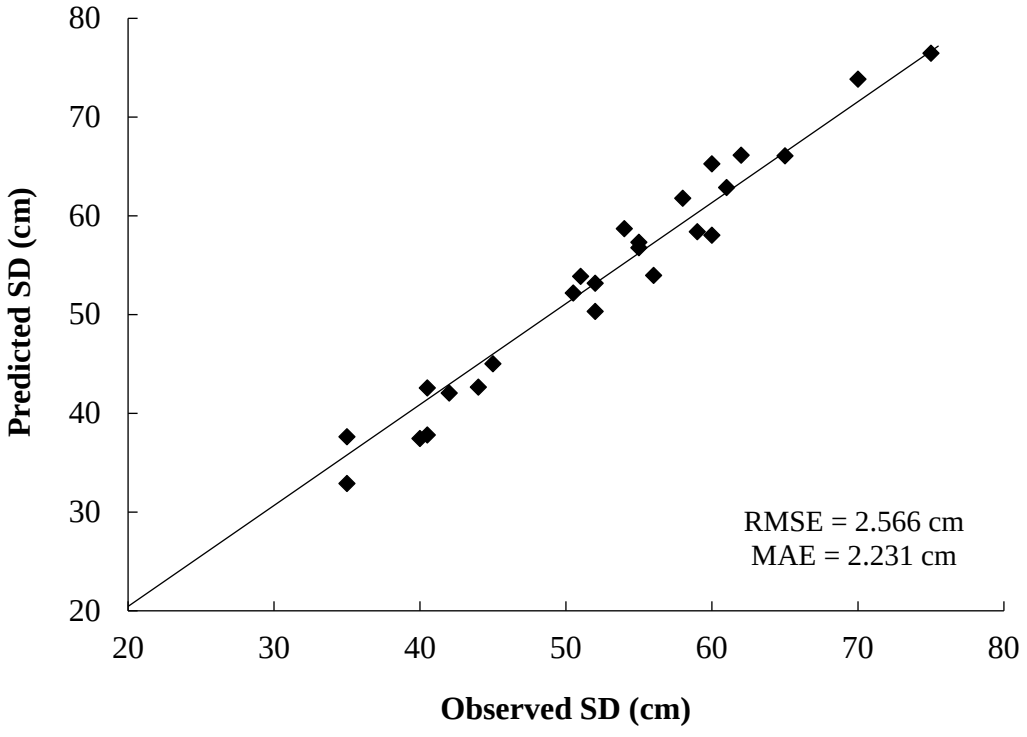

Figure 7.56: Prediction performance of SD-I ANFIS model testing against natural water body data (Trial 2)

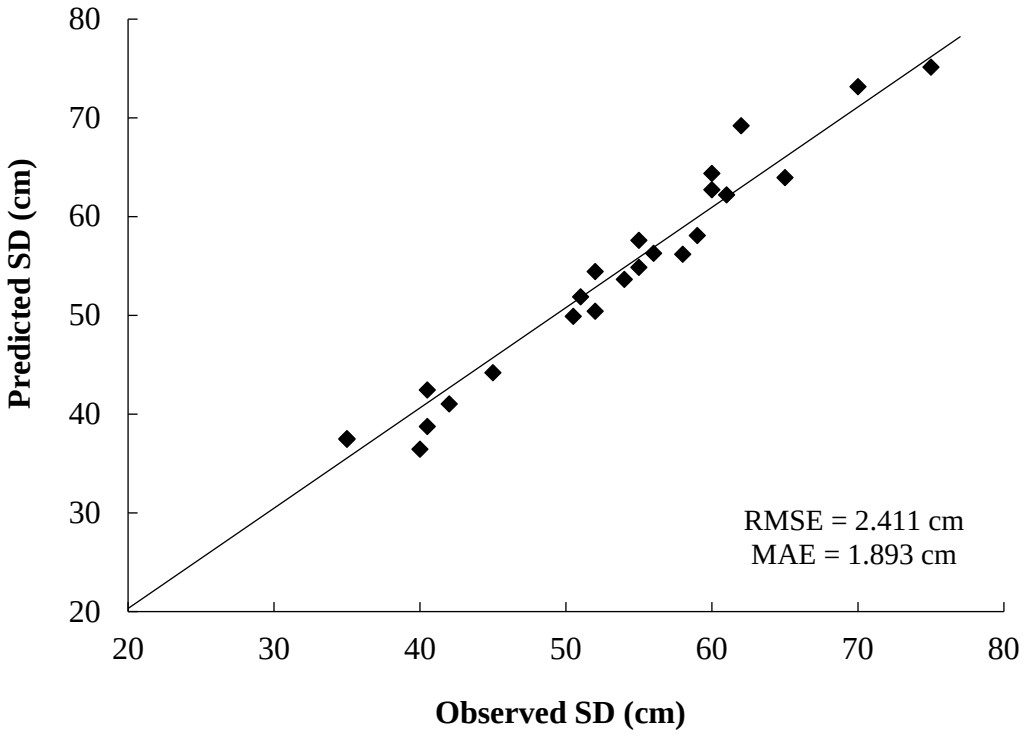

Figure 7.57: Prediction performance of SD-III ANFIS model testing against natural water body data (Trial 2)

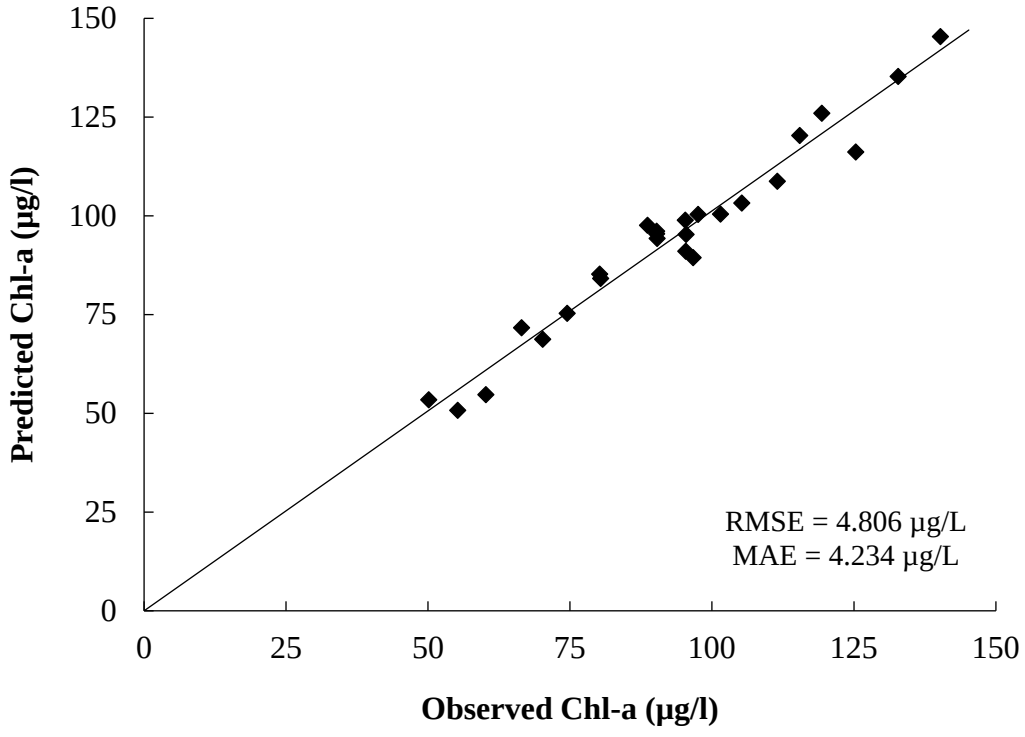

Figure 7.58: Prediction performance of Chl-a-I ANFIS model testing against natural water body data (Trial 2)

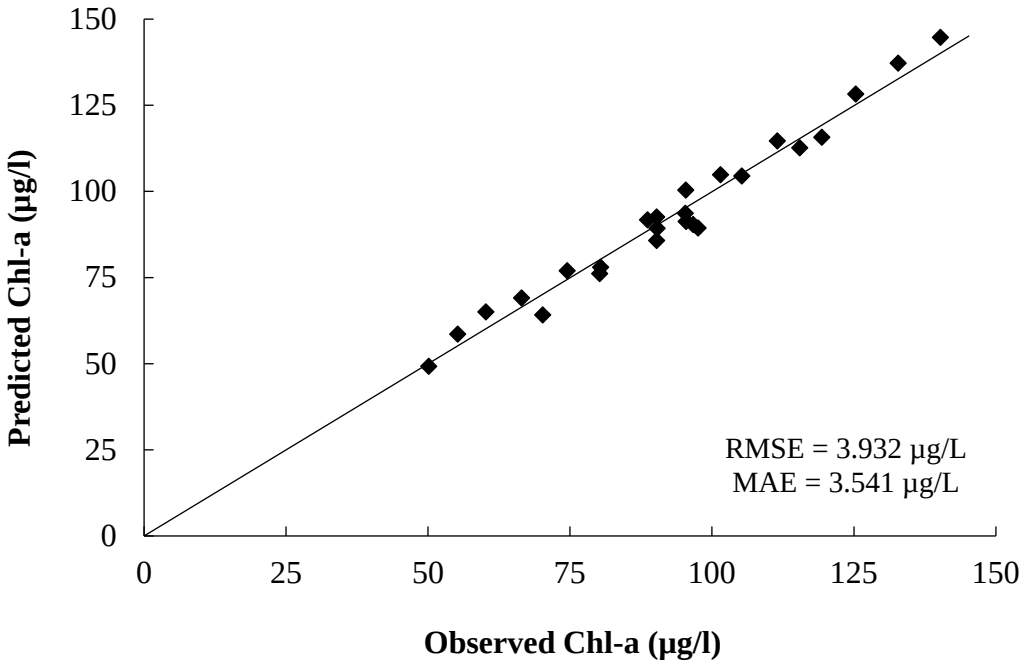

Figure 7.59: Prediction performance of Chl-a-III ANFIS model testing against natural water body data (Trial 2)

The statistical summary of model validation against natural water body data in terms of parameters R^2, E, RMSE and MAE are reported in Table 7.12. Based on these evaluation criteria it was observed that all the considered modelling approaches were successful in prediction of eutrophication indicators in natural water bodies with acceptable accuracy. The performance of ANN and GPR models were quite comparable, however, prediction capacity of ANFIS models were found superior to these two methods as reflected by higher R^2 and E values and lower RMSE and MAE values. The better prediction efficiency of the ANFIS models can be attributed to the fact that it has the added benefit of both neural network as well as fuzzy systems and can yield better prediction accuracy in moderate as well as extreme conditions similar to present validation dataset. Moreover, ANFIS models are quite effective in handling noisy, non-linear, and small dataset problems.

Comparing the models trained with both the prototype lakes, it was observed that, in general the models trained with Tank 3 gave better validation results than the models trained with Tank 1 dataset except for ANN based SD model. The validation results of SD-I and SD-III ANN models are quite comparable and R^2, E, and RMSE being slightly better for SD-I than SD-III. For all other cases R^2, E, RMSE, and MAE

values were superior in case of Tank 3 based models compared to its Tank 1 counterpart. For instance, the R^2, E, RMSE, and MAE values obtained for ANN DO-I model were 0.93, 0.92, 0.31 mg/L, and 0.26 mg/L whereas for ANN DO-III model the corresponding values were 0.96, 0.95, 0.25 mg/L, and 0.19 mg/L respectively. The overall better prediction accuracy of the models from Tank 3 against natural water body can be due to the fact that the dataset of training of these models corresponds to an artificial pond undergoing eutrophication in natural condition whereas in case of Tank 1 the whole process was under controlled environment. Therefore, the Tank 3 based models were able to predict the eutrophication indicators in natural water body with greater accuracy although higher efficiency has been observed for Tank 1 based models during model training.

7.2.6 Sensitivity Analysis

Sensitivity analysis was carried out on the developed models using the data perturbation method, by changing the input values by ±20% one at a time and keeping other parameters unaltered similar to 1st trial. Considering the results of model training and validation, as Tank 3 based models were found to be more accurate for prediction of DO, SD, and Chl-a in eutrophic water bodies, sensitivity analysis was conducted on the DO-III, SD-III, and Chl-a-III models under ANN, GPR, and ANFIS. To identify the most sensitive parameters, the percentage change in output owing to input perturbation was calculated and the results are shown in Figure 7.60. For DO prediction with ANN, GPR, and ANFIS, Chl-a was found to be the most sensitive parameter followed by temperature. Out of the nine input parameters for Chl-a prediction pH, DO, increase in nutrient concentration (TN and TP) and increase in temperature were most influential parameters. Compared to the DO and Chl-a models, input parameters in SD model were less sensitive and changes in input parameters were found to have lesser effect on SD prediction performance. The sensitive parameters indicate that a slight change in these values may cause a significant effect on model prediction. Overall, the sensitivity values of inputs were small for DO, SD, and Chl-a prediction indicating robustness of model performance. Similar trends of sensitivity values were observed for ANN, GPR, and ANFIS models for prediction of DO, SD, and Chl-a respectively. However, the effect of input perturbation on target prediction was smaller in case of ANFIS and GPR based models compared with the ANN models.

Table 7.12: Performance result of model validation with natural waterbody data (Trial 2)

Parameters	DO			SD			Chl-a		
	ANN	GPR	ANFIS	ANN	GPR	ANFIS	ANN	GPR	ANFIS
Tank 1									
R^2	0.936	0.995	0.998	0.941	0.995	0.998	0.991	0.991	0.998
E	0.928	0.850	0.949	0.852	0.870	0.934	0.860	0.830	0.955
RMSE	0.310 mg/L	0.451 mg/L	0.260 mg/L	3.962 cm	3.710 cm	2.566 cm	8.561 µg/L	9.440 µg/L	4.806 µg/L
MAE	0.267 mg/L	0.383 mg/L	0.212 mg/L	3.411 cm	3.180 cm	2.231 cm	7.323 µg/L	7.611 µg/L	4.234 µg/L
Tank 3									
R^2	0.961	0.996	0.999	0.938	0.996	0.998	0.995	0.992	0.998
E	0.951	0.892	0.966	0.834	0.904	0.942	0.902	0.840	0.970
RMSE	0.256 mg/L	0.382 mg/L	0.211 mg/L	4.194 cm	3.190 cm	2.411 cm	7.130 µg/L	8.962 µg/L	3.932 µg/L
MAE	0.192 mg/L	0.282 mg/L	0.177 mg/L	3.417 cm	2.810 cm	1.893 cm	5.662 µg/L	7.061 µg/L	3.541 µg/L

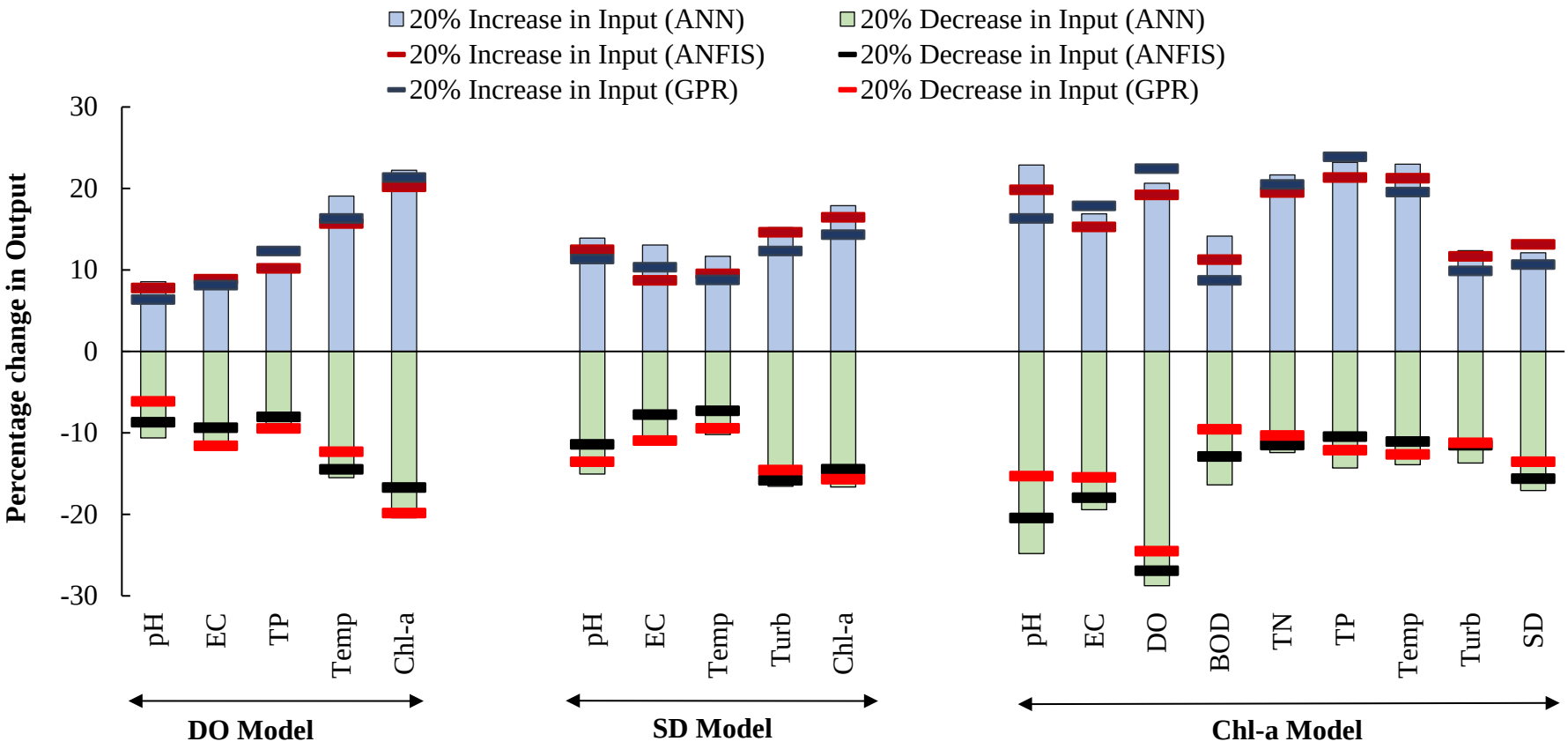

Figure 7.60: Result of sensitivity analysis of developed models (Trial 2)

7.2.7 Major observations on model performance

From the results and discussion of the model performance from the 2nd trial, the following major conclusions can be drawn:

- For DO, SD and Chl-a prediction 5, 5 and 9 input parameters were found optimum respectively in case of ANN models which were also used for GPR and ANFIS model development.
- Using a trial-and-error approach 8, 7 and 18 numbers of neurons in the hidden layer was found to have better performance in DO, SD, and Chl-a prediction by ANN network.
- Model training result revealed that ANN, GPR, and ANFIS models were very accurate in prediction of eutrophication indicators DO, SD, and Chl-a. Models trained with Tank 1 data produced better goodness-of-fit and lesser error compared to Tank 3 based models.
- Model validation with natural water body data revealed suitability of modelling approach for lake eutrophication prediction. Models developed with artificial pond (Tank 3) data were found to give better prediction accuracy in natural water bodies compared to models trained from artificial concrete lake (Tank I) data.
- Most sensitive input parameters for DO prediction was found to be chlorophyll-a followed by temperature. In case of chlorophyll-a prediction pH, increase in nutrient concentration, DO, and increase in temperature were found to have major significance. For SD model, the inputs were found to have lesser sensitivity.
- Overall, the prediction accuracy of the ANFIS models were found to be superior compared to GPR and ANN models both during model training and validation stage.

7.3 SUMMARY ON MODEL PERFORMANCE

In this study, a unique approach was employed to model eutrophication indicators using data from artificially replicated lake eutrophication scenarios. Data-driven modelling tools ANN, SVR, and GPR were employed initially to predict

eutrophication indicators DO and SD. It was found that compared to SVR models, ANN (both MLP and TDNN) and GPR models were able to predict the desired indicators in eutrophic water bodies with acceptable degree of accuracy. From the findings of 1st modelling trial, more robust eutrophication models were trained during 2nd trial for indicators DO, SD, and Chl-a. During the 2nd trial ANFIS methodology was also introduced to train the eutrophication models in addition to ANN and GPR. It was found that the performance of the models during 2nd trial were superior in terms of statistical parameters of evaluation compared to initial investigation. All models trained under ANN, GPR, and ANFIS architecture were able to produce satisfactory results during training and validation stage. ANFIS performance was found to be more robust in terms of prediction accuracy as well as input parameter sensitivity. The better performance of ANFIS models over other machine learning methods is similar to the results reported by earlier works of Pham et. al. [138], Yaseen et. al. [201], Yarar et. al. [200], Sanikhani et. al. [157], Ly et. al. [111], Kovačević et. al. [90], and Elkiran et. al. [44].

From the results of model validation, it was evident that models developed through dataset of Tank 1 and Tank 3 were able to predict the eutrophication indicators in natural water bodies with reasonable accuracy. However, Tank 3 based model's performance was found slightly better than Tank 1 based models. But during experimental investigation it was observed that the time required for replication of lake eutrophication in Tank 3 (artificial pond) was higher compared to Tank 1 (concrete tank). So, with lesser time frame lake modelling may be done through dataset of artificial lakes like Tank 1 under controlled environment with very reasonable prediction accuracy as a rapid management measure. As discussed in the preceding sections that for the concrete bedded prototype lakes, eutrophication process was replicated in around 230 to 270 days' time period under controlled environment whereas under natural condition like in case of Tank 3, it took only around 350 days period for complete change of trophic status of the lake. So, it can be inferred that when the external nutrient loading to the water body is major cause of concern for the water body and factors such as nutrient loading from sediments, sediment mixing etc. are not quite predominant then experimental methodology of prototype lakes under controlled environment can give prompt reliable management outcome.

The adopted modelling methodology based on laboratory investigated data from artificial lakes was successful to predict eutrophication indicators in natural water bodies. The capacity of the created models to be employed as a eutrophication prediction tool in small to large waterbodies under moderate to extreme environmental circumstances was implied by successful model validation with a wide validation dataset. These eutrophication models are extremely valuable tools for predicting the extent of eutrophication in the waterbodies under investigation. Unlike physical process-based models, which require extensive input conditions and are very complicated in nature, the presented data driven modelling technique allows policymakers and lake managers to implement lake management policies quickly and effectively. With limited input parameters, future forecasting of target variables like DO, SD and Chl-a values can be done. The effect of input parameters on target prediction and their interdependency can also be assessed. For instance, input parameters considered for DO prediction in the presented work are pH, EC, TP Chl-a and temperature. So, the effect on DO value of the lake if one or few input parameters increased or decreased can be predicted with the presented models. Moreover, the major parameters responsible for eutrophication in the considered lake can be identified and their control measures can be formulated accordingly.

Case-specific data driven eutrophication models have been investigated previously which are based on large datasets collected over a five to twenty-year period. Current study highlights the usefulness of prevailing data driven neural network modelling approach for lake eutrophication management under circumstances where prolonged water quality data is not available. In developing countries like in India, where surface water quality of lakes is deteriorating at an alarming rate due to cultural eutrophication. This methodology based on artificially simulated lake systems provides a rapid and effective measure to formulate lake restoration policies. The presented models were successful in predicting eutrophication indicators in natural water bodies in Assam, India. Lake eutrophication is a complex process dependent on several factors like morphometry, nutrient loading, thermal stratification, and climatic condition etc. The presented work utilizes periodic monitoring data for development of eutrophication models and so can accommodate the frequent variations in water quality. Considering these aspects, the presented models are more useful for eutrophication management in shallow waterbodies with tropical monsoon climatic conditions and having higher

humidity levels similar to Assam, India. Moreover, the major source of water pollution in the natural water bodies prone to eutrophication in Assam is due to domestic wastes. So, the applicability of the developed models is limited to use in the natural water bodies in Assam, India where major source of contamination is domestic wastes with high nutrient concentration and low BOD.