NOAA Technical Memorandum GLERL-177
DOI: 10.25923/d9mk-jn08

Development of Multiple-variable, Non-linear, Regression Models Linking Lake Erie Water Quality to Environmental and Climate Teleconnection Forcings, 1970s–2010s

Jia Wang[1], Matthew Trumper[2,3], Hongyan Zhang[4], James Kessler[1], Mark Rowe[1], Philip Chu[1], Ting-Yi Yang[2,5], and Brent Lofgren[1]

[1] NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan
[2] Cooperative Institute for Great Lakes Research, University of Michigan
[3] University of Minnesota Twin Cities
[4] Eureka Aquatic Research, Ann Arbor, Michigan
[5] Geodetic Science, School of Earth Science, the Ohio State University, Columbus, Ohio

NOAA Great Lakes Environmental Research Laboratory
4840 S. State Road, Ann Arbor, Michigan

Tuesday, February 15, 2022

UNITED STATES
DEPARTMENT OF COMMERCE

Gina M. Riamondo, Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION

Dr. Richard W. Spinrad, Administrator

NOTICE

This publication is available as a PDF file and can be downloaded from GLERL's web site: www.glerl.noaa.gov or by emailing GLERL Information Services at oar.pubs.glerl@noaa.gov.

TABLE of CONTENTS

1.0 ABSTRACT **1**

2.0 INTRODUCTION **1**

3.0 METHODS **3**

3.1 Data 3

3.2 Statistical analysis, description of tables and figures 4

4.0 RESULTS **5**

4.1 Diatom Models 5

4.2 Standardized Teleconnections Model 10

4.3 Hypolimnetic Oxygen Demand (HOD) Models 15

4.4 Maximum Hypoxic Area Models 23

4.5 Mean Hypoxic Area Models 31

4.6 Mean Dissolved Oxygen (DO) Models 52

4.7 Median Dissolved Oxygen (DO) Models 65

4.8 Total Phosphate Load Models 73

4.9 Hypoxic Factor Models 77

5.0 CORRELATIONS AND TIME SERIES OF TELECONNECTIONS AND WATER LEVEL AND PRECIPITATION **85**

5.1 Erie Annual Average Water Level 85

5.2 Great Lakes Basin Annual Precipitation 87

6.0 SUMMARY **89**

7.0 ACKNOWLEDGMENTS **90**

8.0 REFERENCES **90**

LIST of FIGURES

Figure 2-1. Distribution of air temperature stations around the Great Lakes. 2

Figure 2-2. Hypothesized relationships between teleconnection patterns (ENSO, PDO, NAO, AMO), regional climate (sea level pressure, surface air temperature, precipitation), lake-level processes in Lake Erie (stratification, hypolimnion thickness, water level, nutrient loading), and hypoxia (hypolimnetic oxygen depletion, hypoxic extent). Green arrows indicate a hypothesized positive relationship between two variables, while red arrows indicate a hypothesized negative relationship between two variables. These hypothesized relationships are reflected in the models shown in this report. 3

Figure 4-1. Regression coefficient plot (diatom model). 7

Figure 4-2. Linear correlations between diatom (spring chlorophyll a), biological parameters, and physical forcings. 8

Figure 4-3. Added variable plot for diatom model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the index of the point in the dataset. For example, the number 41 represents the 41st data point out of the total data points. 9

Figure 4-4. Time series plot of modeled vs. observed values (diatom model). 9

Figure 4-5. Regression coefficient plot (diatom model). 12

Figure 4-6. Linear correlations between diatom (spring chlorophyll a), biological parameters, and physical forcings. 13

Figure 4-7. Added variable plot for diatom model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series....... 14

Figure 4-8. Time series plot of modeled vs. observed values (diatom model). 14

Figure 4-9. Regression coefficient plot (HOD model). 16

Figure 4-10. Linear correlations between HOD, biological parameters, and physical forcings... 17

Figure 4-11. Added variable plot for HOD model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series....... 18

Figure 4-12. Time series plot of modeled vs. observed values (HOD model). 18

Figure 4-13. Regression coefficient plot (HOD model). 20

Figure 4-14. Linear correlations between HOD, biological parameters, and physical forcings... 21

Figure 4-15. Added variable plot for HOD model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series....... 22

Figure 4-16. Time series plot of modeled vs. observed values (HOD model). 22

Figure 4-17. Regression coefficient plot (maximum hypoxic extent model. 24

Figure 4-18. Linear correlations between maximum hypoxic extent, biological parameters, and physical forcings. 25

Figure 4-19. Added variable plot for maximum hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data in the data time series. 26

Figure 4-20. Time series plot of modeled vs. observed values (maximum hypoxic area model). 26

Figure 4-21. Regression coefficient plot (maximum hypoxic extent model). 28

Figure 4-22. Linear correlations between maximum hypoxic extent, biological parameters, and physical forcings. 29

Figure 4-23. Added variable plot for maximum hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.. 30

Figure 4-24. Time series plot of modeled vs. observed values (maximum hypoxic area model.. 30

Figure 4-25. Regression coefficient plot (mean hypoxic extent model)..................................... 33

Figure 4-26. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings. 34

Figure 4-27. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 35

Figure 4-28. Time series plot of modeled vs. observed values (mean hypoxic area model)........ 35

Figure 4-29. Regression coefficient plot (mean hypoxic extent model)..................................... 37

Figure 4-30. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings. 38

Figure 4-31. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 39

Figure 4-32. Time series plot of modeled vs. observed values (mean hypoxic area model)........ 39

Figure 4-33. Time series plot showing the contribution of each variable to annual hypoxia prediction. The model used here is a subset of the above model, using only the ratio of PMar / Apr precipitation and March-May air temperature over the Lake Erie basin. This plot allows the user to visualize how hypoxia is related to each variable in different years........ 40

Figure 4-34. Regression coefficient plot (mean hypoxic extent model)........ 41

Figure 4-35. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings. 42

Figure 4-36. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 43

Figure 4-37. Time series plot of modeled vs. observed values (mean hypoxic area model)........ 43

Figure 4-38. Regression coefficient plot (mean hypoxic extent model)........ 45

Figure 4-39. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings. 46

Figure 4-40. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 47

Figure 4-41. Time series plot of modeled vs. observed values (mean hypoxic area model)........ 47

Figure 4-42. Regression coefficient plot (mean hypoxic extent model)........ 49

Figure 4-43. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings. 50

Figure 4-44. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 51

Figure 4-45. Time series plot of modeled vs. observed values (mean hypoxic area model)........ 51

Figure 4-46. Time series plot showing the contribution of each variable to annual hypoxia prediction. The model used here is very similar to the above model, but instead using the ratio of 7yrMarchPrecip / AprMayPrecip rather than a difference. This change was made because it was very difficult to make this style of plot for the above difference term model, simply because the intercept and terms did not add together nicely for visualization. 52

Figure 4-47. Regression coefficient plot (mean dissolved oxygen model). 54

Figure 4-48. Linear correlations between mean dissolved oxygen, biological parameters, and physical forcings. 55

Figure 4-49. Added variable plot for mean dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 56

Figure 4-50. Time series plot of modeled vs. observed values (mean dissolved oxygen model). 56

Figure 4-51. Regression coefficient plot (mean dissolved oxygen model). 58

Figure 4-52. Linear correlations between mean dissolved oxygen, biological parameters, and physical forcings. 59

Figure 4-53. Added variable plot for mean dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 60

Figure 4-54. Time series plot of modeled vs. observed values (mean dissolved oxygen model). 60

Figure 4-55. Regression coefficient plot (mean dissolved oxygen model). 62

Figure 4-56. Linear correlations between mean dissolved oxygen, biological parameters, and physical forcings. 63

Figure 4-57. Added variable plot for mean dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 64

Figure 4-58. Time series plot of modeled vs. observed values (mean dissolved oxygen model). 64

Figure 4-59. Regression coefficient plot (median dissolved oxygen model). 66

Figure 4-60. Linear correlations between median dissolved oxygen, biological parameters, and physical forcings. 67

Figure 4-61. Added variable plot for median dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 68

Figure 4-62. Time series plot of modeled vs. observed values (median dissolved oxygen model). 68

Figure 4-63. Regression coefficient plot (median dissolved oxygen model). 70

Figure 4-64. Linear correlations between median dissolved oxygen, biological parameters, and physical forcings. 71

Figure 4-65. Added variable plot for median dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 72

Figure 4-66. Time series plot of modeled vs. observed values (median dissolved oxygen model). 72

Figure 4-67. Regression coefficient plot (total phosphate model). 74

Figure 4-68. Linear correlations between total phosphate load, biological parameters, and physical forcings. 75

Figure 4-69. Added variable plot for total phosphate model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 76

Figure 4-70. Time series plot of modeled vs. observed values (total phosphate load model). 76

Figure 4-71. Regression coefficient plot (hypoxic factor model). 78

Figure 4-72. Linear correlations between hypoxic factor, biological parameters, and physical forcings. 79

Figure 4-73. Added variable plot for hypoxic factor model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 80

Figure 4-74. Time series plot of modeled vs. observed values (hypoxic factor model). 80

Figure 4-75. Regression coefficient plot (hypoxic factor model). 82

Figure 4-76. Linear correlations between hypoxic factor, biological parameters, and physical forcings. 83

Figure 4-77. Added variable plot for hypoxic factor model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series. 84

Figure 4-78. Time series plot of modeled vs. observed values (hypoxic factor model). 84

Figure 5-1. Erie annual average water level plotted with AMO, PDO, NAO, and ENSO. 86

Figure 5-2. Great Lakes basin precipitation plotted with AMO, PDO, NAO, and ENSO. 88

LIST of TABLES

Table 4.1-1. Correlations and p-values of Diatom with teleconnection patterns. Significant correlations shown in boldface. 5

Table 4.1-2. Regression output for diatom model. 6

Table 4.1-3. Table summarizing the best subsets procedure for the diatom model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 6

Table 4.2-1. Table summarizing the best subsets procedure for the diatom model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 10

Table 4.2-2. Regression output for diatom model. 11

Table 4.2-3. Table summarizing the best subsets procedure for the diatom model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 11

Table 4.3-1. Correlations and p-values of HOD with teleconnection patterns. Significant correlations shown in boldface. 15

Table 4.3-2. Regression output for HOD model. 15

Table 4.3-3. Table summarizing the best subsets procedure for the HOD model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 16

Table 4.3-4 .Regression output for HOD model. 19

Table 4.3-5. Table summarizing the best subsets procedure for the HOD model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 19

Table 4.4-1. Correlations and p-values of maximum hypoxic area with teleconnection patterns. Significant correlations shown in boldface. 23

Table 4.4-2. Regression output for maximum hypoxic area model. 23

Table 4.4-3. Table summarizing the best subsets procedure for the maximum area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 24

Table 4.4-4. Regression output for maximum hypoxic area model. 27

Table 4.4-5. Table summarizing the best subsets procedure for the maximum area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 27

Table 4.5-1. Correlations and p-values of mean hypoxic area with teleconnection patterns. Significant correlations shown in boldface. 31

Table 4.5-2. Regression output for mean hypoxic area model. 32

Table 4.5-3. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R2, R2adj, R2-predicted, and Bayesian information criterion (BIC). 32

Table 4.5-4. Regression output for mean hypoxic area model. 36

Table 4.5-5. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 36

Table 4.5-6. Regression output for mean hypoxic area model. 40

Table 4.5-7. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 41

Table 4.5-8. Regression output for mean hypoxic area model. 44

Table 4.5-9. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 44

Table 4.5-10. Regression output for mean hypoxia area model. 48

Table 4.5-11. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 48

Table 4.6-1. Correlations and p-values of mean dissolved oxygen with teleconnection patterns. Significant correlations shown in boldface. 52

Table 4.6-2. Regression output for mean dissolved oxygen model. 53

Table 4.6-3. Table summarizing the best subsets procedure for the mean dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 53

Table 4.6-4. Regression output for mean dissolved oxygen model. UWindJun and VwindJun represent the u- and v-component of June wind over the Erie basin. This data came from the North American Regional Reanalysis (NARR) and was used because there was no missing data, unlike the wind data from Zhang et al. (2018).. 57

Table 4.6-5. Table summarizing the best subsets procedure for the mean dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 57

Table 4.6-6. Regression output for mean dissolved oxygen model. NARR_TotalPrecipApr is total April precipitation over the Lake Erie basin and comes from the North American Regional Reanalysis (NARR). 61

Table 4.6-7. Table summarizing the best subsets procedure for the mean dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 61

Table 4.7-1. Correlations and p-values of median dissolved oxygen with teleconnection patterns. Significant correlations shown in boldface.. 65

Table 4.7-2. Regression output for median dissolved oxygen model... 65

Table 4.7-3. Table summarizing the best subsets procedure for the median dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 66

Table 4.7-4. Regression output for median dissolved oxygen model... 69

Table 4.7-5. Table summarizing the best subsets procedure for the median dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 69

Table 4.8-1. Correlations and p-values of median dissolved oxygen with teleconnection patterns. Significant correlations shown in boldface.. 73

Table 4.8-2. Regression output for total phosphate model. 73

Table 4.8-3. Table summarizing the best subsets procedure for the total phosphate model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 74

Table 4.9-1. Correlations and p-values of hypoxic factor with teleconnection patterns. Significant correlations shown in boldface. 77

Table 4.9-2. Regression output for hypoxic factor model. 77

Table 4.9-3. Table summarizing the best subsets procedure for the hypoxic factor model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 78

Table 4.9-4. Regression output for hypoxic factor model. 81

Table 4.9-5. Table summarizing the best subsets procedure for the hypoxic factor model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC). 81

Table 5.1-1. Correlations and p-values of Lake Erie water level with teleconnection patterns. Significant correlations shown in boldface. 85

Table 5.2-1. Correlations and p-values of basin annual precipitation with teleconnection patterns. Significant correlations shown in boldface. 87

Development of Multiple-variable, Non-linear, Regression Models Linking Lake Erie Water Quality to Environmental and Climate Teleconnection Forcings, 1970s–2010s

Jia Wang[1], Matthew Trumper[2,3], Hongyan Zhang[4], James Kessler[1], Mark Rowe[1], Philip Chu[1], Ting-Yi Yang[2,5] and Brent Lofgren[1]

[1] NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan
[2] Cooperative Institute for Great Lakes Research, University of Michigan
[3] University of Minnesota Twin City
[4] Eureka Aquatic Research, Ann Arbor, Michigan
[5] Geodetic Science, School of Earth Science, the Ohio State University, Columbus, Ohio

1.0 ABSTRACT

This report uses relationships between biological parameters and environmental forcings to develop statistical prediction models for hypoxia and water quality in Lake Erie. This study is built upon previous studies that showed atmospheric teleconnection patterns such as the El Nino Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), and Atlantic Multidecadal Oscillation (AMO) are associated with anomalous ice cover on the Great Lakes (Bai et al. 2012; Wang et al. 2018) and with water quality in Lake Erie (Zhang et al. 2018). Specifically, we extended the study reported in NOAA Technical Memorandum GLERL-173, which produced relationships between variables (Zhang et al 2018), to establish multiple-variable regression models to quantitatively hindcast and predict year-to-year variations of hypoxia (Zhou et al. 2015; Watson et al. 2016) and other biological parameters using physical/climate variables as predictors.

2.0 INTRODUCTION

The Laurentian Great Lakes, located in the mid-latitude of eastern North America (Fig. 2-1), contain about 95% of the United States' and 20% of the world's fresh surface water supply. Nearly one eighth of the population of the United States and one third of the population of Canada live within their drainage basins. The Great Lakes can be considered a mini climate system—though small compared to the global climate system or Arctic regional climate system—since all five important climate system components are included: regional atmosphere, hydrosphere (hydrodynamics), cryosphere (lake ice), biosphere (aquatic ecosystem and terrestrial ecosystem), and land process (hydrology). We use the Great Lakes resources to meet our social and economic needs, therefore, human dimension is another important component that affects the Great Lakes climate system. In this mini-climate system, there are strong interactions and associations among these components. Because of this concentration of population (human dimension), the year-to-year variability of the ice cover that forms on the Great Lakes each winter affects the regional economy (Niim, 1982). It also affects the lake's abiotic environment and ecosystems (Vanderploeg et al. 1992), and therefore influences summer hypoxia, lake effect snow, water level variability, and the overall hydrologic cycle of the region (Assel et al. 2004).

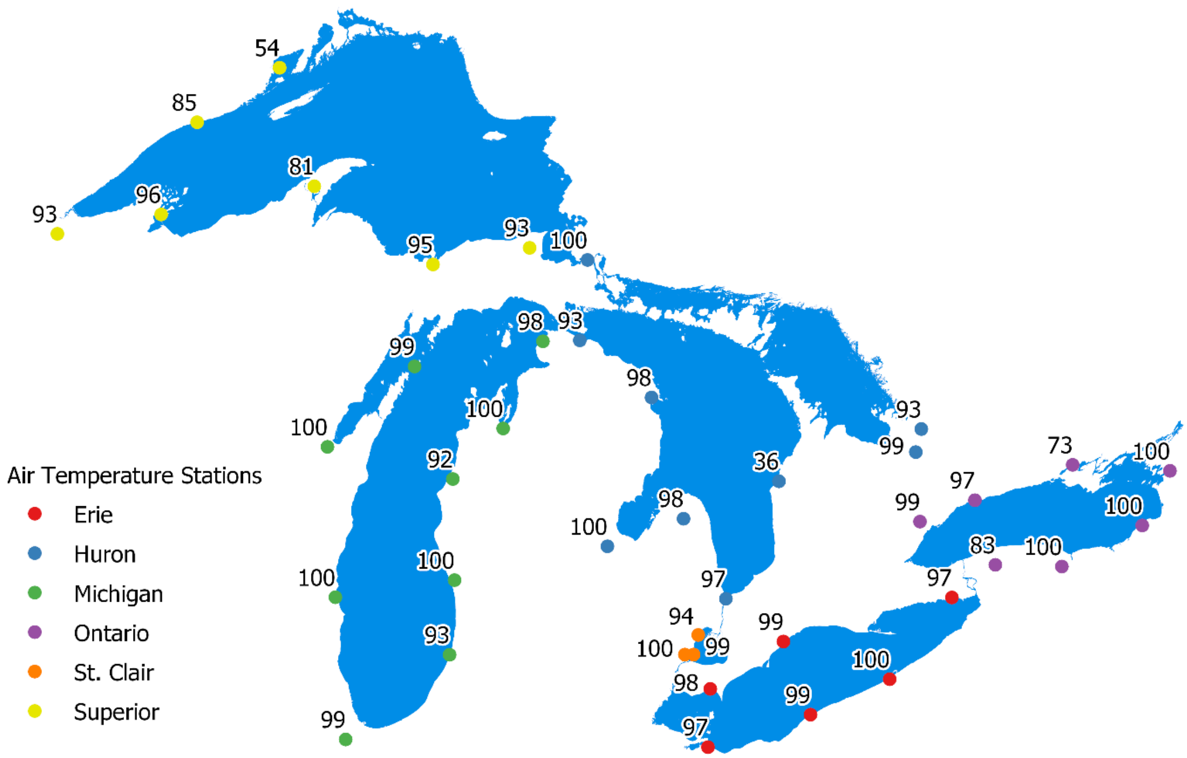

Figure 2-1. Distribution of air temperature stations around the Great Lakes.

Studies show that atmospheric teleconnection patterns such as the North Atlantic Oscillation (NAO), El Nino and Southern Oscillation (ENSO) (Bai et al. 2012), Pacific Decadal Oscillation (PDO), and Atlantic Multidecadal Oscillation (AMO) are associated with anomalous ice cover on the Great Lakes (Wang et al. 2018). Mishra et al. (2011) also shows that lake ice phenology of small lakes around the Great Lakes region is associated with these major climate teleconnection patterns.

Zhang et al. (2018) conducted a comprehensive study that the fundamental relationship between individual water quality variable and single teleconnection pattern has been revealed. Nevertheless, they also found that water quality is influenced by multiple factors, since no single teleconnection pattern (Wang et al. 2018) or environmental forcing can determine its faith. Previous studies show that water quality is related to temperature, wind mixing, runoff, and nutrient loading (Rowe et al. 2019). In this study, we further add the forcing of teleconnection patterns into the water quality equation. We hypothesize that the teleconnection patterns have direct and indirect impacts on the Great Lakes water quality system, as described by the diagram below (Fig. 2-2):

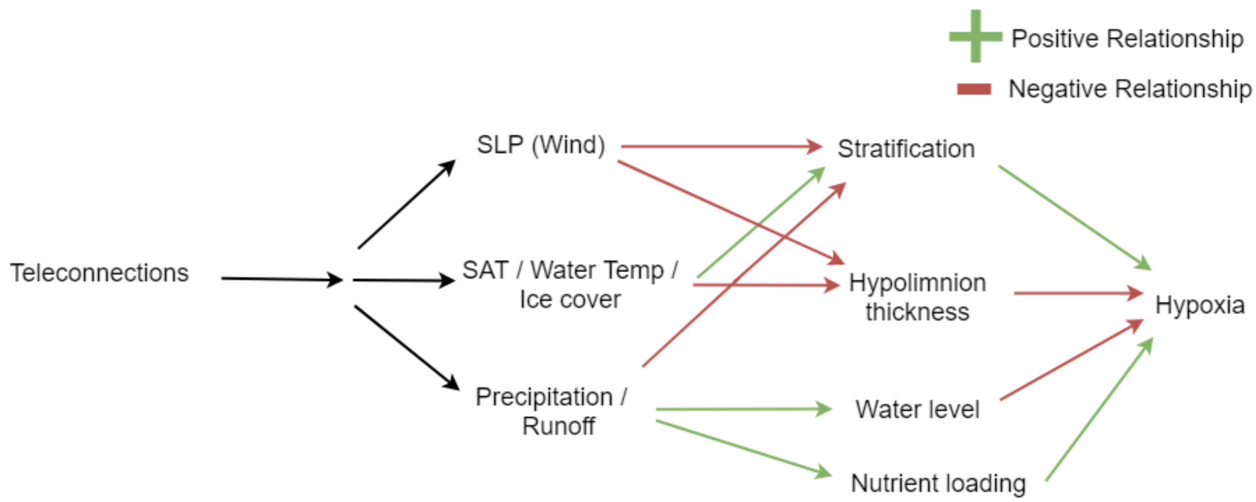

Figure 2-2. Hypothesized relationships between teleconnection patterns (ENSO, PDO, NAO, AMO), regional climate (sea level pressure, surface air temperature, precipitation), lake-level processes in Lake Erie (stratification, hypolimnion thickness, water level, nutrient loading), and hypoxia (hypolimnetic oxygen depletion, hypoxic extent). Green arrows indicate a hypothesized positive relationship between two variables, while red arrows indicate a hypothesized negative relationship between two variables. These hypothesized relationships are reflected in the models shown in this report.

3.0 METHODS

3.1 DATA

Biological parameters, teleconnection indices and other physical forcings used in this report come from NOAA Technical Memorandum 173 (Zhang et al. 2018). In that report, regression models are developed for diatoms (spring chlorophyll a), hypolimnetic oxygen demand (HOD), maximum and mean hypoxic area, mean and median dissolved oxygen, and total phosphate load. Additional monthly climate and hydrological data were downloaded from the Great Lakes Hydrometeorological Database, including (but not limited to) precipitation and temperature over the Lake Erie basin, and Lake Erie water levels. Missing data was a consistent issue in creating multiple regression models, particularly for the last five years when the NOAA Great Lakes Environmental Research Laboratory (NOAA GLERL) last updated their research dataset. In some cases, missing data was filled by using the North American Regional Reanalysis (NARR) dataset subset over the Lake Erie basin. The data were cleaned and formatted, and included relevant climate variables such as air temperature, total precipitation, u- and v-component of wind, soil moisture, evaporation, and runoff. However, the NARR dataset only extends to 1979 so data from NOAA GLERL was preferred for longer hindcast estimates of hypoxia. Thus, most regression models used data from either Zhang et al. (2018) or the NOAA GLERL research data dashboard.

3.2 STATISTICAL ANALYSIS, DESCRIPTION OF TABLES AND FIGURES

To link a water quality variable to multiple forcings, we follow the approach of Wang et al. (2018) to develop multiple-variable, nonlinear regression models to quantitatively hindcast a water quality variable, say hypoxia, using both physically and statistically significant, selective environmental/physical and climate predictors.

The methods used in this research include correlation analysis, scatterplot visualization, and multivariable regression. Model selection and model fit was determined using the commonly used R^2 statistic, as well as R^2-adjusted, which penalizes models that have more predictor variables. In addition, the R^2-predicted statistic was used, whereby one data point in the response variable is withheld from the model, the regression model is calculated using the remaining data, and then that model is used to predict the withheld observation. This process is known as leave-one-out cross validation and is particularly suitable for prediction models and for small sample sizes because it is statistically efficient. In some cases, R^2-predicted decreases sharply when unrelated predictor variables are added to the model and R^2-predicted can go below zero if many unrelated predictor variables are added. Lastly, the commonly used Bayesian Information Criterion (BIC) was calculated for model selection; the model that minimizes BIC is preferred.

Three tables are shown for each prediction model. The first table shows correlations between the response variable and both linear and quadratic teleconnections. The second table includes output for the full regression model, including parameter coefficients (estimates), coefficient 95% confidence intervals (CI), t-statistics, and *p*-values. *T*-statistics are calculated by dividing the coefficients by their standard errors (not shown). The number of observations, R^2, and R^2-adjusted are also shown in the second table. The third table summarizes a method of model selection called best subsets variable selection. This method selects the subset of predictors that do the best at meeting some well-defined objective criteria, such as having the largest R^2 value or the smallest BIC. The third table shows how removing specific predictor variables changes the R^2, R^2-adjusted, R^2-predicted and BIC of each model. The "N" column represents the number of predictor variables used in the model. Notice that R^2 always increases as more predictors are added to the model; however, the other model fit statistics may decrease. This third table may be used to choose models with a reduced number of terms that still retain high explanatory power.

Four figures are shown for each full model in this report. The first figure shows regression coefficients (estimates) as points with 95% confidence interval lines. Red dots indicate negative regression coefficients, while blue dots indicate positive coefficients. This type of plot was included because it gives a useful visualization of the sign (positive or negative) and uncertainty (confidence interval) of coefficients for each predictor in the full model. The second figure is a scatterplot matrix for all the variables included in each model. The correlation coefficient between the two variables are shown in each scatterplot. This plot was included because it is always necessary to look at your data when constructing multiple regression models; allowing the users to identify linear or nonlinear patterns, potential outlier points, and multi-collinearity between predictor variables. The third figure is an added variable plot, which is used to evaluate the importance of each predictor in a model, after accounting for the effects that all the other predictors had in the model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of x to the model already containing the other predictors. The fourth figure is a time series plot of observed vs. modeled data. Red points indicate the predicted values based on the regression model, while the black points are the observed data. Lastly, the regression equation is written out for each full model contained in the report. In various figures, quadratic terms are sometimes denoted with I(^2) syntax, and product (multiply) is denoted by ":"; these are the syntax that *R* software requires for quadratic terms to be specified in linear models. Individual months are denoted with three letter abbreviations (i.e., Jan, Feb, Mar).

4.0 RESULTS

4.1 DIATOM MODELS

Table 4.1-1. Correlations and p-values of Diatom with teleconnection patterns. Significant correlations shown in boldface.

Index	r	p value	Significance (%)
ENSO	0.316	0.089	91.1
$ENSO^2$	0.043	0.823	17.7
NAO	-0.003	0.986	1.4
NAO^2	-0.085	0.655	34.5
AMO	-0.308	0.097	90.3
AMO^2	0.390	0.033	96.7
PDO	0.261	0.164	83.6
PDO^2	0.254	0.176	82.4

Winter Teleconnections Model

$$\text{Diatom} = 17.17 + 0.87\text{ENSO} + -0.92\text{PDO} + 0.46\text{PDO}^2 + 0.49\text{AMO} + 32.93\text{AMO}^2 + 0.09\text{ENSO:PDO} + 1.33\text{AMO:NAO} + -0.79\text{WaterTempSpring} + -0.01\text{AMIC} + -1.51\text{Wind} + e \quad (4.1\text{-}1)$$

Table 4.1-2. Regression output for diatom model.

	Diatom Apr			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	17.17	-14.88 – 49.22	1.05	0.310
ENSO	0.87	-0.28 – 2.02	1.48	0.160
PDO	-0.92	-2.44 – 0.60	-1.19	0.254
PDO^2	0.46	-0.69 – 1.61	0.79	0.441
AMO	0.49	-8.80 – 9.79	0.10	0.918
AMO^2	32.93	-15.18 – 81.05	1.34	0.200
WaterTempSpring	-0.79	-1.81 – 0.22	-1.53	0.147
AMIC	-0.01	-0.06 – 0.04	-0.33	0.743
Wind	-1.51	-5.64 – 2.62	-0.72	0.485
ENSO:PDO	0.09	-1.15 – 1.33	0.15	0.883
AMO:NAO	1.33	-10.62 – 13.28	0.22	0.831
Observations	26			
R^2 / adjusted R^2	0.416 / 0.027			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.1-3. Table summarizing the best subsets procedure for the diatom model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	**Predictors**	**Rsquare**	**AdjRsq**	**PredRsq**	**BIC**
1	AMO^2	0.152	0.122	-0.053	43.951
2	PDO^2 AMO^2	0.252	0.196	0.026	40.396
3	PDO^2 AMO^2 WaterTempSpring	0.311	0.231	0.027	38.134
4	ENSO PDO^2 AMO^2 WaterTempSpring	0.360	0.258	0.008	36.113
5	ENSO PDO PDO^2 AMO^2 WaterTempSpring	0.395	0.269	0.055	34.680
6	ENSO PDO PDO^2 AMO^2 WaterTempSpring Wind	0.404	0.216	-0.218	31.633
7	ENSO PDO PDO^2 AMO^2 WaterTempSpring AMIC Wind	0.411	0.181	-0.255	31.524
8	ENSO PDO PDO^2 AMO^2 WaterTempSpring AMIC.... Wind AMO:NAO	0.414	0.139	-0.569	31.536
9	ENSO PDO PDO^2 AMO^2 WaterTempSpring AMIC.... Wind ENSO:PDO AMO:NAO	0.416	0.087	-0.911	31.634
10	ENSO PDO PDO^2 AMO AMO^2 WaterTempSpring AMIC.... Wind ENSO:PDO AMO:NAO	0.416	0.027	-1.219	31.781

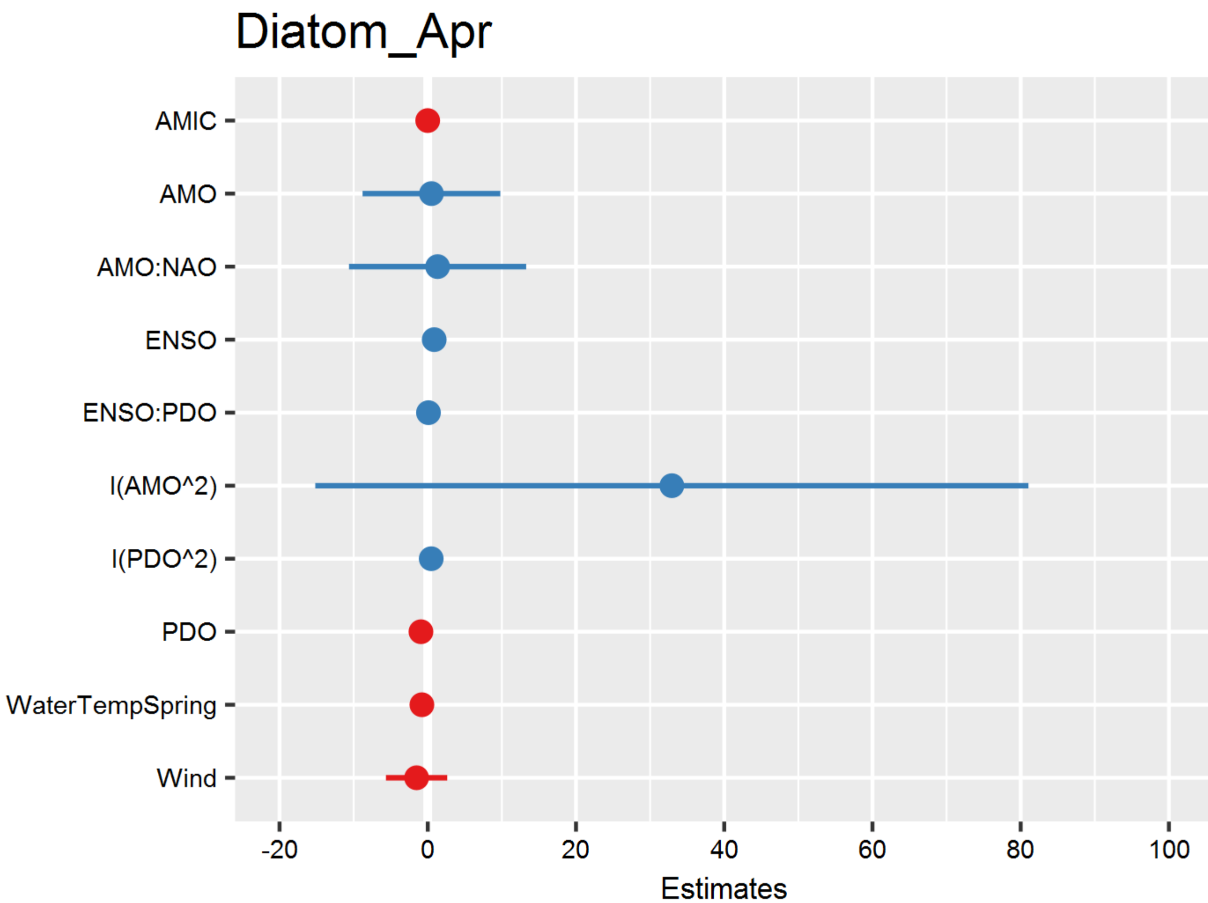

Figure 4-1. Regression coefficient plot (diatom model).

Figure 4-2. Linear correlations between diatom (spring chlorophyll a), biological parameters, and physical forcings.

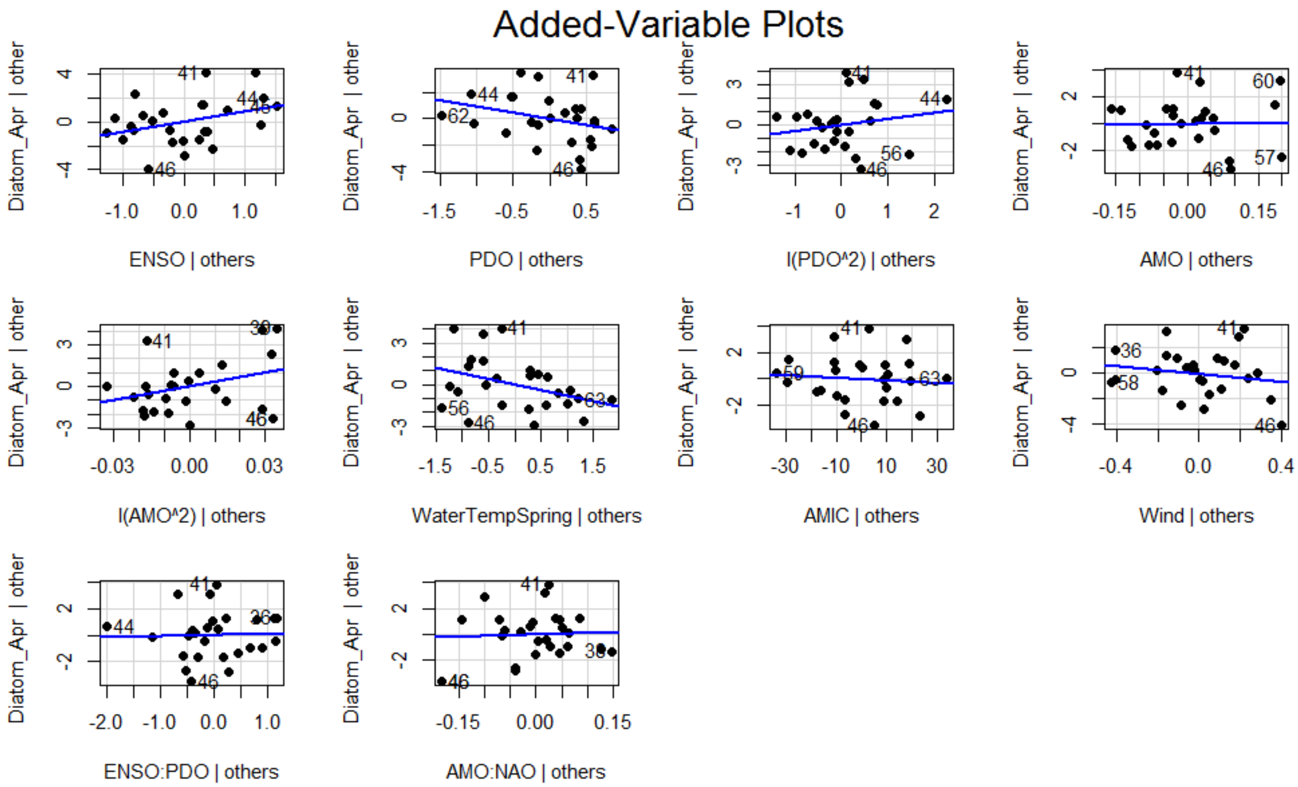

Figure 4-3. Added variable plot for diatom model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the index of the point in the dataset. For example, the number 41 represents the 41st data point out of the total data points.

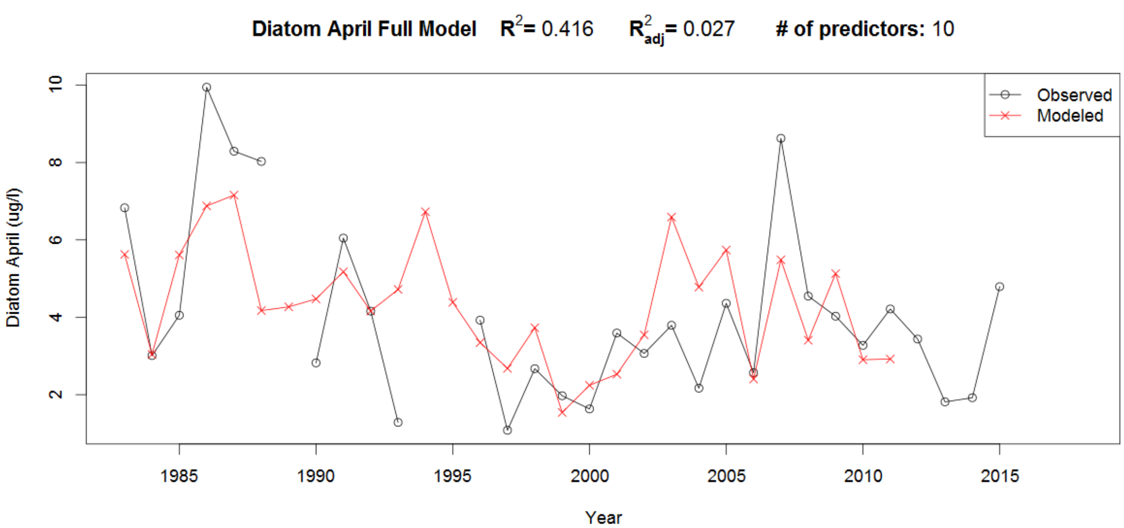

Figure 4-4. Time series plot of modeled vs. observed values (diatom model).

4.2 STANDARDIZED TELECONNECTIONS MODEL

Here, regression summary tables are included for the same model as above, but with standardized predictor variables rather than the original variables. The "scale" function in *R* was used to standardize (normalize) each variable with respect to its mean and variance. These tables show that standardizing the predictor variables does not have a significant effect on the R^2 or estimated regression coefficients of the model. Comparing the following two tables with the preceding two tables, it is clear that standardizing the predictors has virtually no effect on the model results.

Table 4.2-1. Table summarizing the best subsets procedure for the diatom model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	scale(AMO)^2	0.106	0.074	-0.079	-3.882
2	scale(PDO)^2 scale(AMO)^2	0.226	0.169	0.005	-8.007
3	scale(PDO)^2 scale(AMO)^2 scale(WaterTempSpring)	0.294	0.213	0.006	-10.541
4	scale(ENSO) scale(PDO)^2 scale(AMO)^2 scale(WaterTempSpring)	0.336	0.230	-0.028	-12.141
5	scale(ENSO) scale(PDO) scale(PDO)^2 scale(AMO)^2 scale(WaterTempSpring)	0.394	0.268	0.051	-14.588
6	scale(ENSO) scale(PDO) scale(PDO)^2 scale(AMO)^2 scale(WaterTempSpring) scale(Wind)	0.408	0.220	-0.209	-11.218
7	scale(ENSO) scale(PDO) scale(PDO)^2 scale(AMO) scale(AMO)^2 scale(WaterTempSpring) scale(Wind)	0.410	0.181	-0.400	-11.153
8	scale(ENSO) scale(PDO) scale(PDO)^2 scale(AMO) scale(AMO)^2 scale(WaterTempSpring) scale(AMIC) scale(Wind)	0.414	0.138	-0.459	-11.123
9	scale(ENSO) scale(PDO) scale(PDO)^2 scale(AMO) scale(AMO)^2 scale(WaterTempSpring) scale(AMIC) scale(Wind) scale(AMO):scale(NAO)	0.416	0.087	-0.773	-11.015
10	scale(ENSO) scale(PDO) scale(PDO)^2 scale(AMO) scale(AMO)^2 scale(WaterTempSpring) scale(AMIC) scale(Wind) scale(ENSO):scale(PDO) scale(AMO):scale(NAO)	0.416	0.027	-1.138	-10.869

Monthly Teleconnections Model

$$\begin{aligned}\text{Diatom} = {} & 1.75 + 0.43\text{ENSO.monthlyJan} + -0.34\text{PDO.monthlyJan} + 0.19\text{PDO.monthlyJan}^2 + \\ & 0.1\text{AMO.monthlyJan} + 20.11\text{AMO.monthlyJan}^2 + \\ & 0.15\text{ENSO.monthlyJan:PDO.monthlyJan} + \\ & -0.66\text{AMO.monthlyJan:NAO.monthlyJan} + \\ & -0.56\text{WaterTempSpring} + -0.01\text{AMIC} + -2.07\text{monthlyErieWindsMAY} + e \qquad (4.2\text{-}1)\end{aligned}$$

Table 4.2-2. Regression output for diatom model.

	Diatom_Apr			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	17.75 *	2.26 – 33.23	2.25	0.040
ENSO.monthlyJan	0.43	-0.62 – 1.48	0.80	0.434
PDO.monthlyJan	-0.34	-1.46 – 0.77	-0.60	0.556
PDO.monthlyJan^2	0.19	-0.63 – 1.01	0.45	0.658
AMO.monthlyJan	0.10	-8.66 – 8.85	0.02	0.983
AMO.monthlyJan^2	20.11	-20.44 – 60.67	0.97	0.346
WaterTempSpring	-0.56	-1.50 – 0.37	-1.18	0.256
AMIC	-0.01	-0.06 – 0.04	-0.57	0.575
monthlyErieWinds MAY	-2.07	-4.51 – 0.38	-1.66	0.118
ENSO.monthlyJan:PDO.monthlyJan	0.15	-0.86 – 1.16	0.29	0.773
AMO.monthlyJan:NAO.monthlyJan	-0.66	-5.47 – 4.16	-0.27	0.792
Observations	26			
R^2 / adjusted R^2	0.454 / 0.090			
***** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.2-3. Table summarizing the best subsets procedure for the diatom model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	monthlyErieWindsMAY	0.282	0.252	0.166	35.802
2	AMO.monthlyJan^2 monthlyErieWindsMAY	0.348	0.292	0.121	33.548
3	ENSO.monthlyJan AMO.monthlyJan^2 monthlyErieWindsMAY	0.372	0.286	0.067	32.845
4	ENSO.monthlyJan AMO.monthlyJan^2 WaterTempSpring monthlyErieWindsMAY	0.400	0.286	0.044	31.921
5	ENSO.monthlyJan PDO.monthlyJan AMO.monthlyJan^2 WaterTempSpring monthlyErieWindsMAY	0.421	0.276	0.036	31.319
6	ENSO.monthlyJan PDO.monthlyJan PDO.monthlyJan^2 AMO.monthlyJan^2 WaterTempSpring monthlyErieWindsMAY	0.437	0.259	-0.085	30.876
7	ENSO.monthlyJan PDO.monthlyJan PDO.monthlyJan^2 AMO.monthlyJan^2 WaterTempSpring AMIC monthlyErieWindsMAY	0.448	0.233	-0.135	30.642
8	ENSO.monthlyJan PDO.monthlyJan PDO.monthlyJan^2 AMO.monthlyJan^2 WaterTempSpring AMIC monthlyErieWindsMAY ENSO.monthlyJan:PDO.monthlyJan	0.451	0.193	-0.379	30.741
9	ENSO.monthlyJan PDO.monthlyJan PDO.monthlyJan^2 AMO.monthlyJan^2 WaterTempSpring AMIC monthlyErieWindsMAY ENSO.monthlyJan:PDO.monthlyJan AMO.monthlyJan:NAO.monthlyJan	0.454	0.147	-0.633	30.879
10	ENSO.monthlyJan PDO.monthlyJan PDO.monthlyJan^2 AMO.monthlyJan AMO.monthlyJan^2 WaterTempSpring AMIC monthlyErieWindsMAY ENSO.monthlyJan:PDO.monthlyJan AMO.monthlyJan:NAO.monthlyJan	0.454	0.090	-0.945	31.127

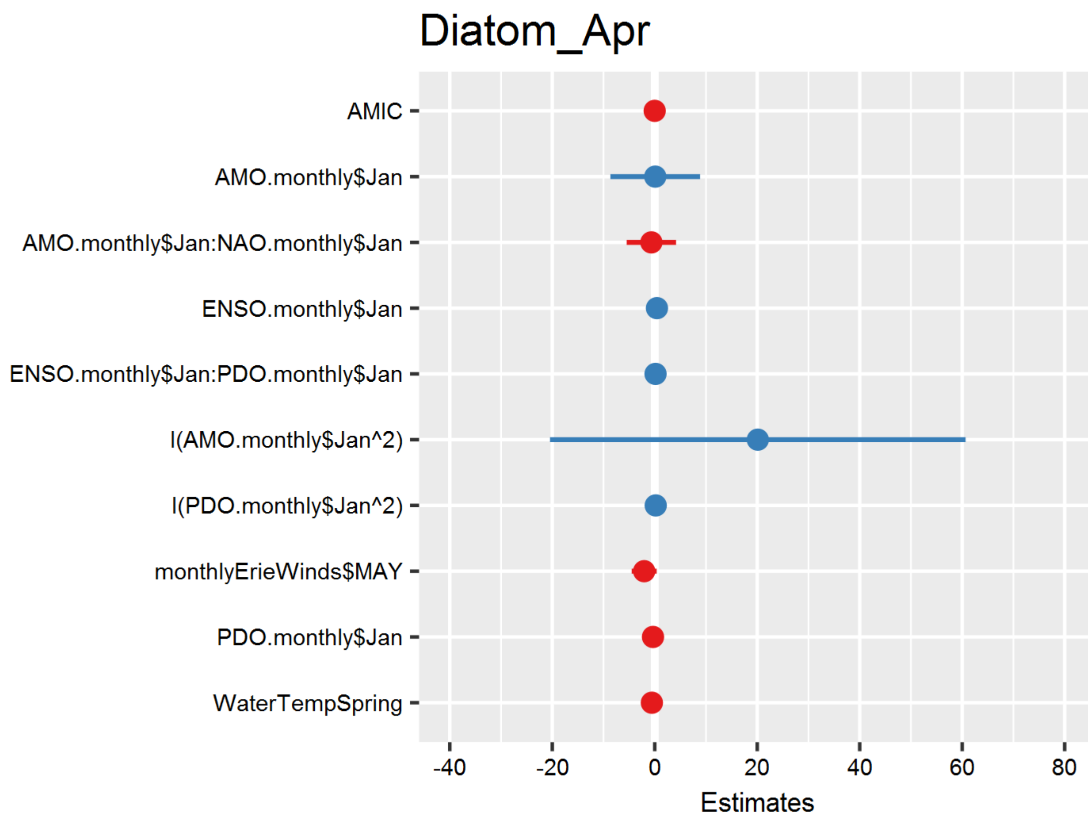

Figure 4-5. Regression coefficient plot (diatom model).

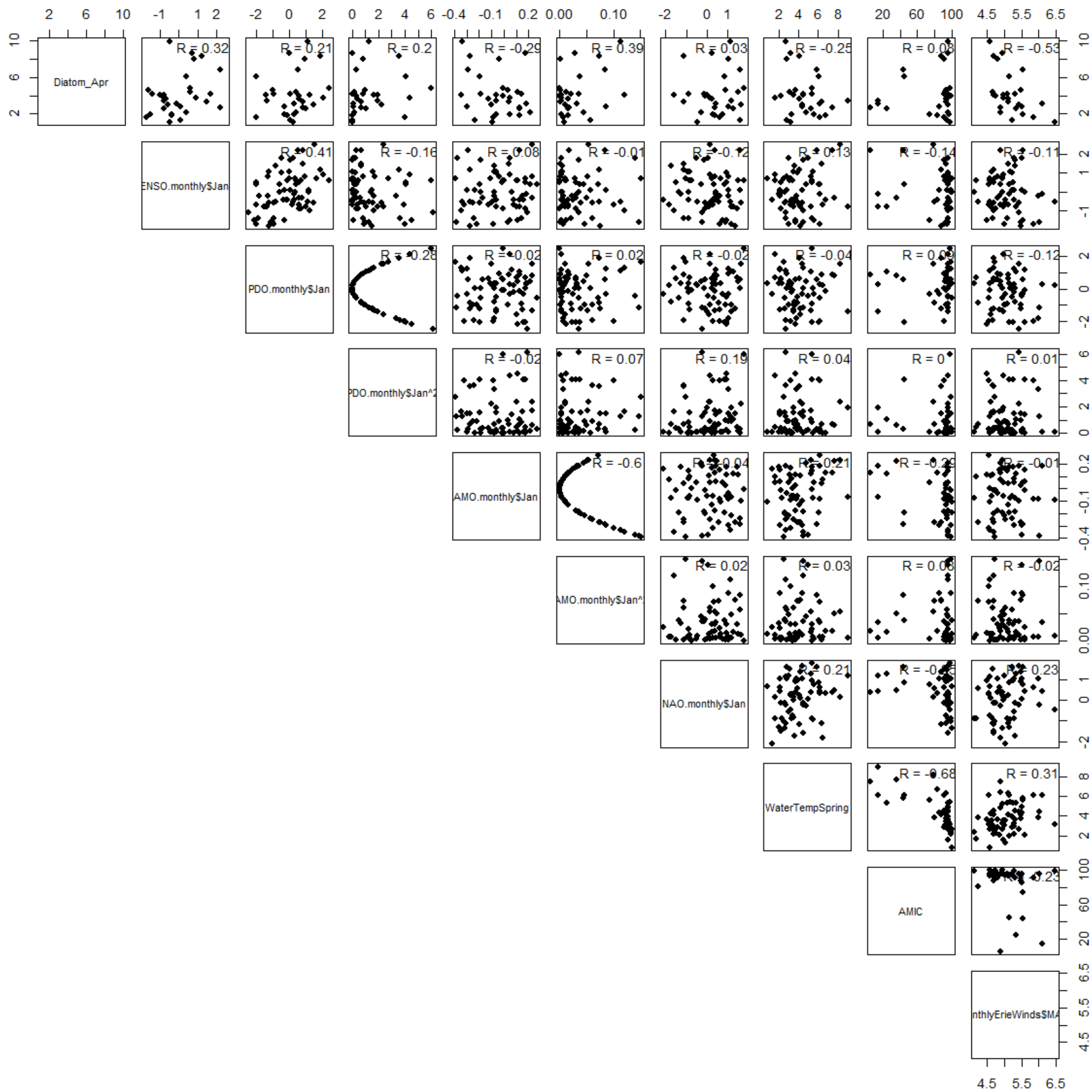

Figure 4-6. Linear correlations between diatom (spring chlorophyll a), biological parameters, and physical forcings.

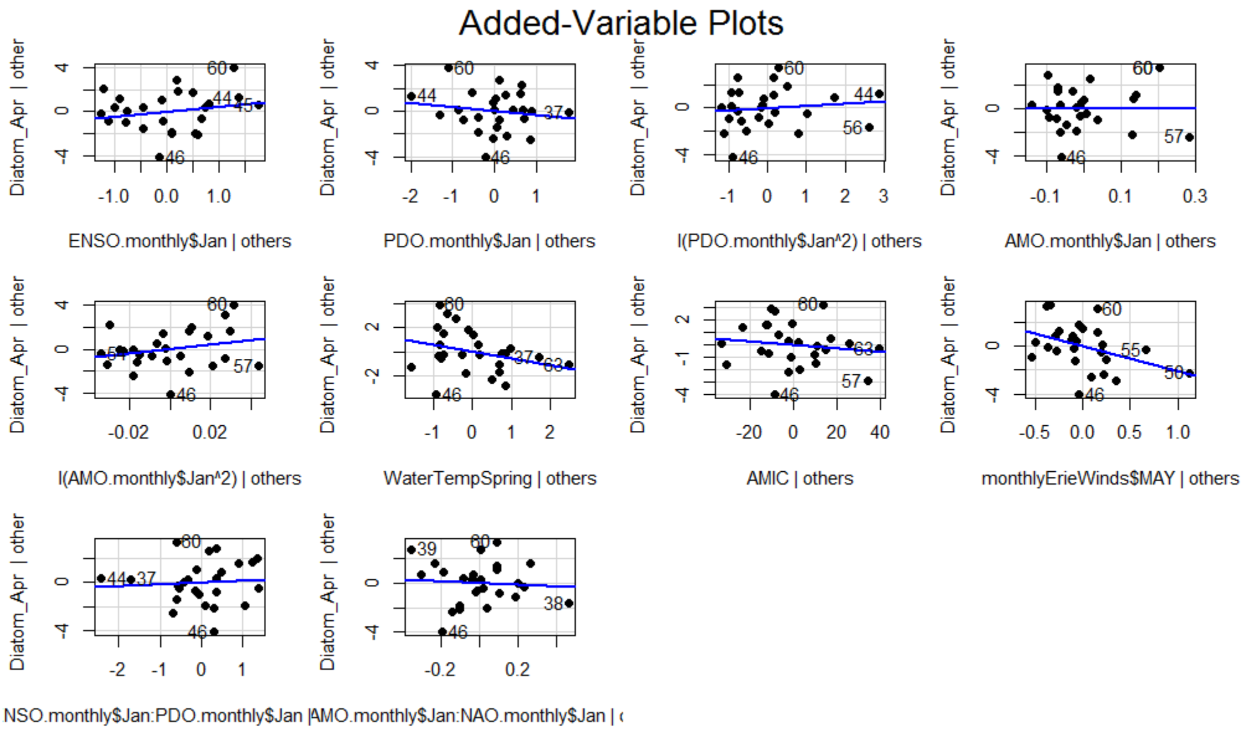

Figure 4-7. Added variable plot for diatom model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

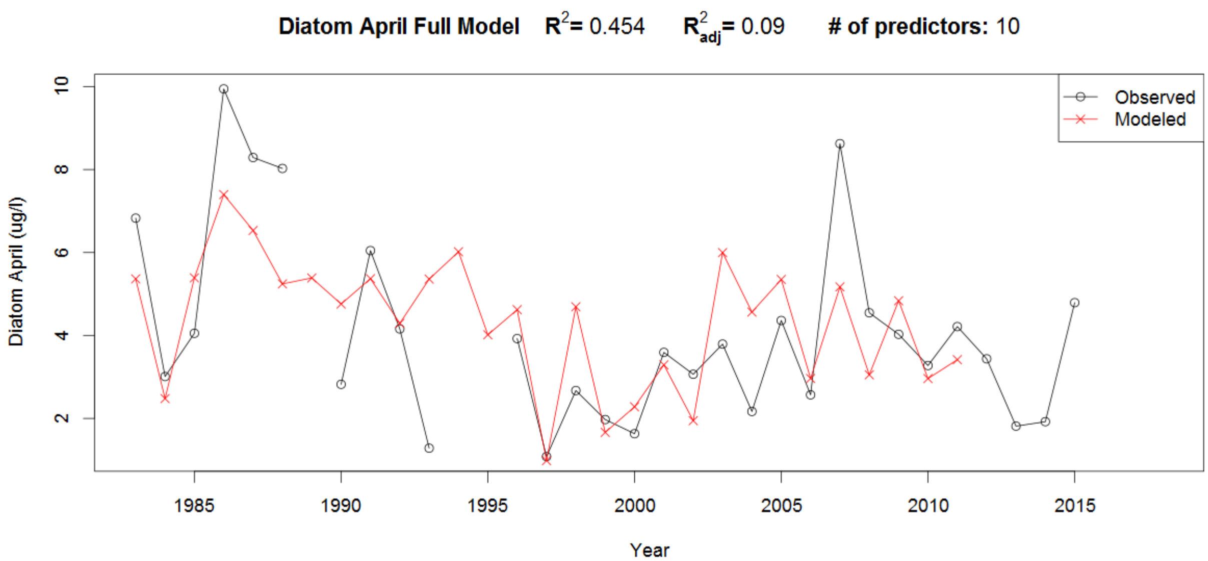

Figure 4-8. Time series plot of modeled vs. observed values (diatom model).

4.3 HYPOLIMNETIC OXYGEN DEMAND (HOD) MODELS

Table 4.3-1. Correlations and p-values of HOD with teleconnection patterns. Significant correlations shown in boldface.

Index	r	p value	Significance (%)
ENSO	0.084	0.575	42.5
$ENSO^2$	0.119	0.426	57.4
NAO	**-0.307**	**0.036**	**96.4**
NAO^2	-0.206	0.166	83.4
AMO	-0.081	0.588	41.2
AMO^2	**0.375**	**0.009**	**99.1**
PDO	0.101	0.499	50.1
PDO^2	0.144	0.335	66.5

Winter Teleconnections Model

$$HOD = -1.26 + -0.25NAO + 0.40AMO + 5.87AMO^2$$
$$+ -0.03WaterTempSummer + -0.001DurationDays + -0.19U\text{-}WindApr$$
$$+ 0.59\log(Tpload) + e \qquad (4.3\text{-}1)$$

Table 4.3-2. Regression output for HOD model.

	HOD			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	-1.26	-8.05 – 5.53	-0.36	0.718
NAO	-0.25 *	-0.43 – -0.06	-2.59	0.015
AMO	0.40	-0.61 – 1.40	0.78	0.445
AMO^2	5.87 *	0.55 – 11.19	2.16	0.040
WaterTempSummer	-0.03	-0.20 – 0.15	-0.29	0.778
Duration..days.	-0.001	-0.01 – 0.00	-1.01	0.322
U-WindApr	-0.19 *	-0.36 – -0.02	-2.15	0.041
log(Tpload)	0.59 *	0.05 – 1.13	2.12	0.043
Observations	34			
R^2 / adjusted R^2	0.455 / 0.308			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.3-3. Table summarizing the best subsets procedure for the HOD model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	log(Tpload)	0.220	0.200	0.144	-72.775
2	NAO log(Tpload)	0.293	0.256	0.191	-74.722
3	NAO AMO^2 U-WindApr	0.361	0.302	0.099	-64.815
4	NAO AMO^2 U-WindApr log(Tpload)	0.437	0.361	0.161	-65.828
5	NAO AMO AMO^2 U-WindApr log(Tpload)	0.458	0.364	0.165	-64.146
6	NAO AMO AMO^2 WaterTempSummer U-WindApr log(Tpload)	0.458	0.342	0.122	-61.569
7	NAO AMO AMO^2 WaterTempSummer Duration..days. U-WindApr log(Tpload)	0.455	0.308	0.004	-56.798

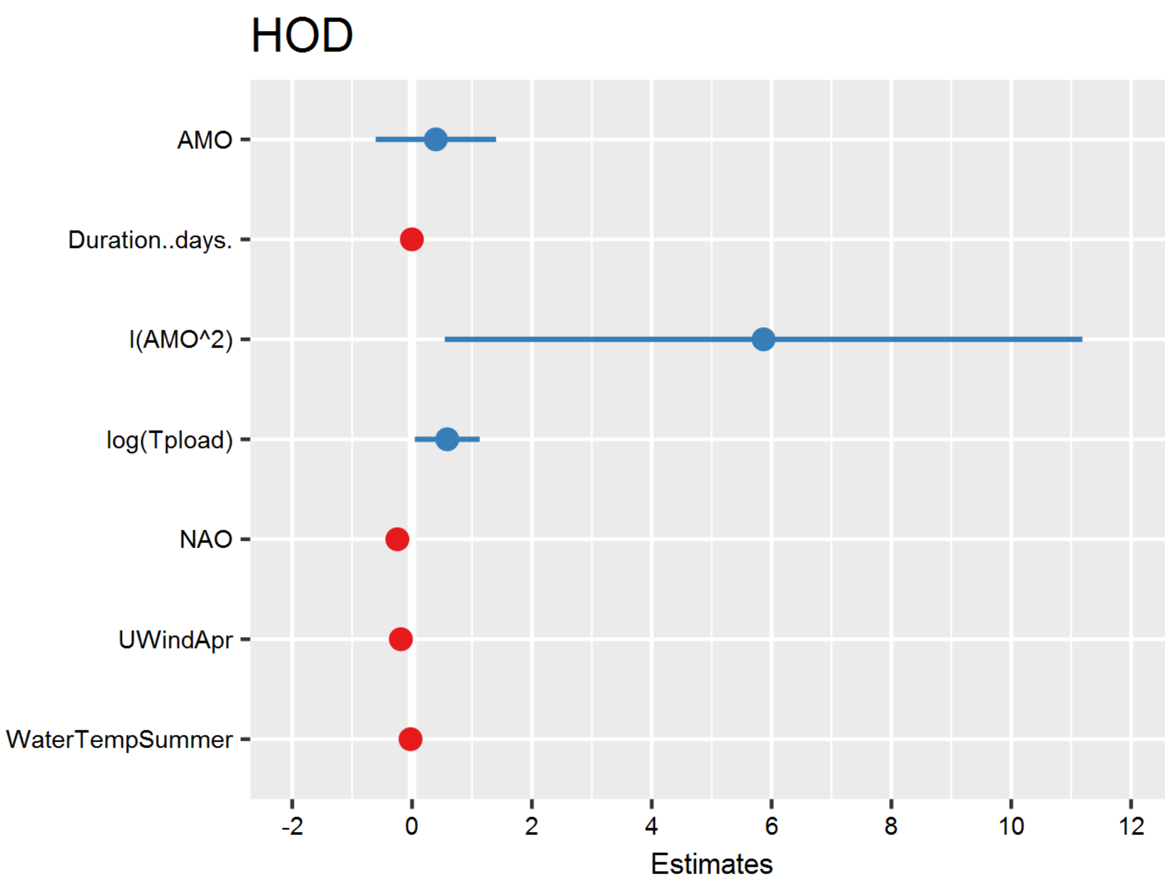

Figure 4-9. Regression coefficient plot (HOD model).

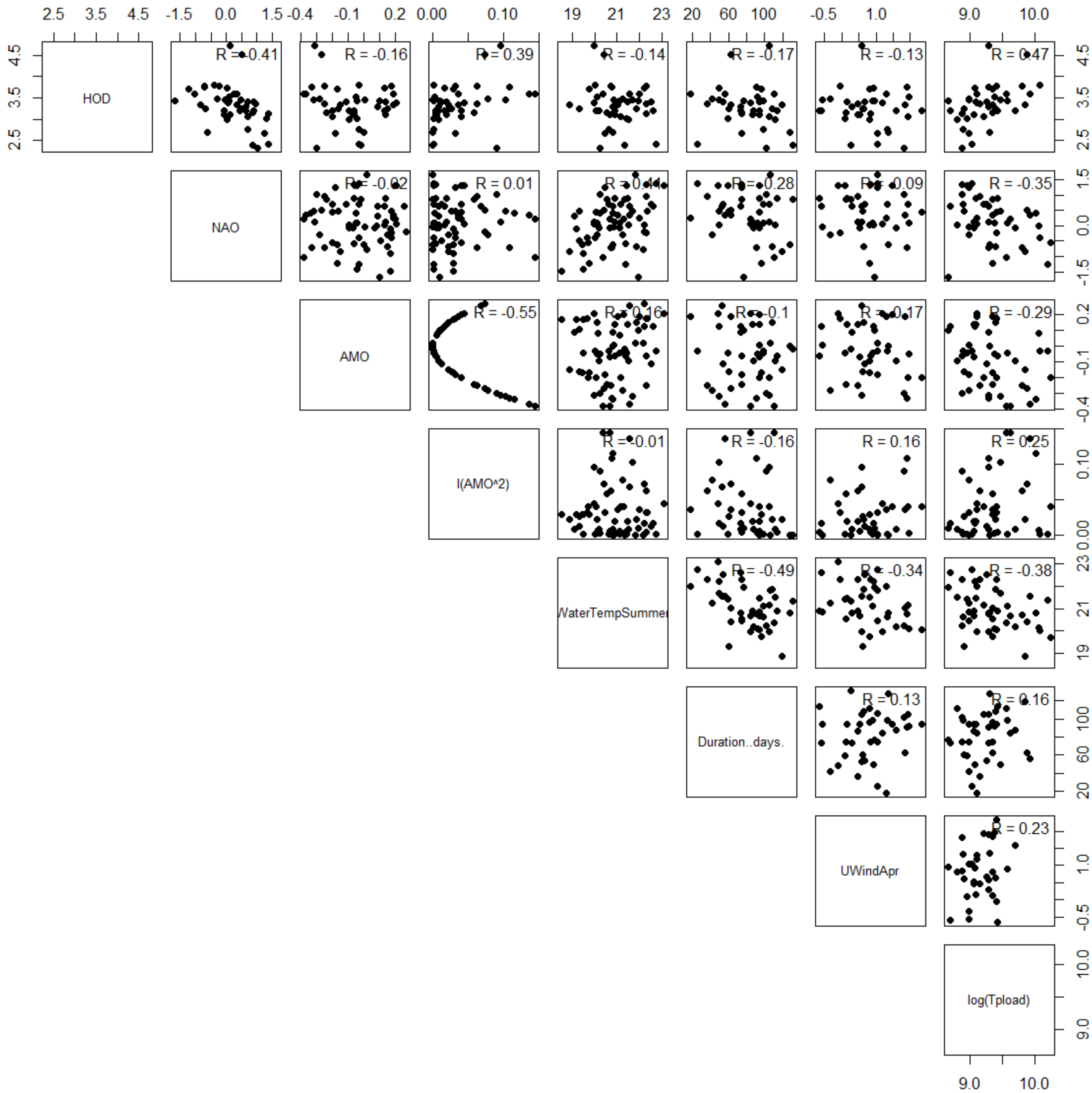

Figure 4-10. Linear correlations between HOD, biological parameters, and physical forcings.

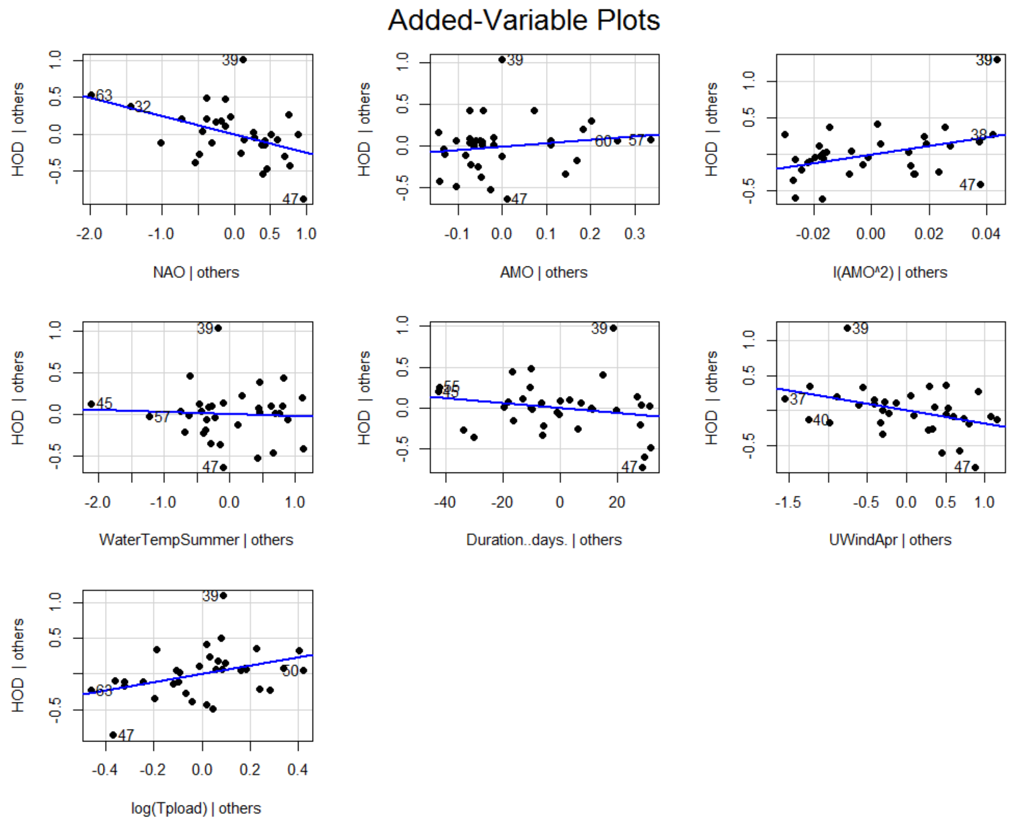

Figure 4-11. Added variable plot for HOD model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

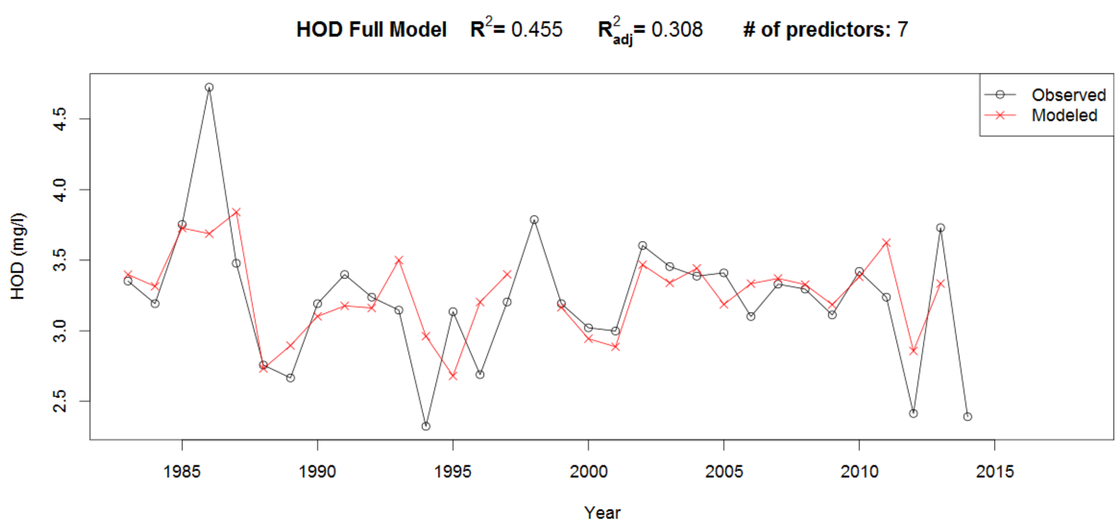

Figure 4-12. Time series plot of modeled vs. observed values (HOD model).

Monthly Teleconnections Model

$$HOD = -0.14 + -0.29NAO + 4.25AMO.monthlyMar^2{:}PDO.monthlyMar^2 + -0.01DurationDays + 0.43\log(Tpload) + e \quad (4.3\text{-}2)$$

Table 4.3-4 .Regression output for HOD model.

	HOD			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	-0.14	-2.77 – 2.49	-0.11	0.916
NAO	-0.29 ***	-0.42 – -0.16	-4.43	<0.001
Duration..days.	-0.01 ***	-0.01 – -0.00	-3.92	<0.001
log(Tpload)	0.43 **	0.14 – 0.72	2.95	0.006
AMO.monthlyMar^2:PDO.monthlyMar^2	4.25 ***	2.93 – 5.58	6.30	<0.001
Observations	39			
R^2 / adjusted R^2	0.684 / 0.647			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.3-5. Table summarizing the best subsets procedure for the HOD model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	**Predictors**	**Rsquare**	**AdjRsq**	**PredRsq**	**BIC**
1	AMO.monthlyMar^2:PDO.monthlyMar^2	0.268	0.250	0.141	NA
2	log(Tpload) AMO.monthlyMar^2:PDO.monthlyMar^2	0.409	0.378	0.235	NA
3	NAO Duration..days. AMO.monthlyMar^2:PDO.monthlyMar^2	0.635	0.605	0.518	NA
4	NAO Duration..days. log(Tpload) AMO.monthlyMar^2:PDO.monthlyMar^2	0.684	0.647	0.573	NA

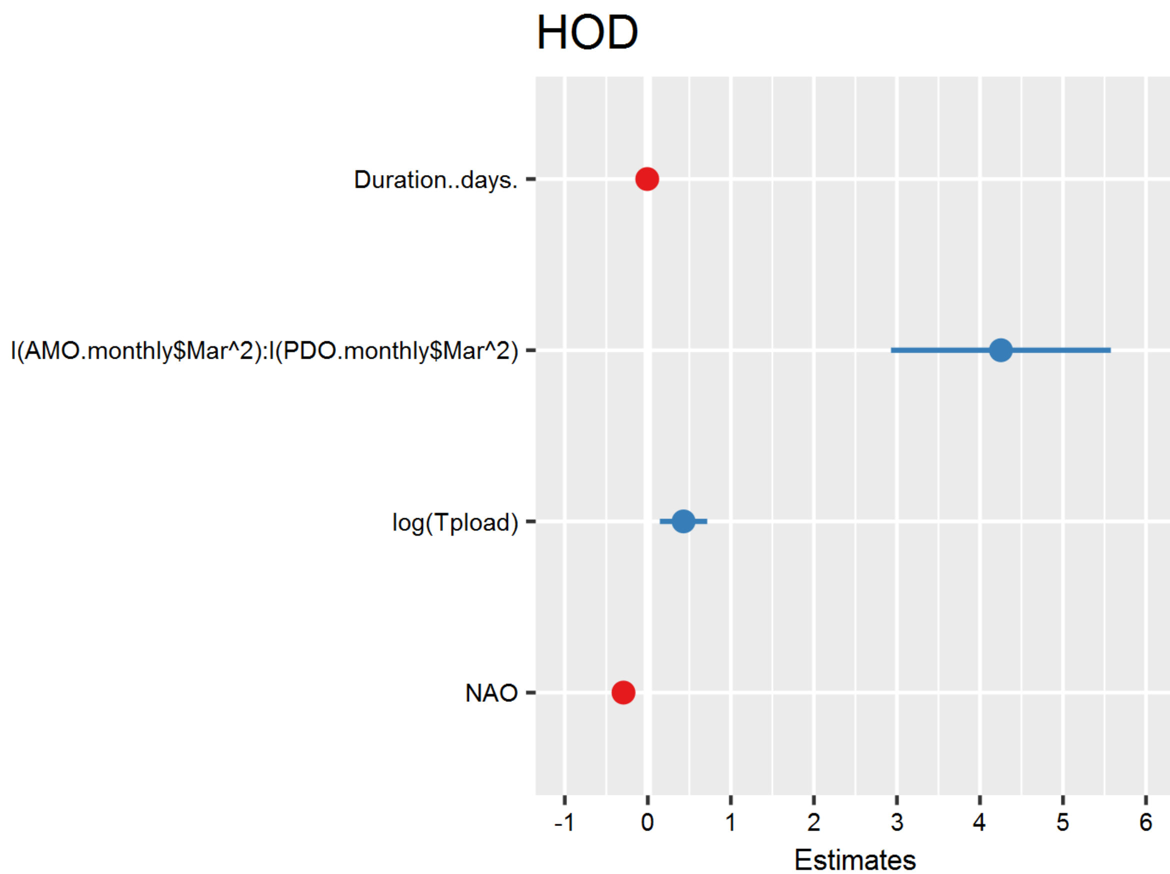

Figure 4-13. Regression coefficient plot (HOD model).

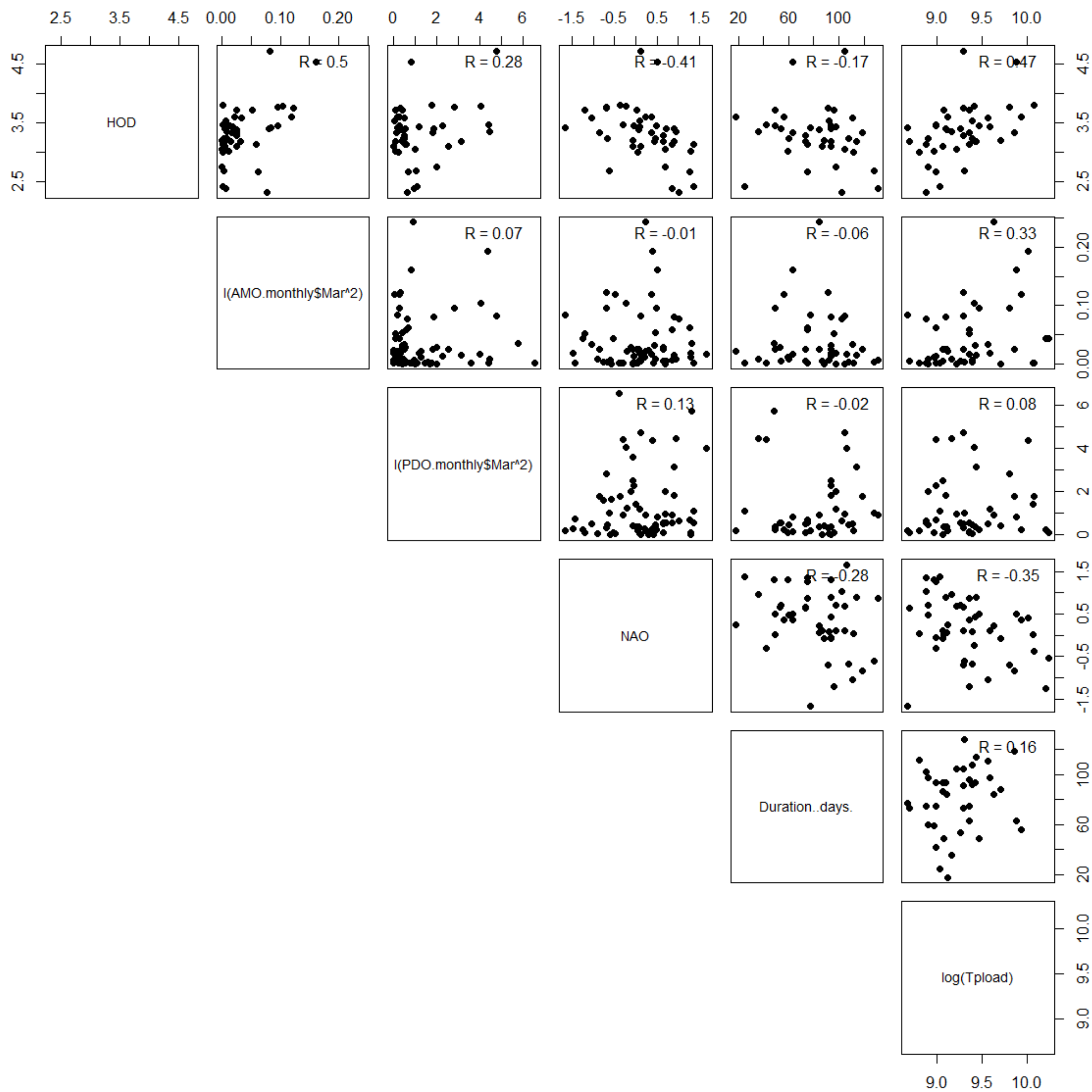

Figure 4-14. Linear correlations between HOD, biological parameters, and physical forcings.

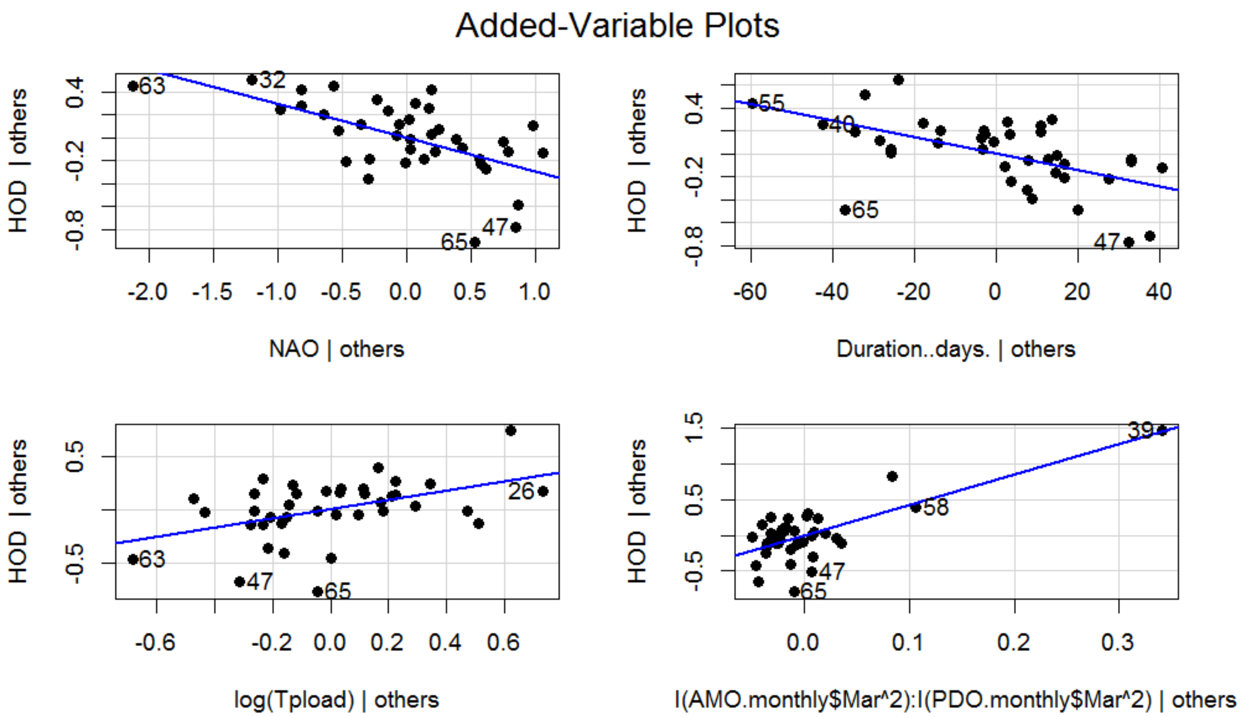

Figure 4-15. Added variable plot for HOD model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

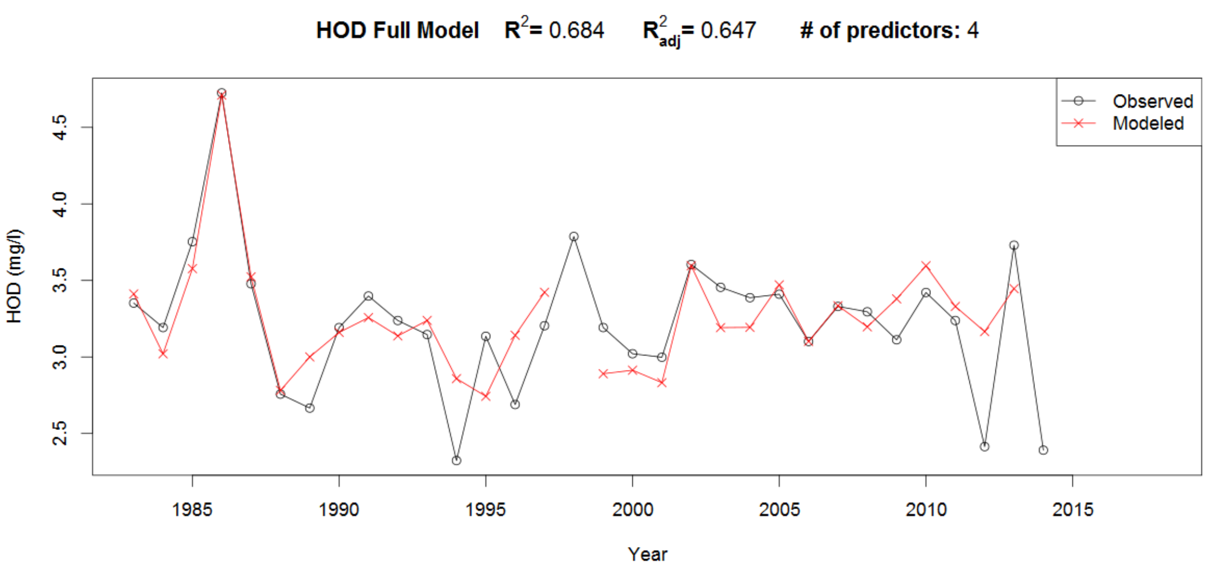

Figure 4-16. Time series plot of modeled vs. observed values (HOD model).

4.4 MAXIMUM HYPOXIC AREA MODELS

Table 4.4-1. Correlations and p-values of maximum hypoxic area with teleconnection patterns. Significant correlations shown in boldface.

Index	**r**	**p value**	**Significance (%)**
ENSO	-0.022	0.913	8.7
ENSO^2	0.142	0.480	52.0
NAO	-0.129	0.522	47.8
NAO^2	0.009	0.966	3.4
AMO	0.017	0.933	6.7
AMO^2	-0.042	0.836	16.4
PDO	-0.011	0.956	4.4
PDO^2	0.403	0.037	96.3

Winter Teleconnections Model

$$\text{MaxArea} = 35.98 + 0.95\text{ENSO}^2 + 1.3\text{PDO}^2 + -0.85\text{ENSO}^2\text{:PDO}^2 + 0.04\text{WaterTempSpring} + -0.02\text{Duration..days.} + -5.05\text{monthlyErieWindsJUL} + -0.35\text{DiatomApr} + e \quad (4.4\text{-}1)$$

Table 4.4-2. Regression output for maximum hypoxic area model.

	MaxArea			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	35.98 *	10.99 – 60.97	2.82	0.014
ENSO^2	0.95	-0.77 – 2.66	1.08	0.299
PDO^2	1.30	-0.27 – 2.88	1.62	0.127
WaterTempSpring	0.04	-1.07 – 1.15	0.07	0.944
Duration..days.	-0.02	-0.08 – 0.04	-0.52	0.608
Diatom_Apr	-0.35	-1.01 – 0.31	-1.04	0.317
monthlyErieWindsJUL	-5.05 *	-8.81 – -1.28	-2.63	0.020
ENSO^2:PDO^2	-0.85	-2.31 – 0.61	-1.14	0.274
Observations	22			
R^2 / adjusted R^2	0.453 / 0.179			
***** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.4-3. Table summarizing the best subsets procedure for the maximum area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	monthlyErieWindsJUL	0.185	0.151	0.069	56.146
2	PDO^2 monthlyErieWindsJUL	0.268	0.205	0.151	56.549
3	PDO^2 Diatom_Apr monthlyErieWindsJUL	0.346	0.242	0.122	50.790
4	ENSO^2 PDO^2 Diatom_Apr monthlyErieWindsJUL	0.391	0.255	0.071	53.565
5	ENSO^2 PDO^2 Diatom_Apr monthlyErieWindsJUL ENSO^2:PDO^2	0.425	0.256	0.064	56.963
6	ENSO^2 PDO^2 Duration..days. Diatom_Apr monthlyErieWindsJUL ENSO^2:PDO^2	0.453	0.234	-0.039	59.551
7	ENSO^2 PDO^2 WaterTempSpring Duration..days. Diatom_Apr monthlyErieWindsJUL ENSO^2:PDO^2	0.453	0.179	-0.148	63.626

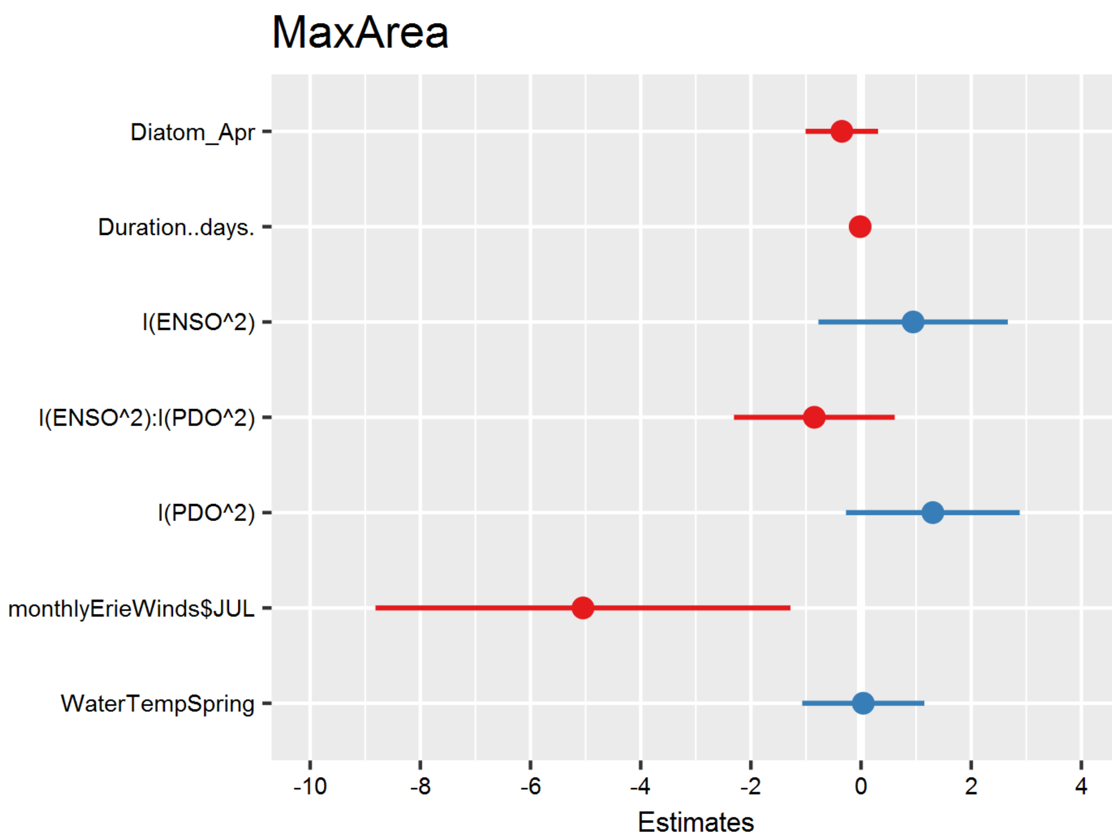

Figure 4-17. Regression coefficient plot (maximum hypoxic extent model.

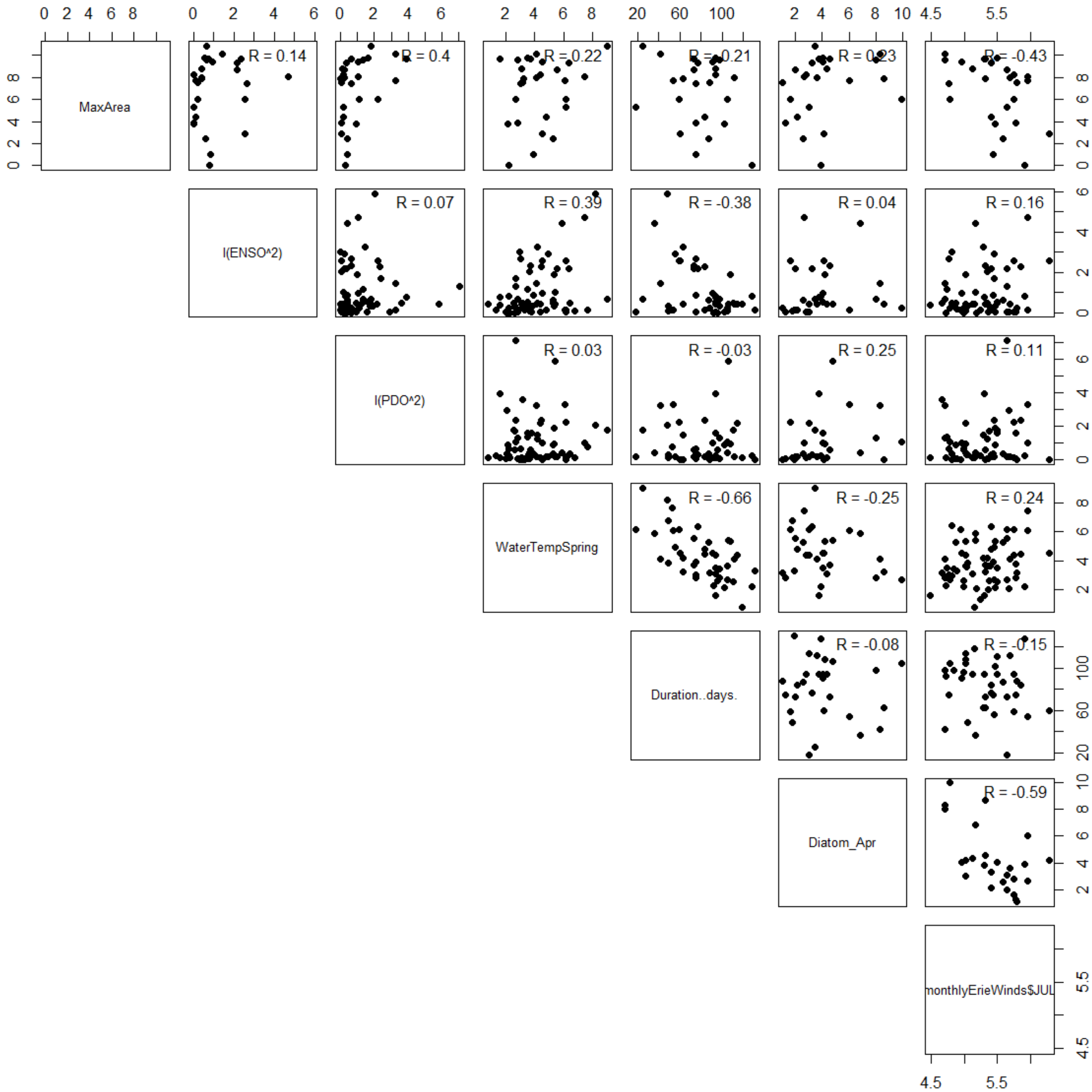

Figure 4-18. Linear correlations between maximum hypoxic extent, biological parameters, and physical forcings.

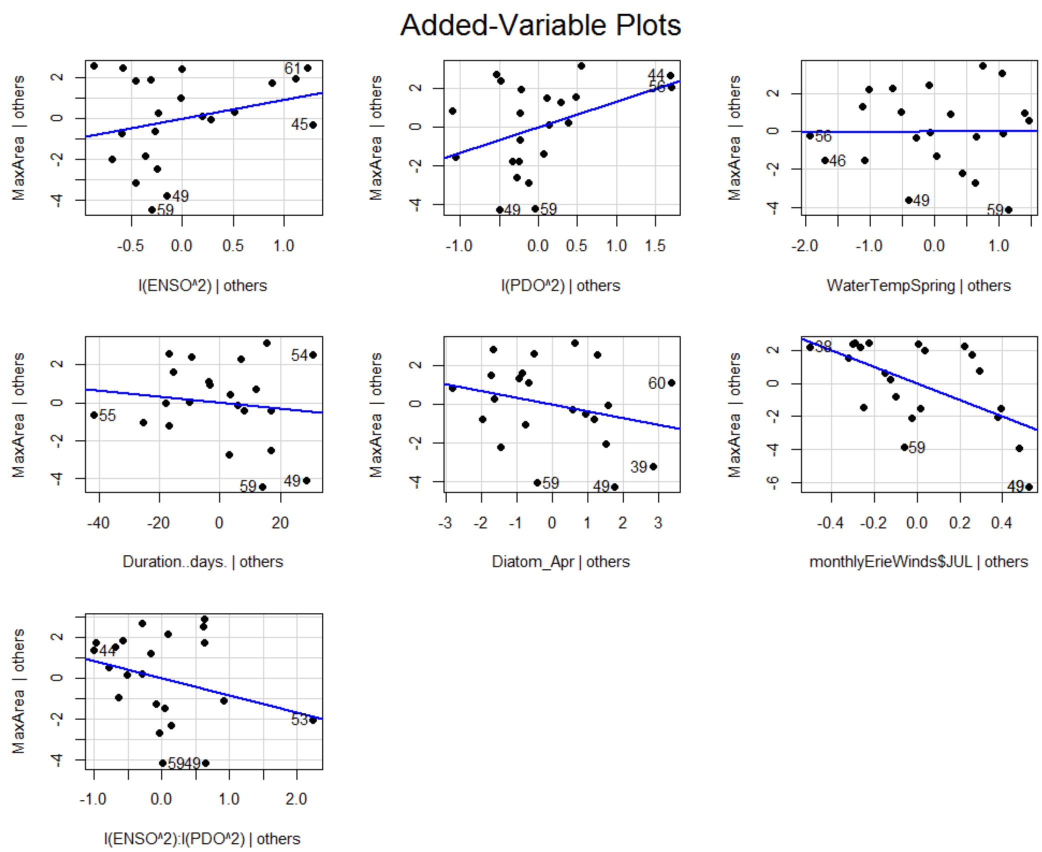

Figure 4-19. Added variable plot for maximum hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data in the data time series.

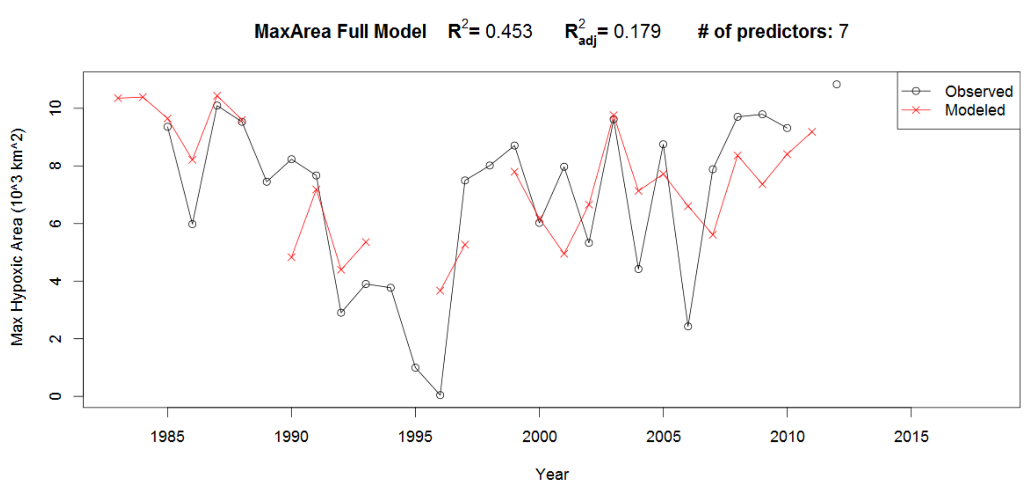

Figure 4-20. Time series plot of modeled vs. observed values (maximum hypoxic area model).

Monthly Teleconnections Model

$$MaxArea = 2.70 + 0.97PDO^2 + 3.51AMO + 3.29EriePrecip\_PMarAprRatio + 0.04\log(Tpload) + e \quad (4.4\text{-}2)$$

Table 4.4-4. Regression output for maximum hypoxic area model.

	MaxArea			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	2.70	-32.29 – 37.69	0.15	0.881
PDO^2	0.97 *	0.09 – 1.86	2.15	0.043
eriePrecip_PMarAprRatio	3.29 **	1.33 – 5.25	3.29	0.003
AMO	3.51	-2.18 – 9.20	1.21	0.239
log(Tpload)	0.04	-3.79 – 3.88	0.02	0.983
Observations	27			
R^2 / adjusted R^2	0.449 / 0.348			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.4-5. Table summarizing the best subsets procedure for the maximum area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	**Predictors**	**Rsquare**	**AdjRsq**	**PredRsq**	**BIC**
1	eriePrecip_PMarAprRatio	0.316	0.288	0.188	53.324
2	PDO^2 eriePrecip_PMarAprRatio	0.412	0.363	0.272	52.145
3	PDO^2 eriePrecip_PMarAprRatio AMO	0.449	0.377	0.240	53.252
4	PDO^2 eriePrecip_PMarAprRatio AMO log(Tpload)	0.449	0.348	0.188	55.706

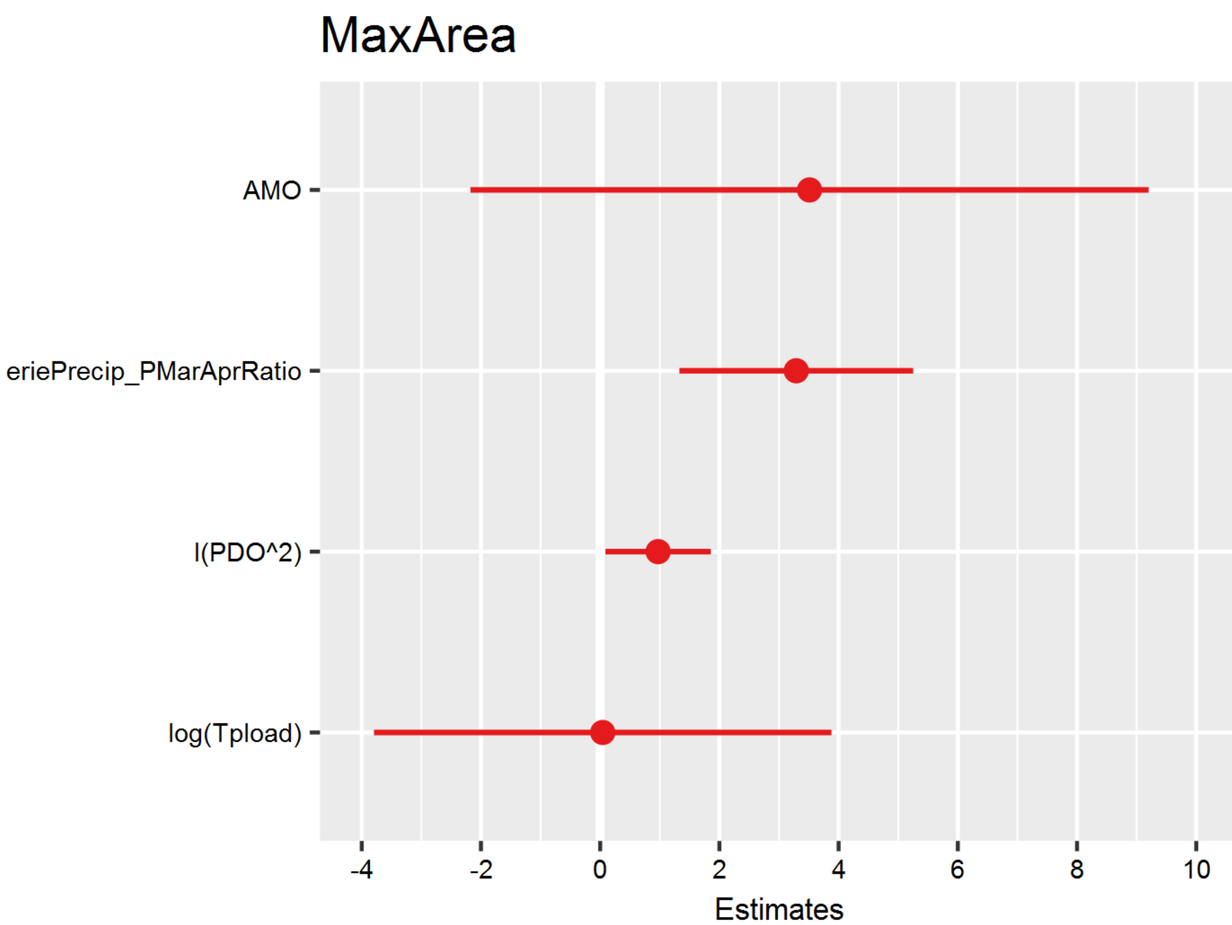

Figure 4-21. Regression coefficient plot (maximum hypoxic extent model).

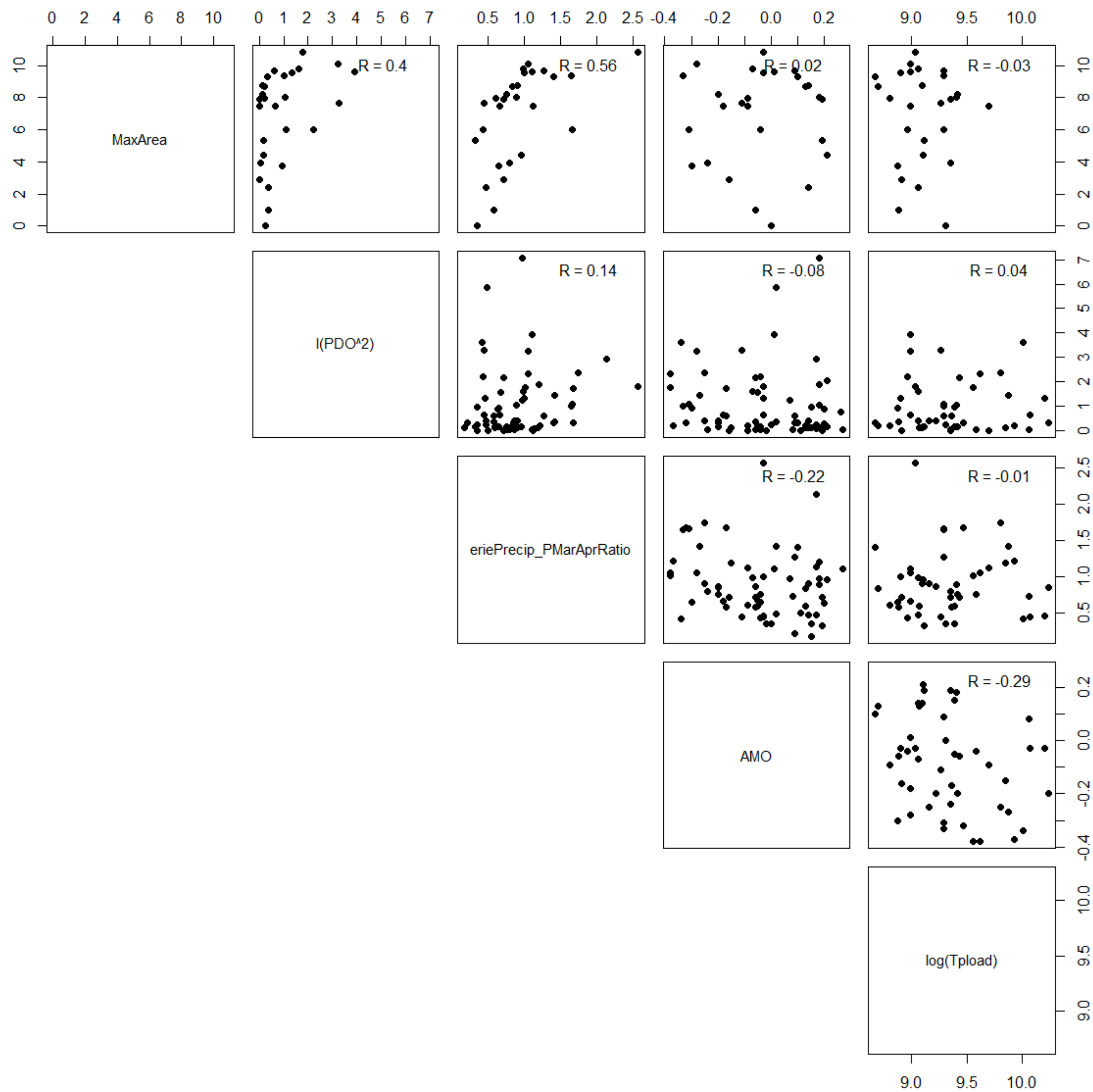

Figure 4-22. Linear correlations between maximum hypoxic extent, biological parameters, and physical forcings.

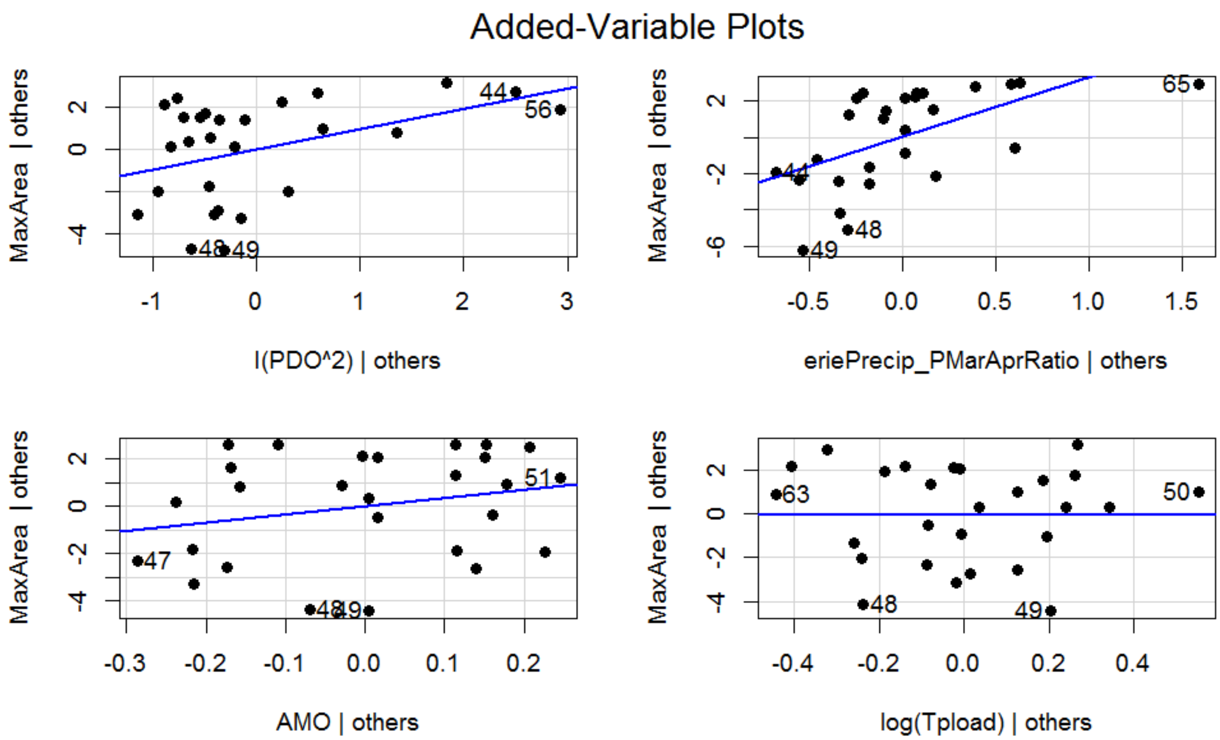

Figure 4-23. Added variable plot for maximum hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

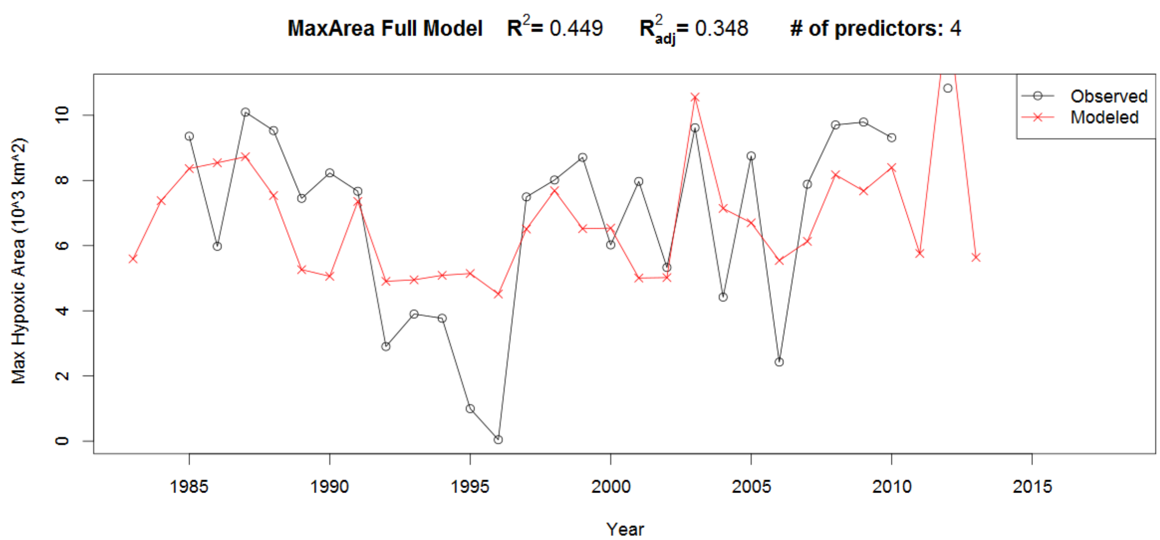

Figure 4-24. Time series plot of modeled vs. observed values (maximum hypoxic area model.

4.5 MEAN HYPOXIC AREA MODELS

Mean hypoxic area is referred to as "MeanAreaFull" because this variable includes additional data for 2013-2015 that we digitized from Del Giudice et al. (2018). This data was missing from the NOAA Technical memorandum 173. Notably, the correlation with PDO^2 is no longer significant after including this data for 2013-2015. In particular, the year 2015 worsens the quadratic relationship between PDO and mean hypoxic area. Thus, it no longer appears that quadratic PDO is a significant predictor of mean hypoxic area. This should also be taken into account later in the report for Hypoxic Factor models because Hypoxic Factor is missing 2015 data and as a result is still significantly correlated with PDO^2, which is somewhat misleading.

Table 4.5-1. Correlations and p-values of mean hypoxic area with teleconnection patterns. Significant correlations shown in boldface.

Index	r	p value	Significance (%)
ENSO	0.159	0.447	55.3
$ENSO^2$	0.184	0.380	62.0
NAO	-0.152	0.467	53.3
NAO^2	0.144	0.493	50.7
AMO	-0.319	0.120	88.0
AMO^2	0.182	0.384	61.6
PDO	-0.051	0.810	19.0
PDO^2	0.319	0.120	88.0

Winter Teleconnections Model

$$\text{MeanArea} = -2.74 + 0.84\text{ENSO} + 1.26\text{ENSO}^2 + -0.68\text{NAO}^2 + 0.44\text{PDO}^2 + -4.61\text{AMO}$$
$$+ 0.67\text{WaterTempSpring} + 0.01\text{Duration..days.} + -1.68\text{monthlyErieWindsAPR}$$
$$+ 0.24\text{Diatom\_Apr} + 1.11\log(\text{Tpload}) + e \qquad (4.5\text{-}1)$$

Table 4.5-2. Regression output for mean hypoxic area model.

	MeanAreaFull			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	-2.74	-40.22 – 34.75	-0.14	0.890
ENSO	0.84	-0.50 – 2.18	1.23	0.255
ENSO^2	1.26	-0.30 – 2.82	1.58	0.152
NAO^2	-0.68	-2.36 – 1.00	-0.79	0.453
PDO^2	0.44	-0.25 – 1.13	1.24	0.250
AMO	-4.61	-9.42 – 0.20	-1.88	0.097
WaterTempSpring	0.67	-0.10 – 1.43	1.71	0.125
Duration..days.	0.01	-0.03 – 0.05	0.46	0.660
monthlyErieWindsAPR	-1.68	-3.86 – 0.50	-1.51	0.168
Diatom_Apr	0.24	-0.13 – 0.60	1.25	0.246
log(Tpload)	1.11	-2.58 – 4.79	0.59	0.572
Observations	19			
R^2 / adjusted R^2	0.805 / 0.562			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.5-3. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R2, R2adj, R2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	Duration..days.	0.187	0.150	-0.065	35.921
2	PDO^2 monthlyErieWindsAPR	0.451	0.390	0.266	21.922
3	PDO^2 WaterTempSpring monthlyErieWindsAPR	0.545	0.465	0.318	21.280
4	PDO^2 WaterTempSpring monthlyErieWindsAPR Diatom_Apr	0.661	0.570	0.493	20.624
5	PDO^2 AMO WaterTempSpring monthlyErieWindsAPR Diatom_Apr	0.731	0.635	0.238	22.396
6	ENSO^2 PDO^2 AMO WaterTempSpring monthlyErieWindsAPR Diatom_Apr	0.762	0.653	0.240	25.982
7	ENSO^2 PDO^2 AMO WaterTempSpring Duration..days. monthlyErieWindsAPR Diatom_Apr	0.768	0.621	0.009	30.065
8	ENSO ENSO^2 PDO^2 AMO WaterTempSpring Duration..days. monthlyErieWindsAPR Diatom_Apr	0.786	0.614	0.003	34.797
9	ENSO ENSO^2 NAO^2 PDO^2 AMO WaterTempSpring Duration..days. monthlyErieWindsAPR Diatom_Apr	0.797	0.594	-0.577	39.892
10	ENSO ENSO^2 NAO^2 PDO^2 AMO WaterTempSpring Duration..days. monthlyErieWindsAPR Diatom_Apr log(Tpload)	0.805	0.562	-0.711	45.286

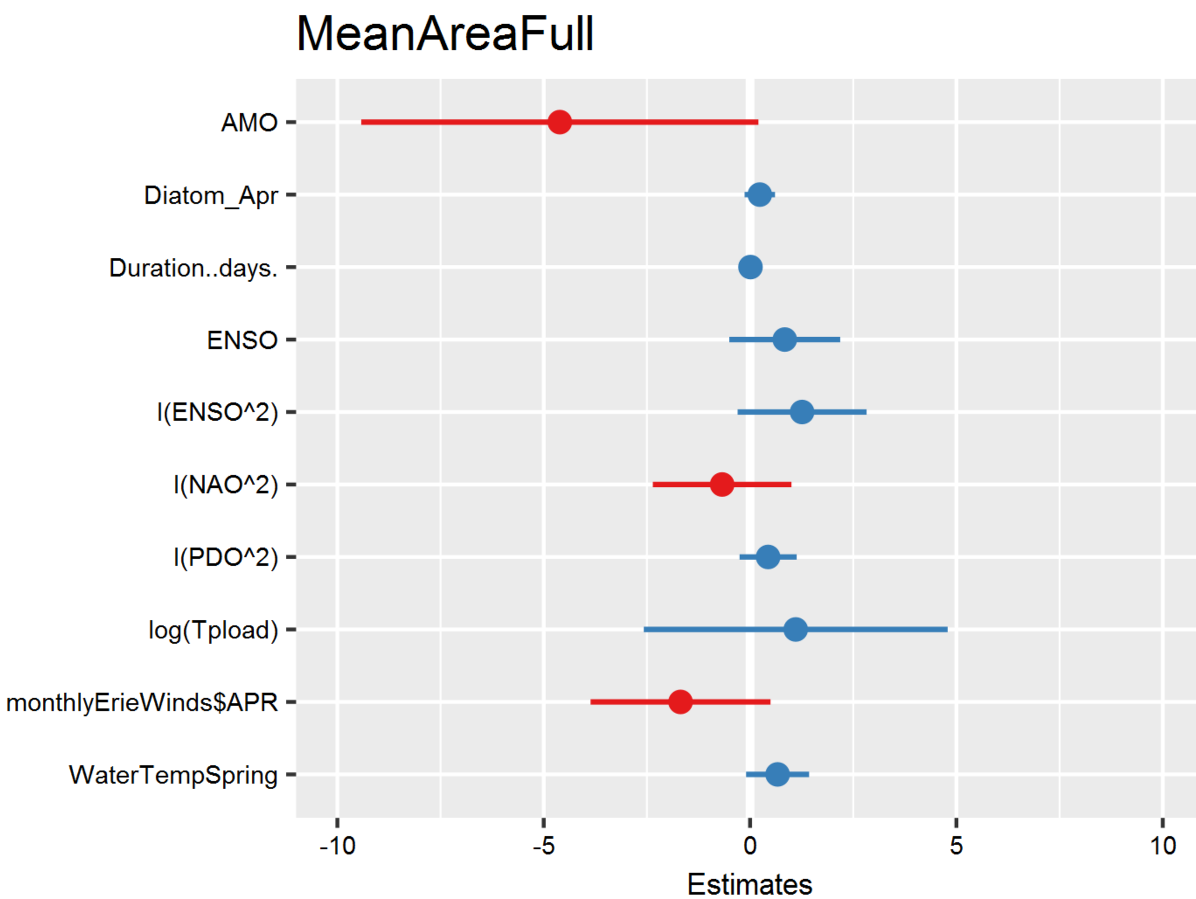

Figure 4-25. Regression coefficient plot (mean hypoxic extent model).

Figure 4-26. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings.

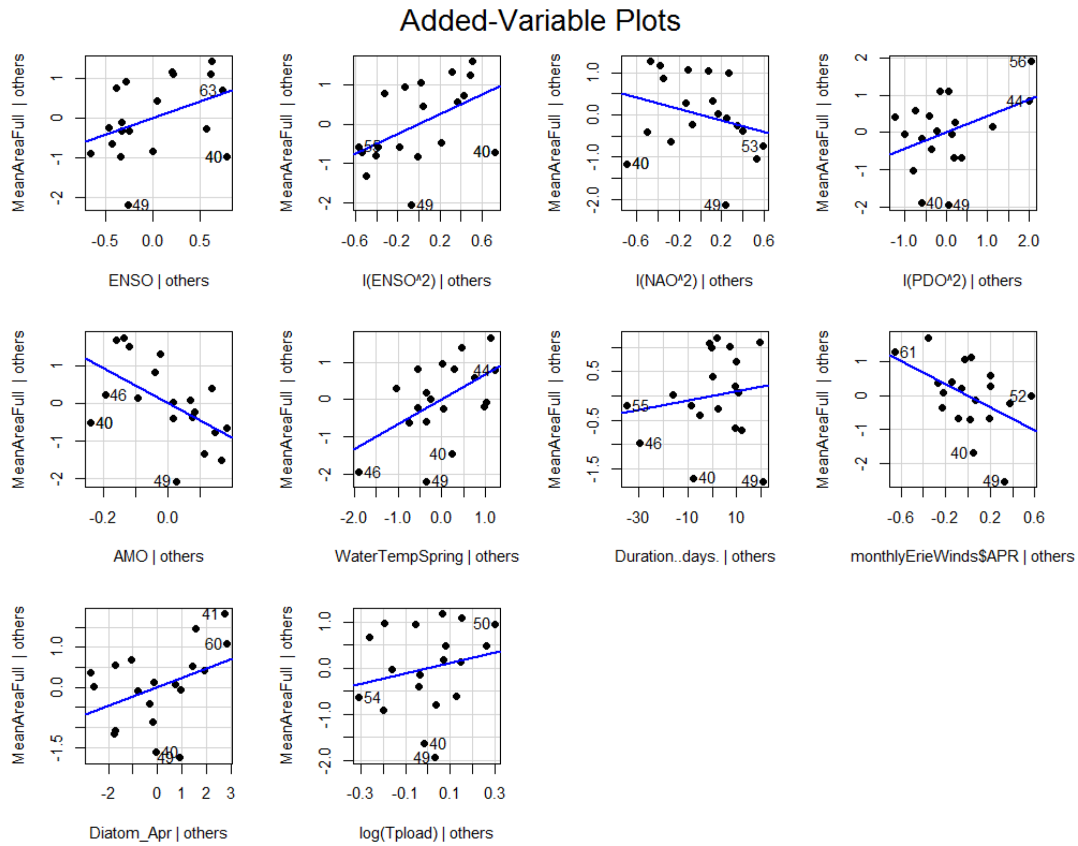

Figure 4-27. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

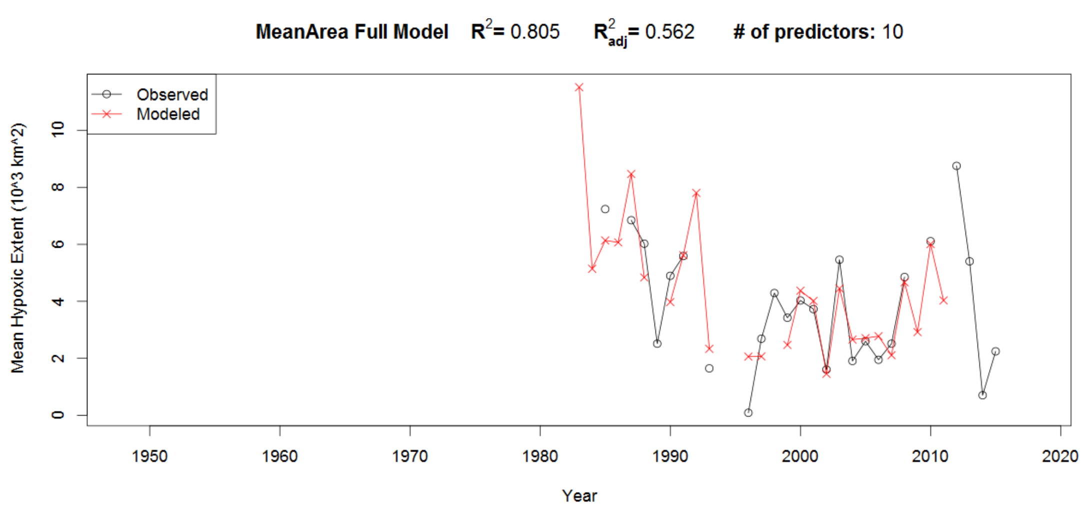

Figure 4-28. Time series plot of modeled vs. observed values (mean hypoxic area model).

Monthly Teleconnections Model (4 terms)

"eriePrecip_PMarAprRatio" is calculated as March precipitation in the previous year divided by April precipitation in the current year (precipitation over the Erie basin). This index was calculated based on the hypothesis that a combination of wet prior years (increases nutrient load to Lake Erie) and warm/dry current year (earlier onset of thermal stratification) is a major contributor to Lake Erie hypoxia. High values of eriePrecip_PMarAprRatio represent this combination of wet conditions in the prior year and dry conditions in the current year.

$$\text{MeanArea} = 2.51 + -3.79\text{AMO} + 0.78\text{ErieAirTemp\_MarAprMay} + 1.92\text{eriePrecip\_PMarAprRatio} + -0.87\text{UWindMay} + e \quad (4.5\text{-}2)$$

Table 4.5-4. Regression output for mean hypoxic area model.

	MeanAreaFull			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	2.51 ***	1.24 – 3.78	3.87	0.001
erieAirTemp_MarAprMay	0.78 **	0.33 – 1.23	3.42	0.003
eriePrecip_PMarAprRatio	1.92 **	0.81 – 3.03	3.38	0.003
AMO	-3.79 *	-6.96 – -0.62	-2.34	0.030
U-WindMay	-0.87	-1.76 – 0.01	-1.94	0.066
Observations	25			
R^2 / adjusted R^2	0.762 / 0.715			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.5-5. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	eriePrecip_PMarAprRatio	0.523	0.502	0.452	23.891
2	erieAirTemp_MarAprMay eriePrecip_PMarAprRatio	0.676	0.647	0.577	17.495
3	erieAirTemp_MarAprMay eriePrecip_PMarAprRatio AMO	0.717	0.677	0.607	17.000
4	erieAirTemp_MarAprMay eriePrecip_PMarAprRatio AMO UWindMay	0.762	0.715	0.622	16.642

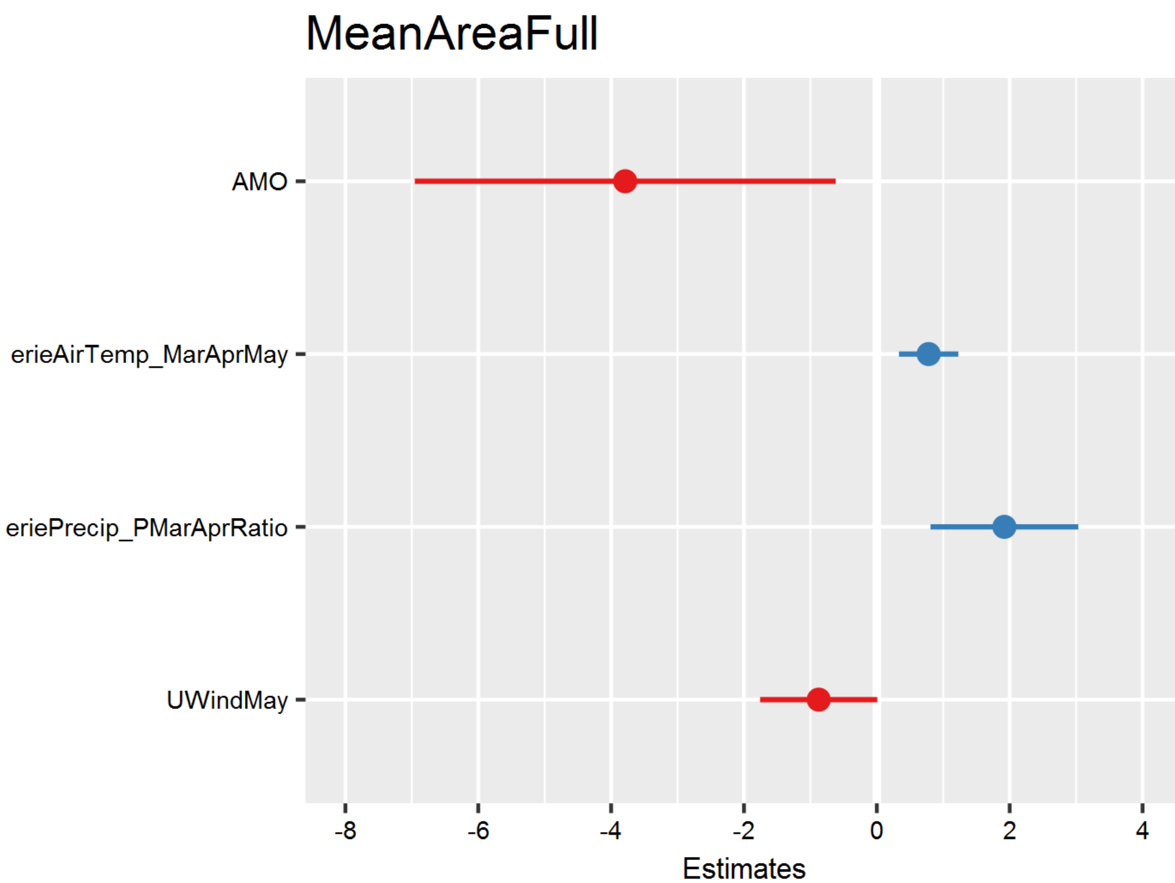

Figure 4-29. Regression coefficient plot (mean hypoxic extent model).

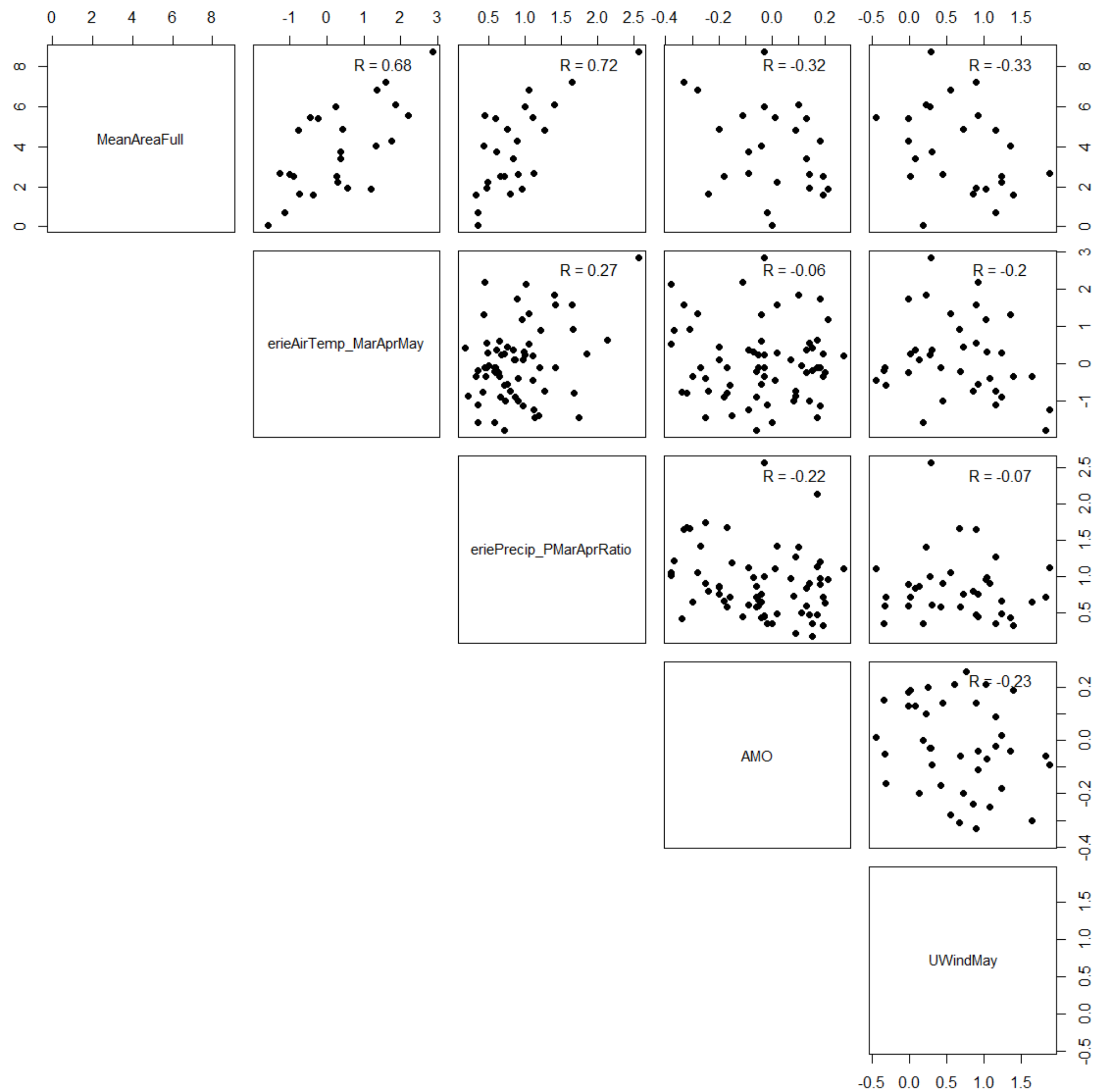

Figure 4-30. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings.

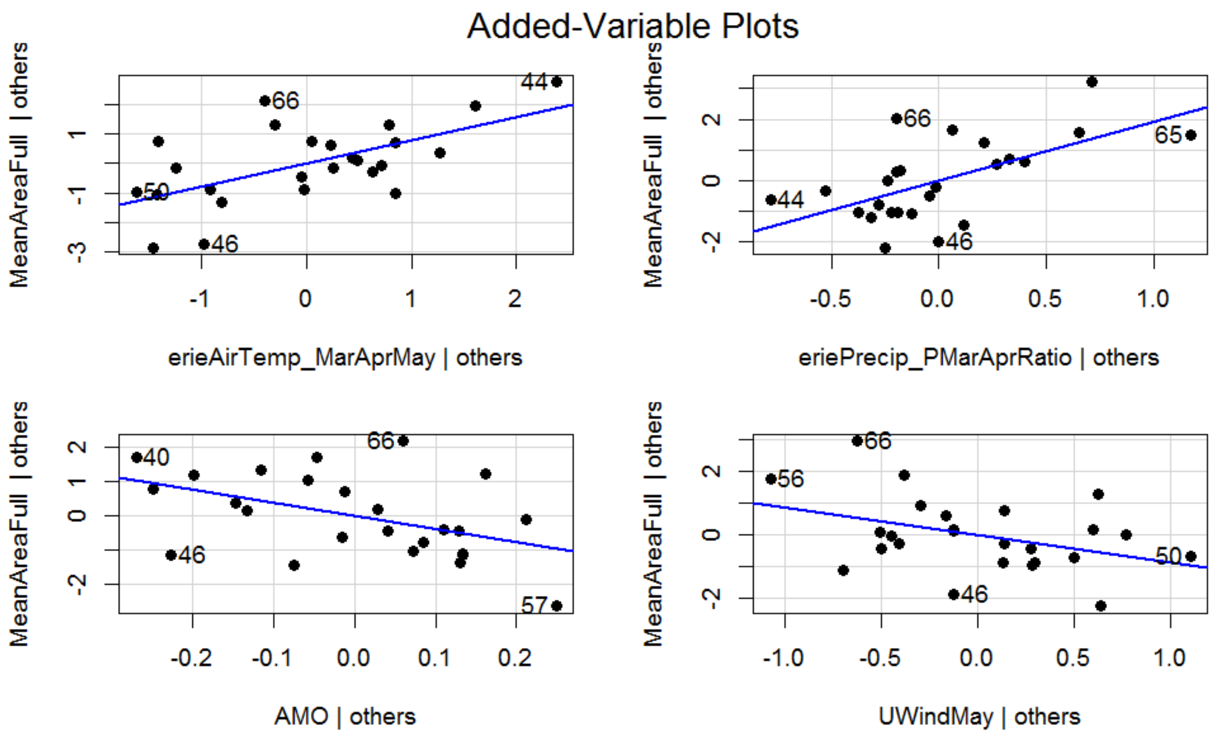

Figure 4-31. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

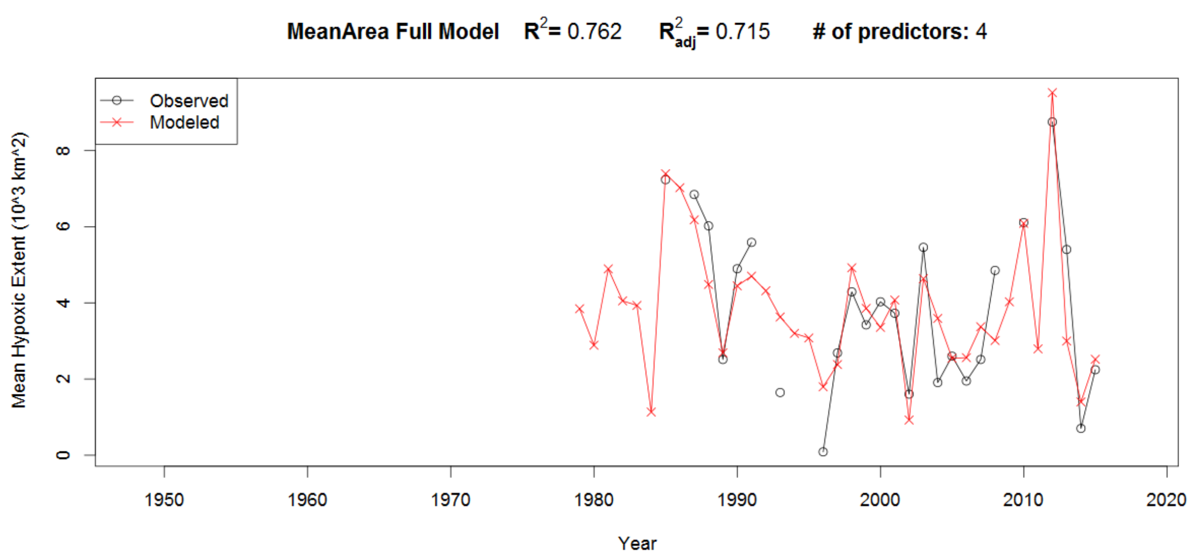

Figure 4-32. Time series plot of modeled vs. observed values (mean hypoxic area model).

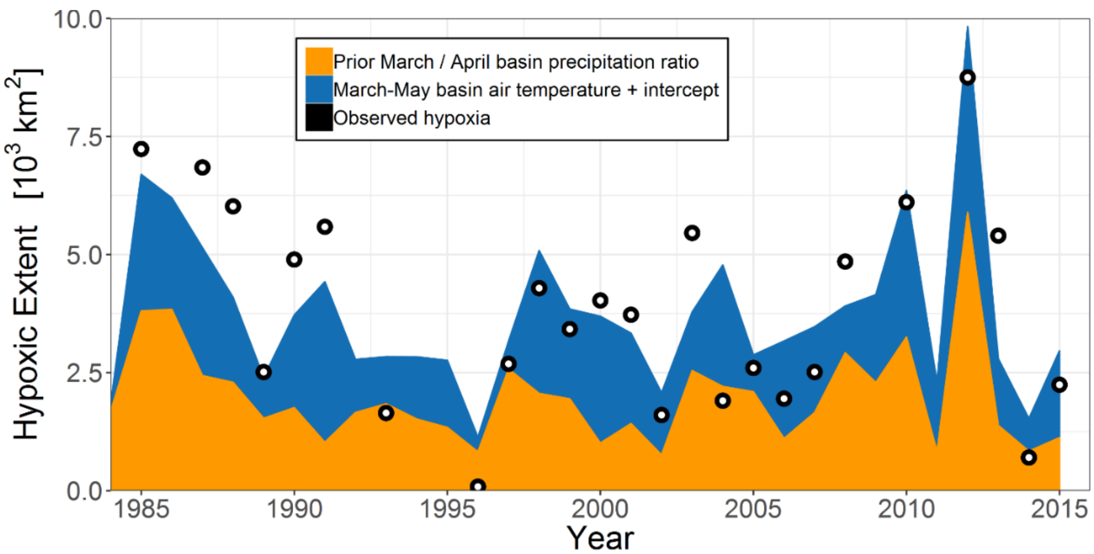

Figure 4-33. Time series plot showing the contribution of each variable to annual hypoxia prediction. The model used here is a subset of the above model, using only the ratio of PMar / Apr precipitation and March-May air temperature over the Lake Erie basin. This plot allows the user to visualize how hypoxia is related to each variable in different years.

Monthly Teleconnections Model (6 terms)

$$\text{MeanArea} = 3.77 + -3.76\text{AMO} + 0.23\text{PDO}^2$$
$$+ 0.50\text{erieAirTemp\_AprMay} + 2.17\text{eriePrecip\_PMarAprRatio} + -0.02\text{DurationDays}$$
$$+ -0.90\text{UWindMay} + e \qquad (4.5\text{-}3)$$

Table 4.5-6. Regression output for mean hypoxic area model.

	MeanAreaFull			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	3.77 **	1.51 – 6.03	3.27	0.005
erieAirTemp_AprMay	0.50	0.03 – 0.98	2.08	0.053
eriePrecip_PMarAprRatio	2.17 ***	1.13 – 3.21	4.10	0.001
AMO	-3.76 *	-7.05 – -0.46	-2.23	0.039
PDO^2	0.23	-0.09 – 0.55	1.42	0.174
Duration..days.	-0.02 *	-0.04 – -0.00	-2.21	0.041
UWindMay	-0.90	-1.80 – 0.01	-1.95	0.068
Observations	24			
R^2 / adjusted R^2	0.811 / 0.745			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.5-7. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	eriePrecip_PMarAprRatio	0.523	0.502	0.452	23.660
2	erieAirTemp_AprMay eriePrecip_PMarAprRatio	0.661	0.630	0.590	17.990
3	erieAirTemp_AprMay eriePrecip_PMarAprRatio PDO^2	0.705	0.663	0.523	17.195
4	erieAirTemp_AprMay eriePrecip_PMarAprRatio AMO Duration..days.	0.745	0.692	0.619	17.247
5	erieAirTemp_AprMay eriePrecip_PMarAprRatio AMO Duration..days. UWindMay	0.789	0.730	0.641	17.302
6	erieAirTemp_AprMay eriePrecip_PMarAprRatio AMO PDO^2 Duration..days. UWindMay	0.811	0.745	0.626	19.033

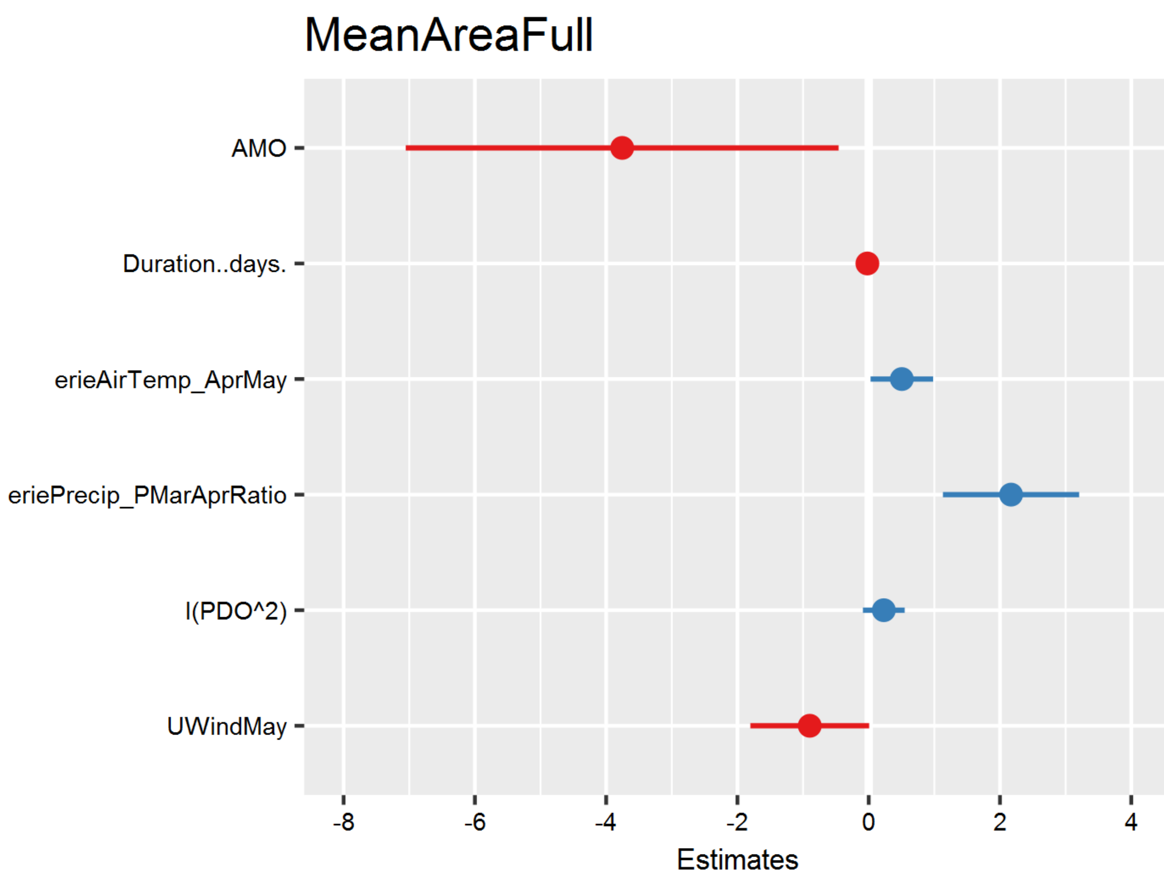

Figure 4-34. Regression coefficient plot (mean hypoxic extent model).

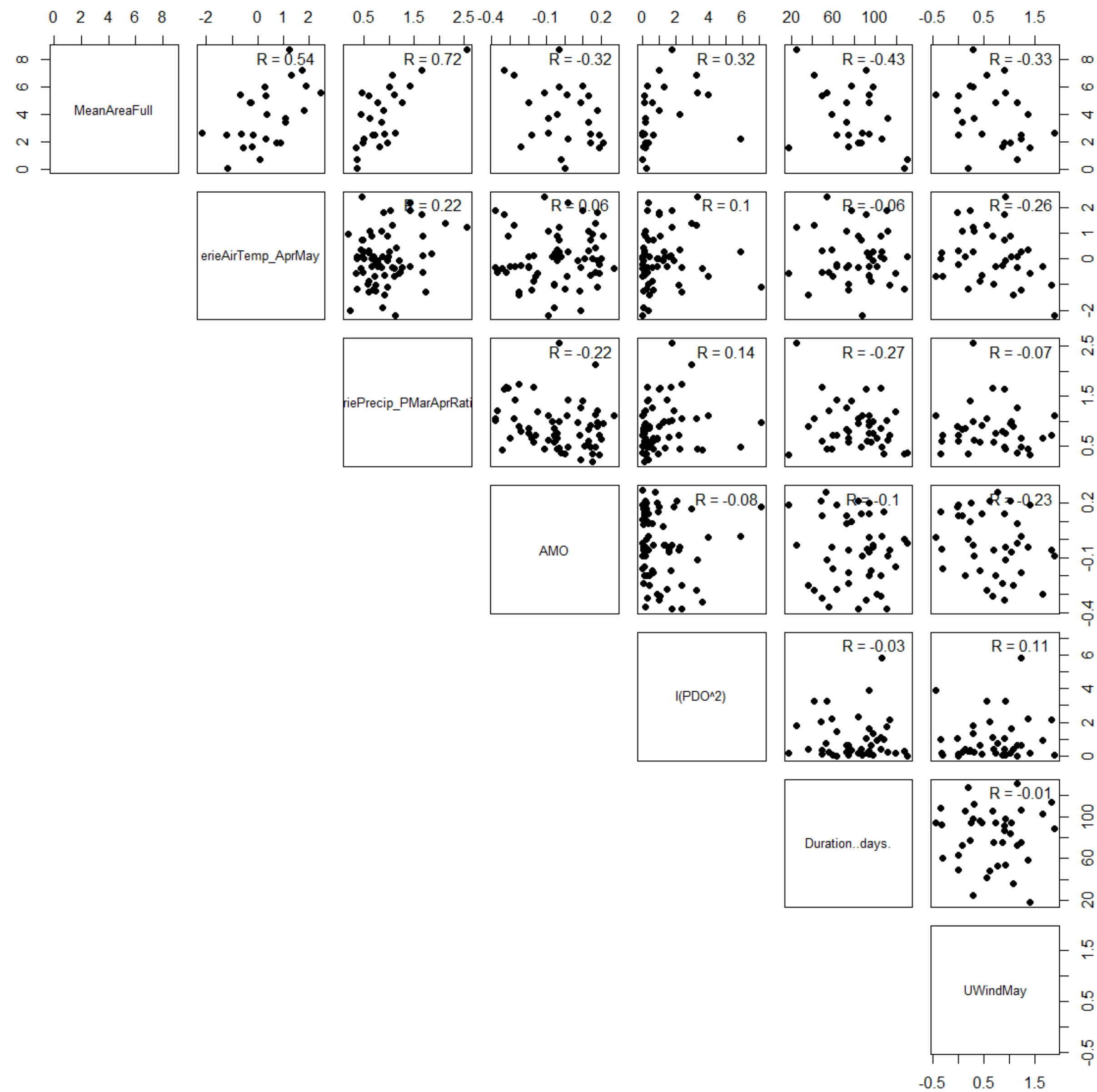

Figure 4-35. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings.

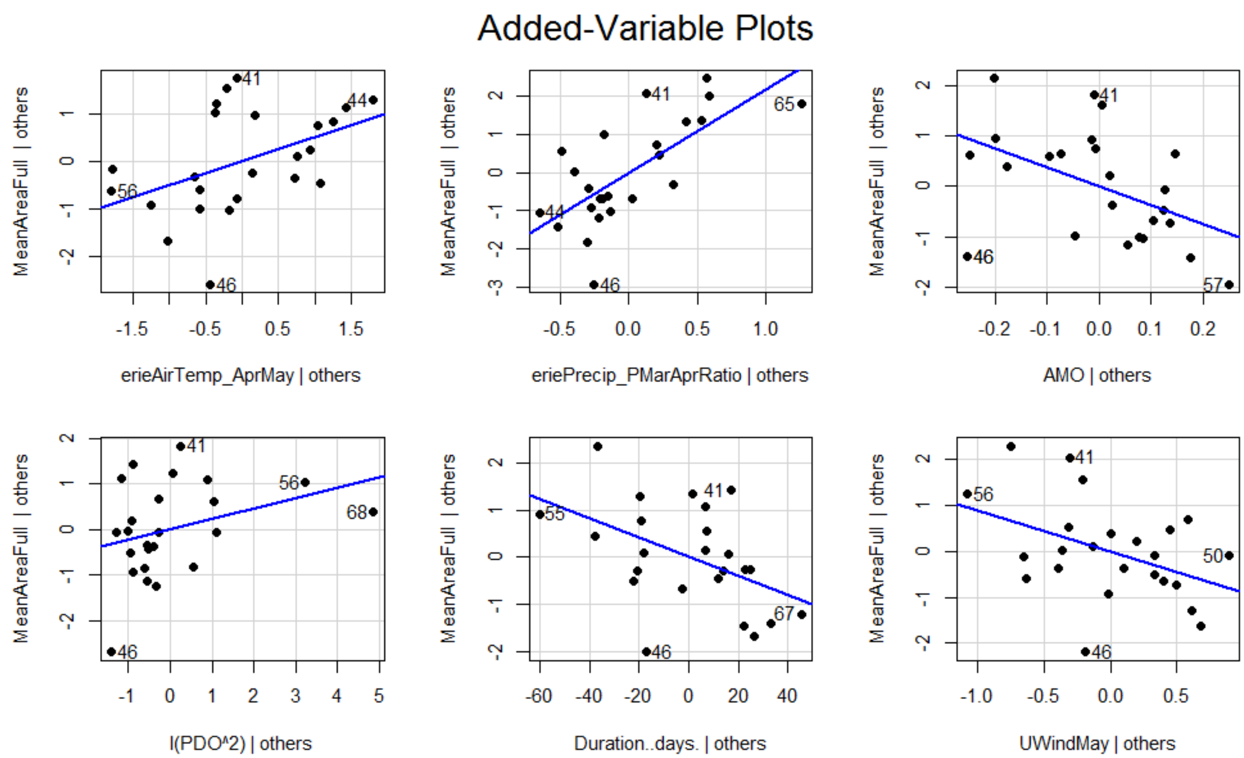

Figure 4-36. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

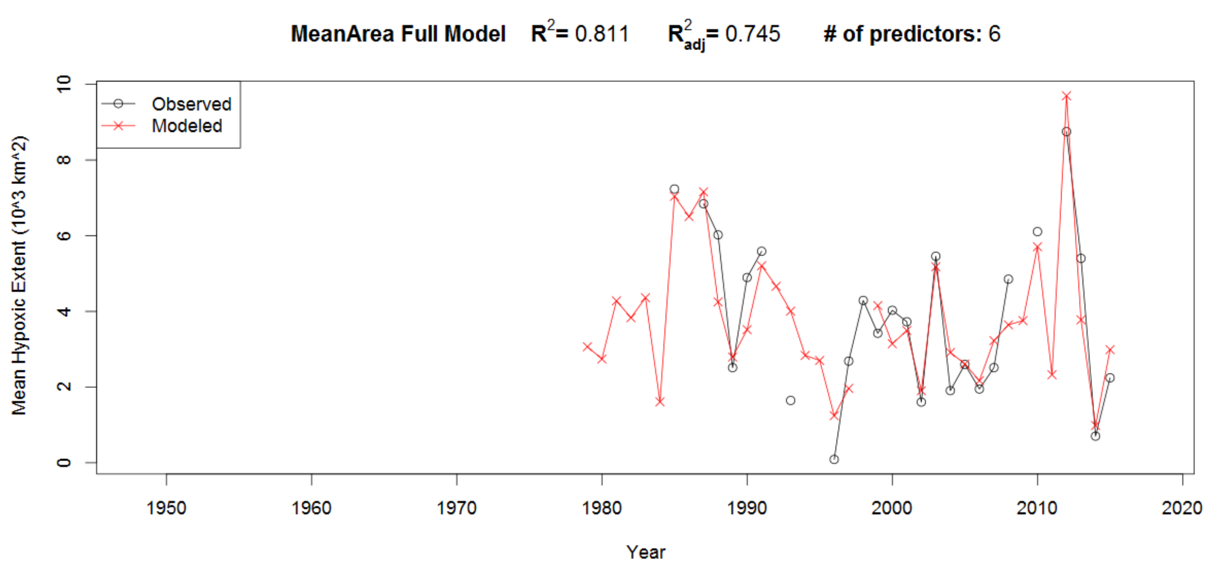

Figure 4-37. Time series plot of modeled vs. observed values (mean hypoxic area model).

Monthly Teleconnections Model (3 terms)

$$\text{MeanArea} = -4.20 + 0.11\text{erie7yrPrecipMar} + -0.04\text{eriePrecip\_BasinAprMay} + 0.86\text{erieBasinAirTempMarApr} + e \quad (4.5\text{-}4)$$

Table 4.5-8. Regression output for mean hypoxic area model.

	MeanAreaFull			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	-4.20	-8.26 – -0.14	-2.03	0.056
erie7yrPrecipMar	0.11 ***	0.06 – 0.15	4.28	<0.001
eriePrecip_BasinAprMay	-0.04 **	-0.06 – -0.01	-3.21	0.004
erieBasinAirTempMarApr	0.86 ***	0.56 – 1.16	5.61	<0.001
Observations	24			
R^2 / adjusted R^2	0.804 / 0.774			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.5-9. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	erieBasinAirTempMarApr	0.472	0.448	0.381	24.899
2	erie7yrPrecipMar erieBasinAirTempMarApr	0.702	0.674	0.574	14.628
3	erie7yrPrecipMar eriePrecip_BasinAprMay erieBasinAirTempMarApr	0.804	0.774	0.682	9.501

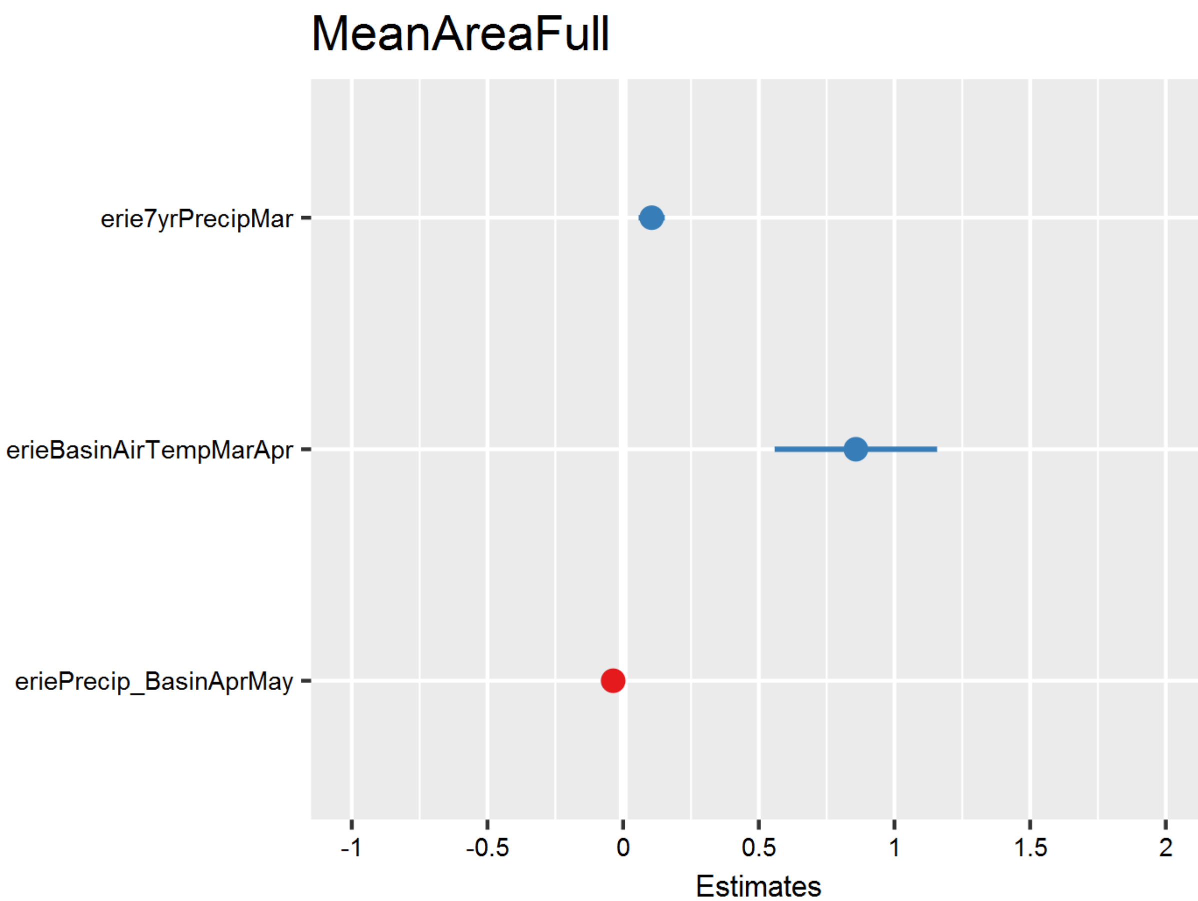

Figure 4-38. Regression coefficient plot (mean hypoxic extent model).

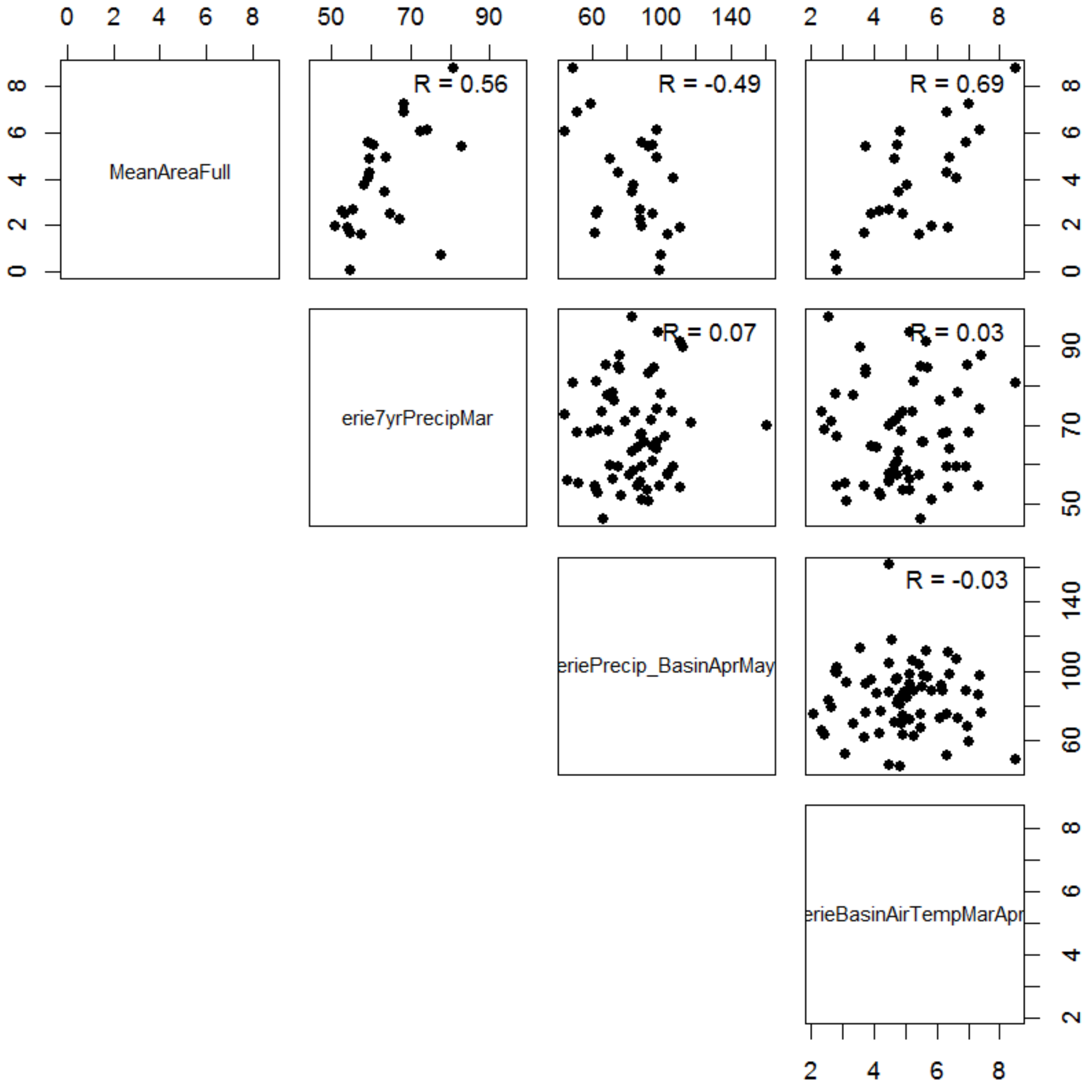

Figure 4-39. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings.

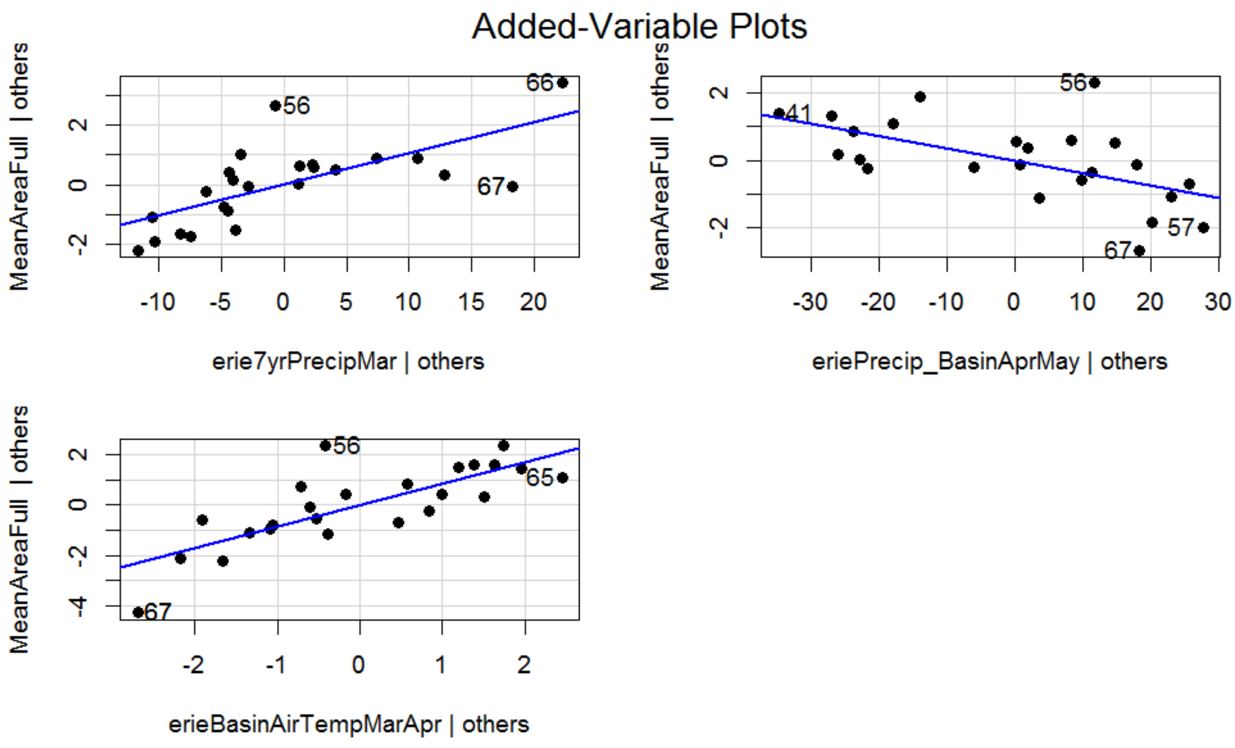

Figure 4-40. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

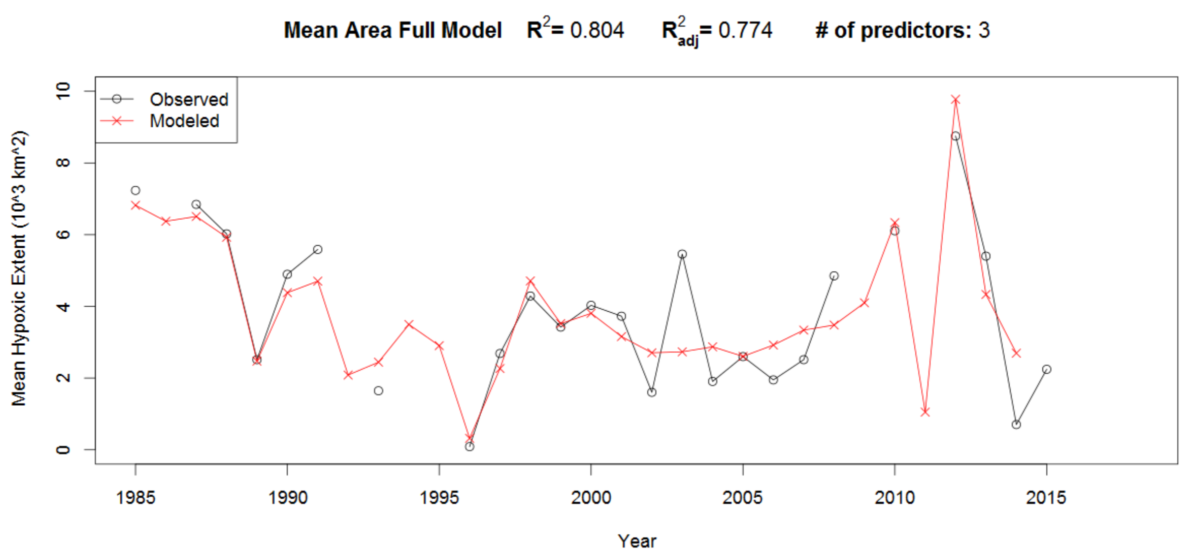

Figure 4-41. Time series plot of modeled vs. observed values (mean hypoxic area model).

Monthly Teleconnections Model (2 terms)

Here, "erie7yrMarPrecip_AprMayPrecip_Difference" was used as a predictor. This variable was calculated by taking the difference between the prior 7-year average of March precipitation and the current year Apr/May average precipitation over the Lake Erie basin.

$$MeanArea = 0.14 + 0.05 erie7yrMarPrecip\_AprMayPrecip\_Difference + 0.90\ erieAirTempMarApr + e \quad (4.5\text{-}5)$$

Table 4.5-10. Regression output for mean hypoxia area model.

	MeanAreaFull			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	0.14	-1.75 – 2.02	0.14	0.890
erie7yrMarPrecip_AprMayPrecip_Difference	0.05 ***	0.03 – 0.07	4.84	<0.001
erieAirTempMarApr	0.90 ***	0.57 – 1.22	5.42	<0.001
Observations	25			
R^2 / adjusted R^2	0.749 / 0.726			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.5-11. Table summarizing the best subsets procedure for the mean area model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	erieAirTempMarApr	0.480	0.458	0.395	25.629
2	erie7yrMarPrecip_AprMayPrecip_Difference erieAirTempMarApr	0.749	0.726	0.691	12.467

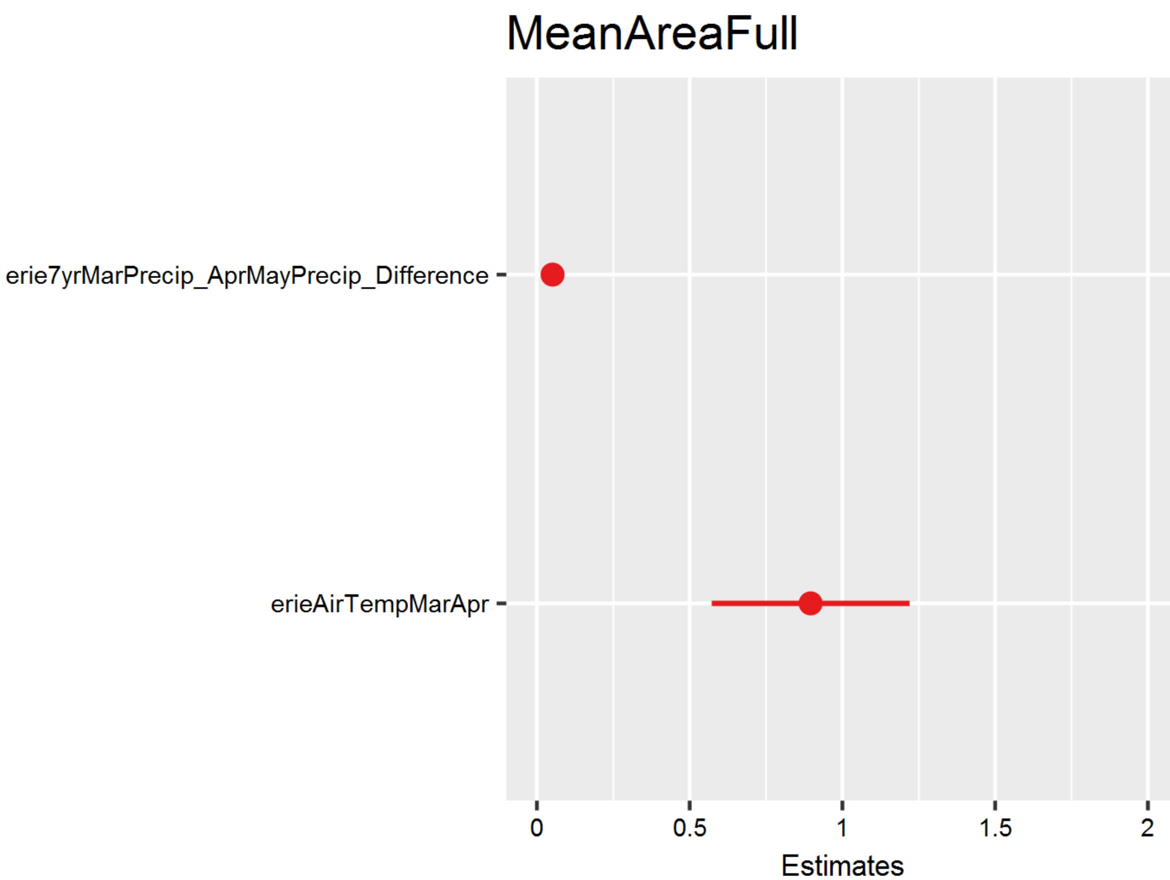

Figure 4-42. Regression coefficient plot (mean hypoxic extent model).

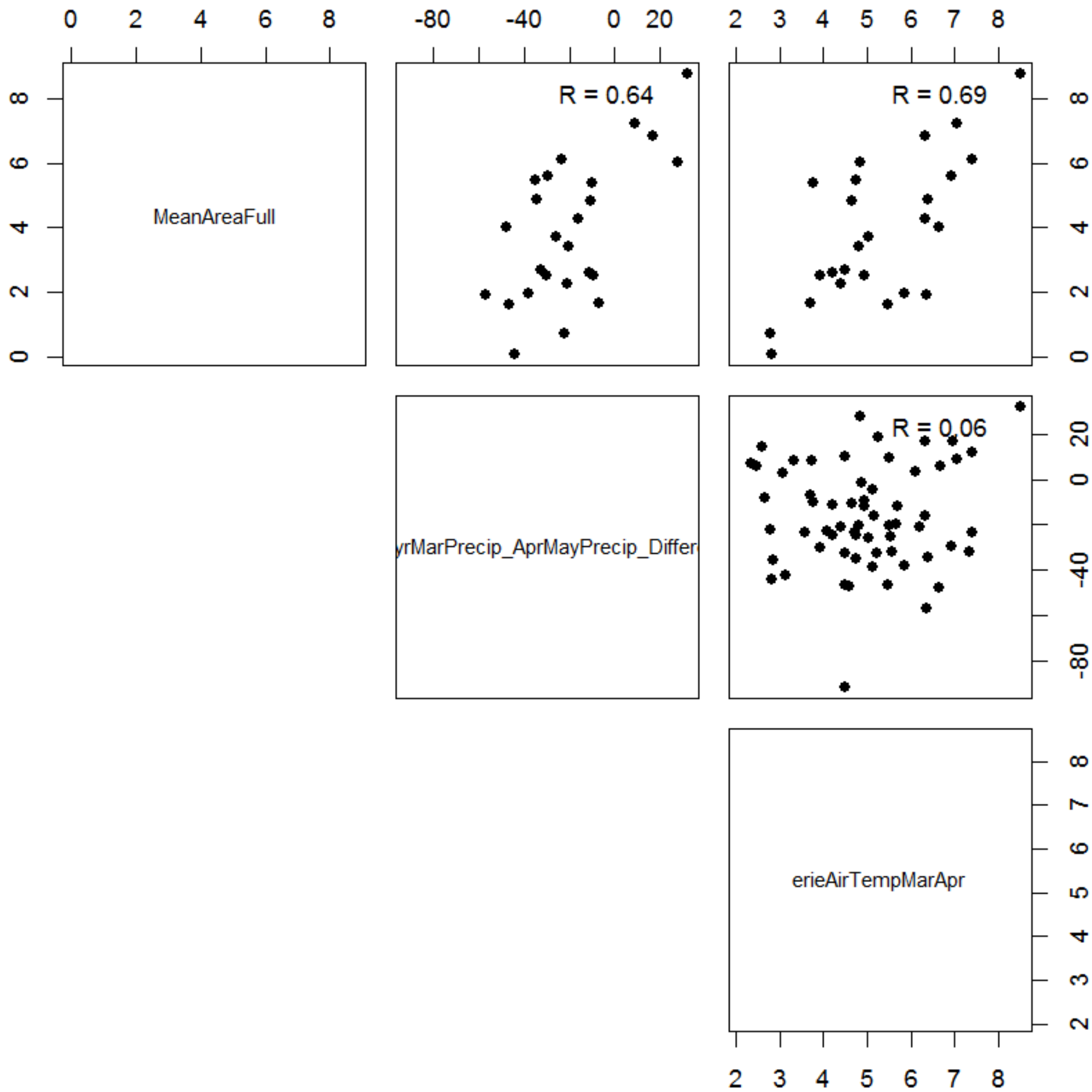

Figure 4-43. Linear correlations between mean hypoxic extent, biological parameters, and physical forcings.

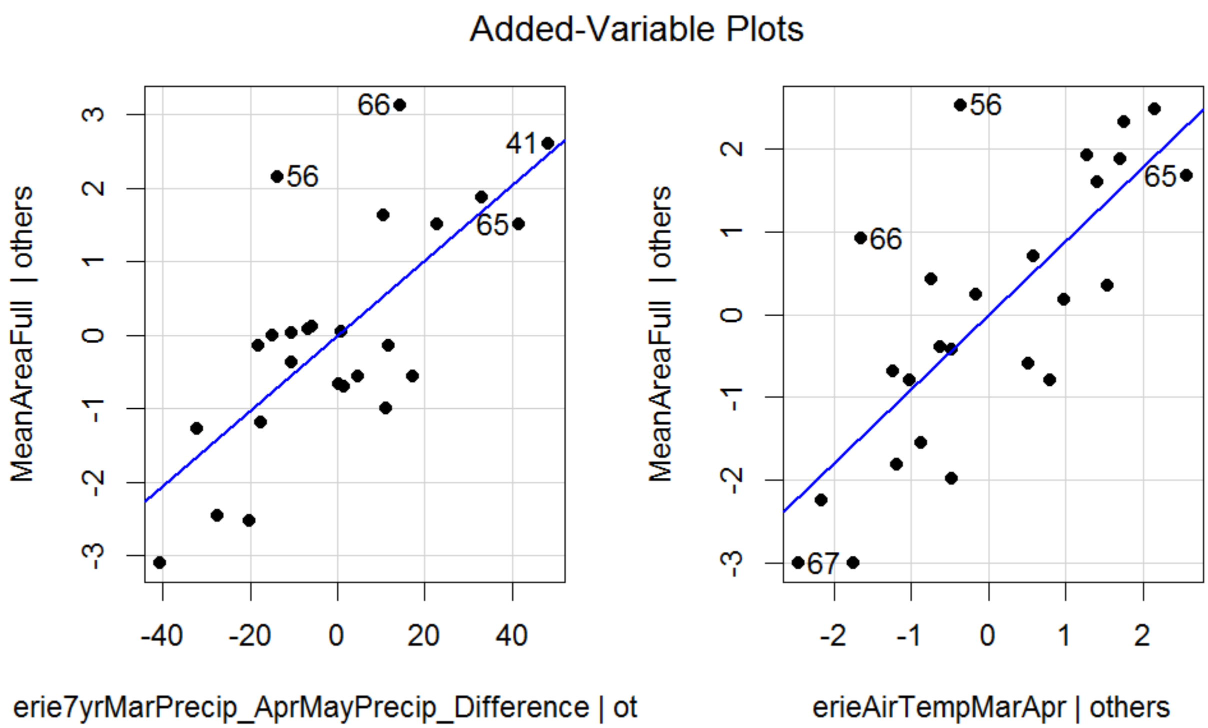

Figure 4-44. Added variable plot for mean hypoxic extent model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

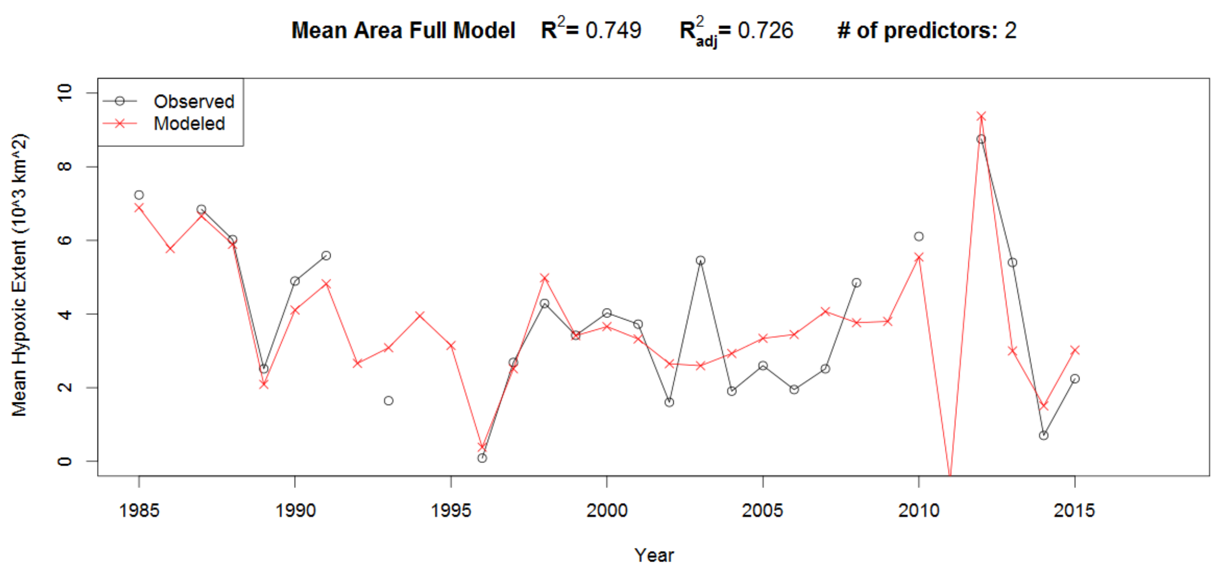

Figure 4-45. Time series plot of modeled vs. observed values (mean hypoxic area model).

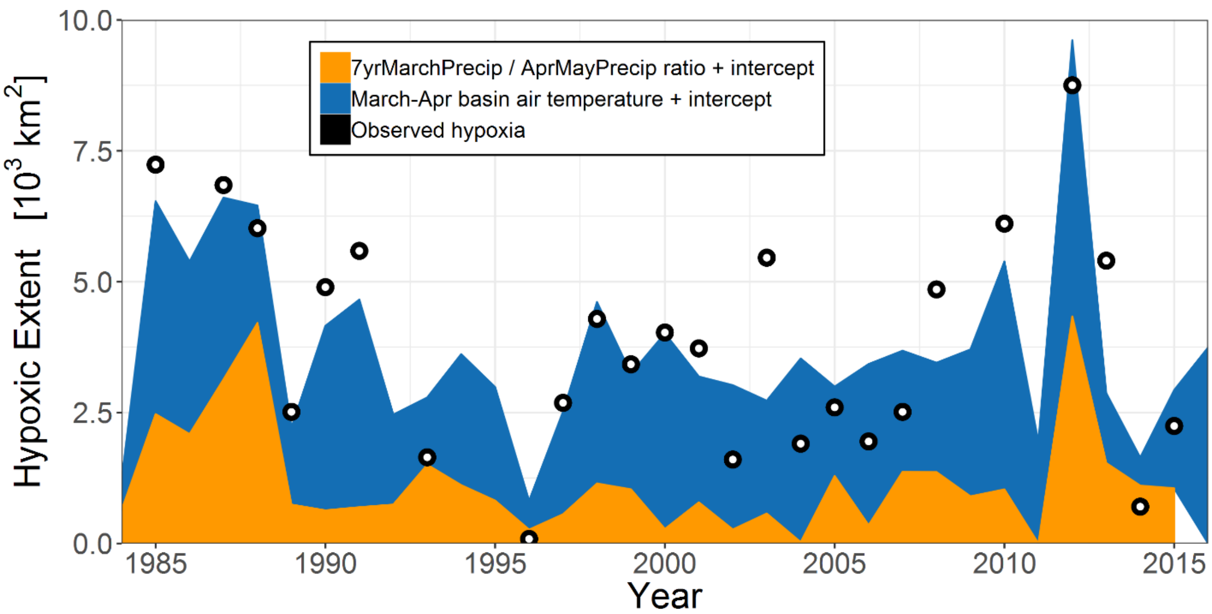

Figure 4-46. Time series plot showing the contribution of each variable to annual hypoxia prediction. The model used here is very similar to the above model, but instead using the ratio of 7yrMarchPrecip / AprMayPrecip rather than a difference. This change was made because it was very difficult to make this style of plot for the above difference term model, simply because the intercept and terms did not add together nicely for visualization.

4.6 MEAN DISSOLVED OXYGEN (DO) MODELS

Table 4.6-1. Correlations and p-values of mean dissolved oxygen with teleconnection patterns. Significant correlations shown in boldface.

Index	r	p value	Significance (%)
ENSO	-0.157	0.433	56.7
ENSO^2	-0.028	0.891	10.9
NAO	0.001	0.997	0.3
NAO^2	-0.094	0.642	35.8
AMO	0.413	0.032	96.8
AMO^2	-0.432	0.024	97.6
PDO	-0.024	0.904	9.6
PDO^2	-0.345	0.078	92.2

Winter Teleconnections Model

$$MeanDO = -9.34 + -0.46PDO^2 + 3.66AMO$$
$$+ -0.45WaterTempSpring + -0.01DurationDays + -0.81UWindApr$$
$$+ -0.17DiatomApr + 1.89\log(Tpload) + e \quad (4.6\text{-}1)$$

Table 4.6-2. Regression output for mean dissolved oxygen model.

	MeanDO			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	-9.34	-33.09 – 14.41	-0.77	0.452
PDO^2	-0.46	-1.07 – 0.15	-1.48	0.158
AMO	3.66	-0.15 – 7.48	1.88	0.078
WaterTempSpring	-0.45	-0.94 – 0.05	-1.77	0.097
Duration..days.	-0.01	-0.04 – 0.02	-0.81	0.431
UWindApr	-0.81	-1.67 – 0.05	-1.84	0.085
Diatom_Apr	-0.17	-0.44 – 0.11	-1.18	0.254
log(Tpload)	1.89	-0.60 – 4.38	1.49	0.156
Observations	24			
R^2 / adjusted R^2	0.567 / 0.377			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.6-3. Table summarizing the best subsets procedure for the mean dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	AMO	0.171	0.137	0.063	26.357
2	AMO WaterTempSpring	0.353	0.299	0.197	22.534
3	AMO WaterTempSpring Diatom_Apr	0.446	0.370	0.207	21.674
4	PDO^2 WaterTempSpring UWindApr Diatom_Apr	0.468	0.367	0.154	23.200
5	PDO^2 AMO WaterTempSpring UWindApr Diatom_Apr	0.525	0.406	0.146	24.006
6	PDO^2 AMO WaterTempSpring Duration..days. UWindApr Diatom_Apr	0.534	0.379	0.059	26.386
7	PDO^2 AMO WaterTempSpring Duration..days. UWindApr Diatom_Apr log(Tpload)	0.567	0.377	0.043	27.739

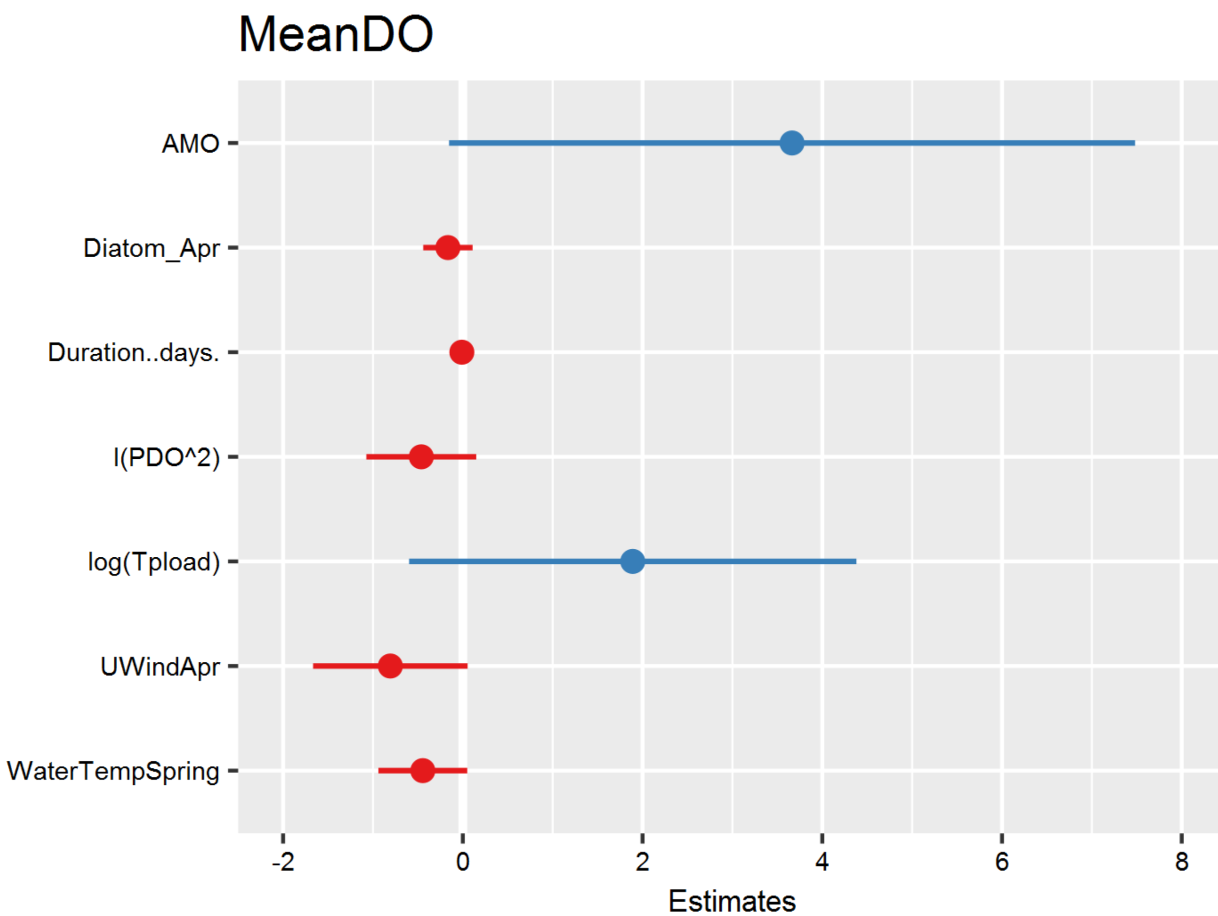

Figure 4-47. Regression coefficient plot (mean dissolved oxygen model).

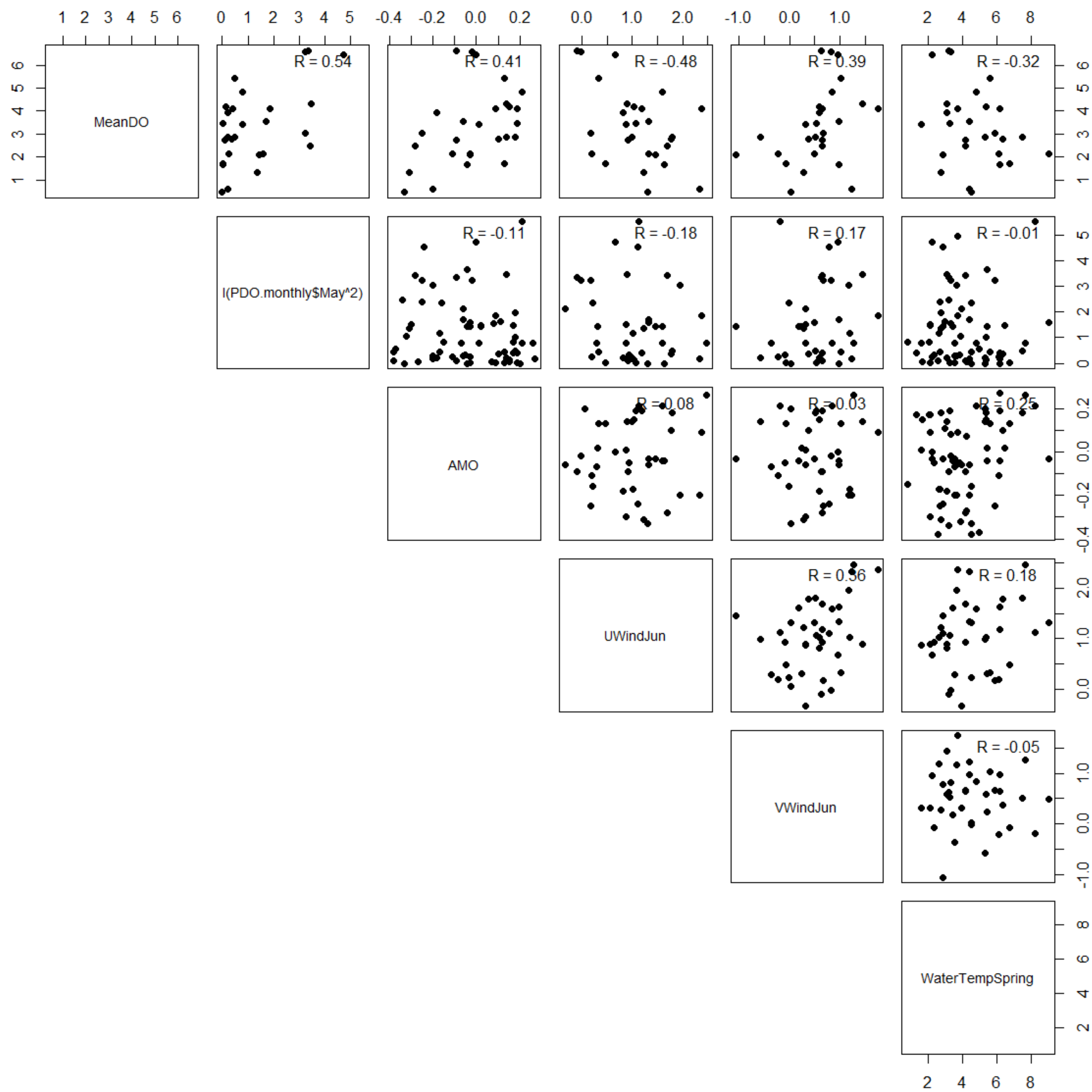

Figure 4-48. Linear correlations between mean dissolved oxygen, biological parameters, and physical forcings.

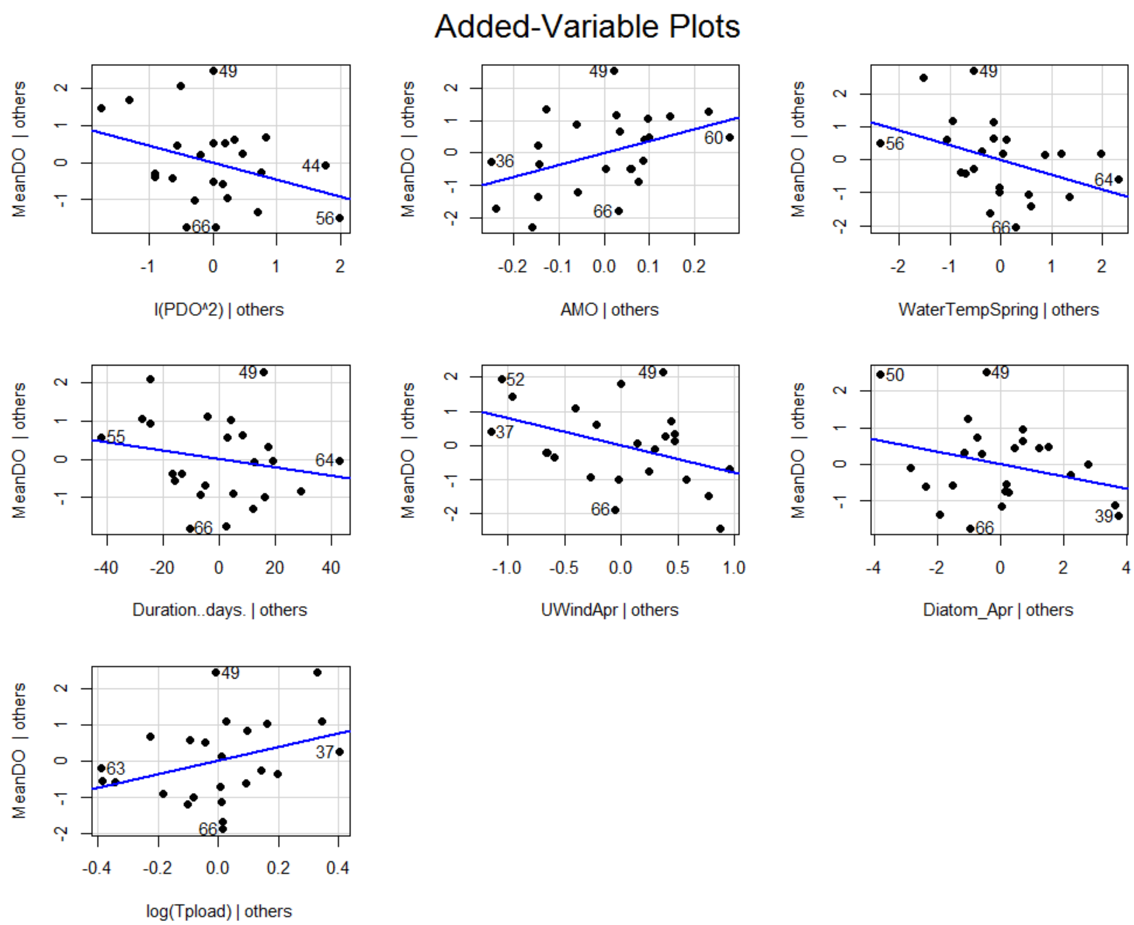

Figure 4-49. Added variable plot for mean dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

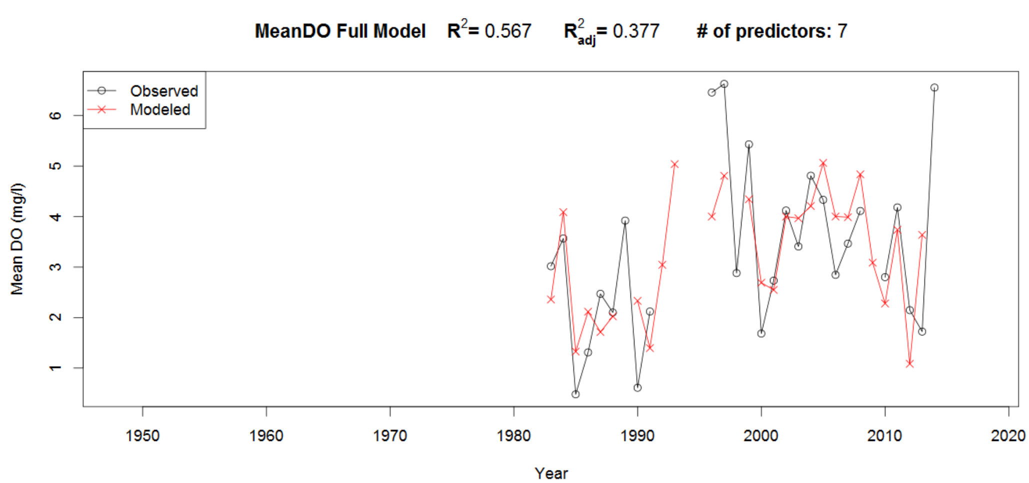

Figure 4-50. Time series plot of modeled vs. observed values (mean dissolved oxygen model).

Monthly Teleconnections Model (including WaterTempSpring)

$$MeanDO = 4.93 + 0.53PDO.monthlyMay^2 + 5.3AMO + -0.91UWindJun + 0.56VWindJun + -0.23WaterTempSpring + -0.12Diatom\_Apr + e \quad (4.6\text{-}2)$$

Table 4.6-4. Regression output for mean dissolved oxygen model. UWindJun and VwindJun represent the u- and v-component of June wind over the Erie basin. This data came from the North American Regional Reanalysis (NARR) and was used because there was no missing data, unlike the wind data from Zhang et al. (2018).

	MeanDO			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	4.93 ***	3.46 – 6.40	6.58	<0.001
PDO.monthlyMay^2	0.53 **	0.26 – 0.80	3.82	0.001
AMO	5.30 ***	3.30 – 7.30	5.18	<0.001
UWindJun	-0.91 **	-1.43 – -0.39	-3.42	0.003
VWindJun	0.56	-0.05 – 1.17	1.79	0.089
WaterTempSpring	-0.23 *	-0.42 – -0.03	-2.30	0.033
Diatom_Apr	-0.12	-0.28 – 0.03	-1.58	0.130
Observations	26			
R^2 / adjusted R^2	0.850 / 0.802			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.6-5. Table summarizing the best subsets procedure for the mean dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	**Predictors**	**Rsquare**	**AdjRsq**	**PredRsq**	**BIC**
1	PDO.monthlyMay^2	0.289	0.261	0.162	19.558
2	PDO.monthlyMay^2 AMO	0.579	0.544	0.467	7.748
3	PDO.monthlyMay^2 AMO Diatom_Apr	0.704	0.663	0.579	2.189
4	PDO.monthlyMay^2 AMO UWindJun Diatom_Apr	0.776	0.733	0.664	-1.290
5	PDO.monthlyMay^2 AMO UWindJun WaterTempSpring Diatom_Apr	0.824	0.780	0.728	-2.795
6	PDO.monthlyMay^2 AMO UWindJun VWindJun WaterTempSpring Diatom_Apr	0.850	0.802	0.741	-1.959

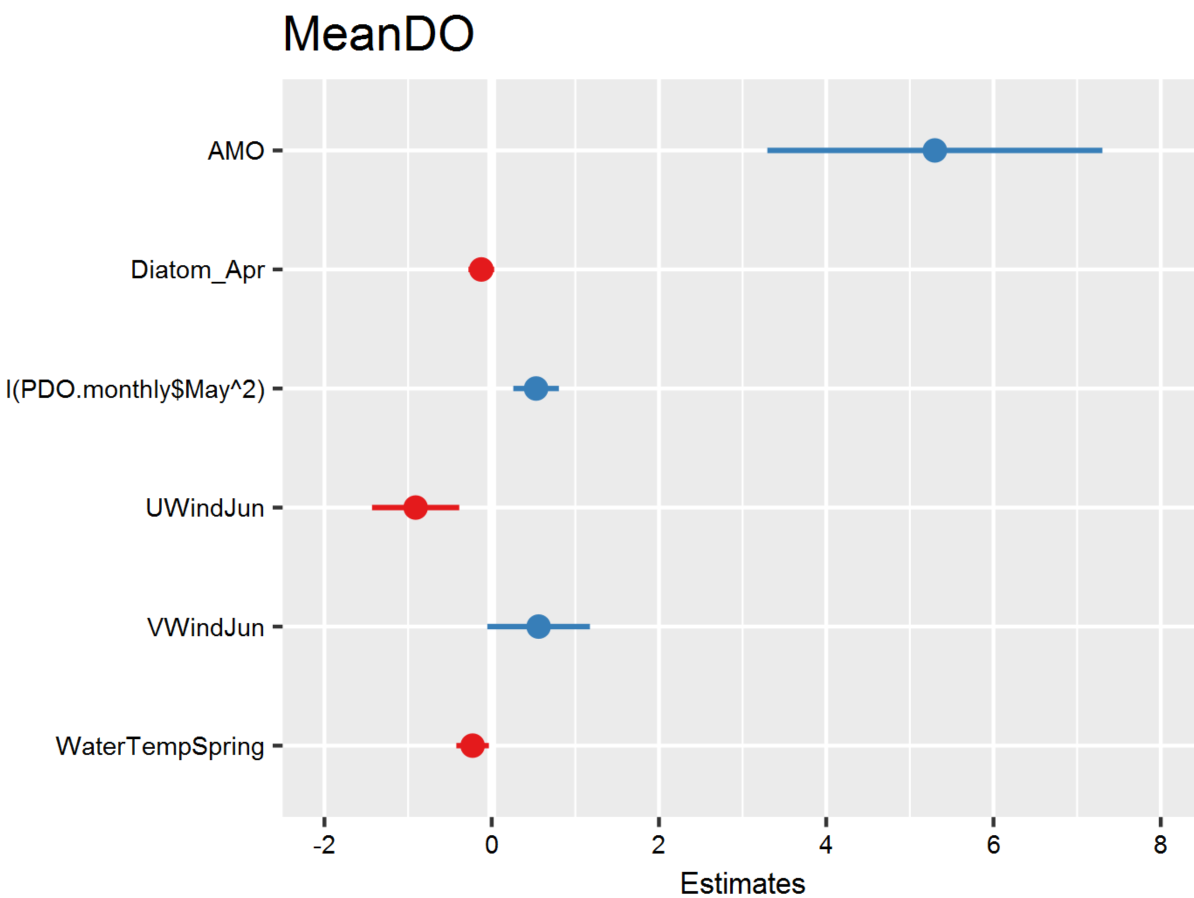

Figure 4-51. Regression coefficient plot (mean dissolved oxygen model).

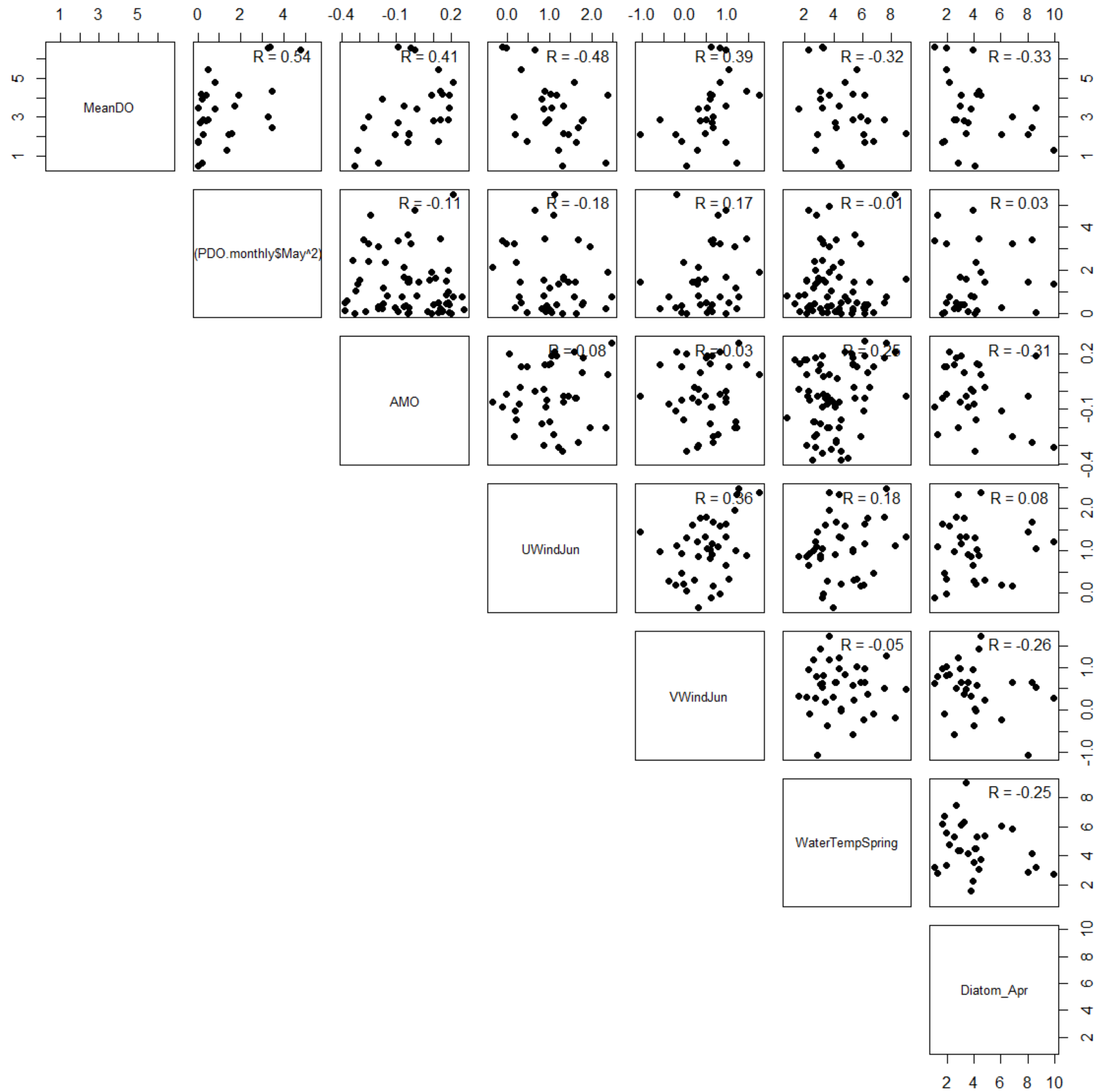

Figure 4-52. Linear correlations between mean dissolved oxygen, biological parameters, and physical forcings.

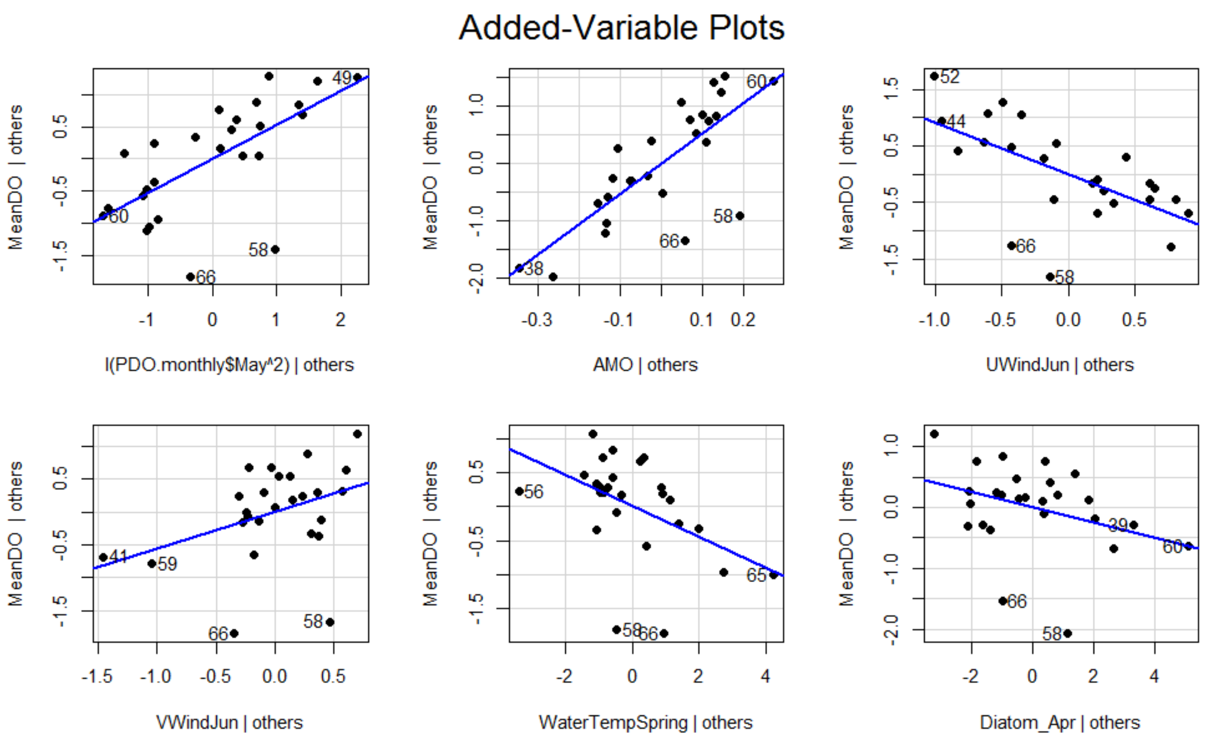

Figure 4-53. Added variable plot for mean dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

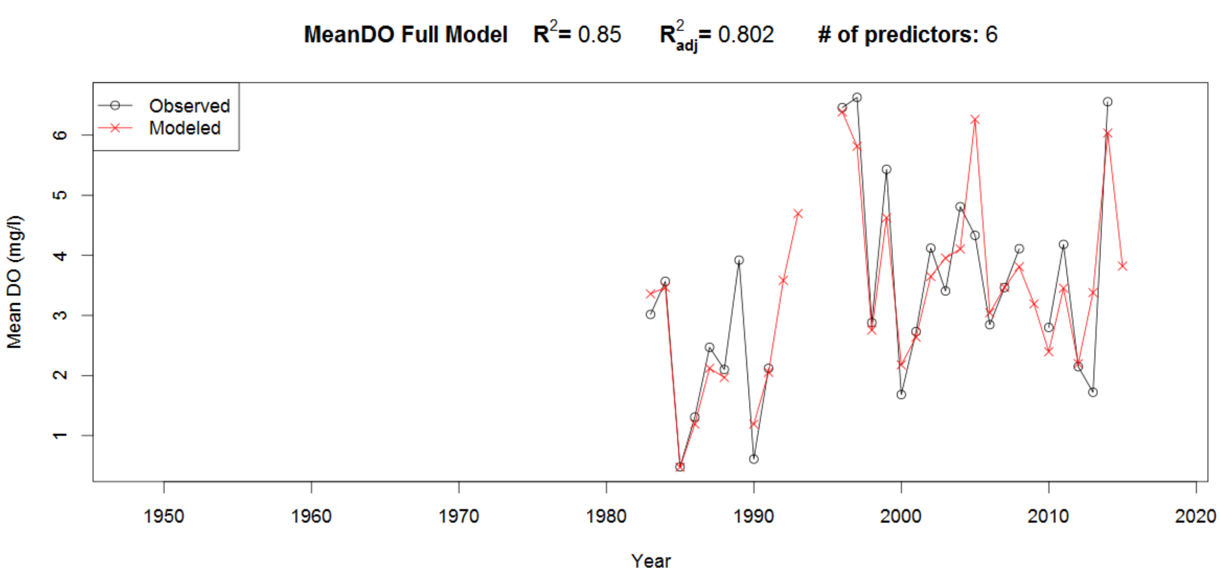

Figure 4-54. Time series plot of modeled vs. observed values (mean dissolved oxygen model).

Monthly Teleconnections Model (including April precipitation)

$$MeanDO = 4.80 + 0.44PDO.monthlyMay^2 + 5.88AMO$$

$$+ -1.42UWindJun + 0.83VWindJun + -2.69NARR\_TotalPrecipApr + e \qquad (4.6\text{-}3)$$

Table 4.6-6. Regression output for mean dissolved oxygen model. NARR_TotalPrecipApr is total April precipitation over the Lake Erie basin and comes from the North American Regional Reanalysis (NARR).

	MeanDO			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	4.80 ***	3.80 – 5.80	9.40	<0.001
PDO.monthlyMay^2	0.44 **	0.20 – 0.68	3.62	0.002
AMO	5.88 ***	4.04 – 7.72	6.25	<0.001
UWindJun	-1.42 ***	-1.92 – -0.91	-5.52	<0.001
VWindJun	0.83 **	0.29 – 1.37	2.99	0.007
NARR_TotalPrecipApr	-2.69 **	-4.26 – -1.13	-3.38	0.003
Observations	27			
R^2 / adjusted R^2	0.844 / 0.807			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.6-7. Table summarizing the best subsets procedure for the mean dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	PDO.monthlyMay^2	0.289	0.261	0.162	19.438
2	PDO.monthlyMay^2 AMO	0.579	0.544	0.467	7.511
3	PDO.monthlyMay^2 AMO UWindJun	0.691	0.651	0.567	1.804
4	PDO.monthlyMay^2 AMO UWindJun NARR_TotalPrecipApr	0.778	0.738	0.677	-3.092
5	PDO.monthlyMay^2 AMO UWindJun VWindJun NARR_TotalPrecipApr	0.844	0.807	0.743	-6.399

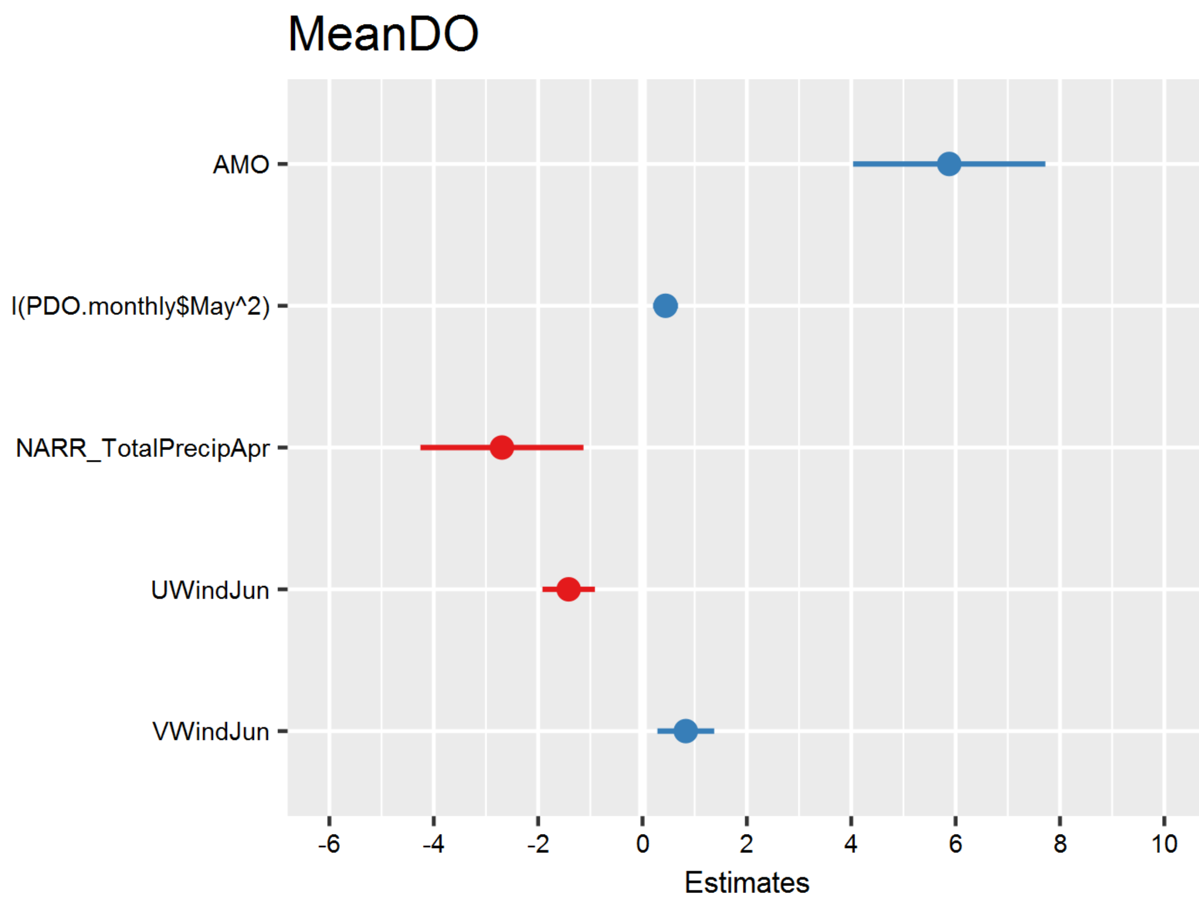

Figure 4-55. Regression coefficient plot (mean dissolved oxygen model).

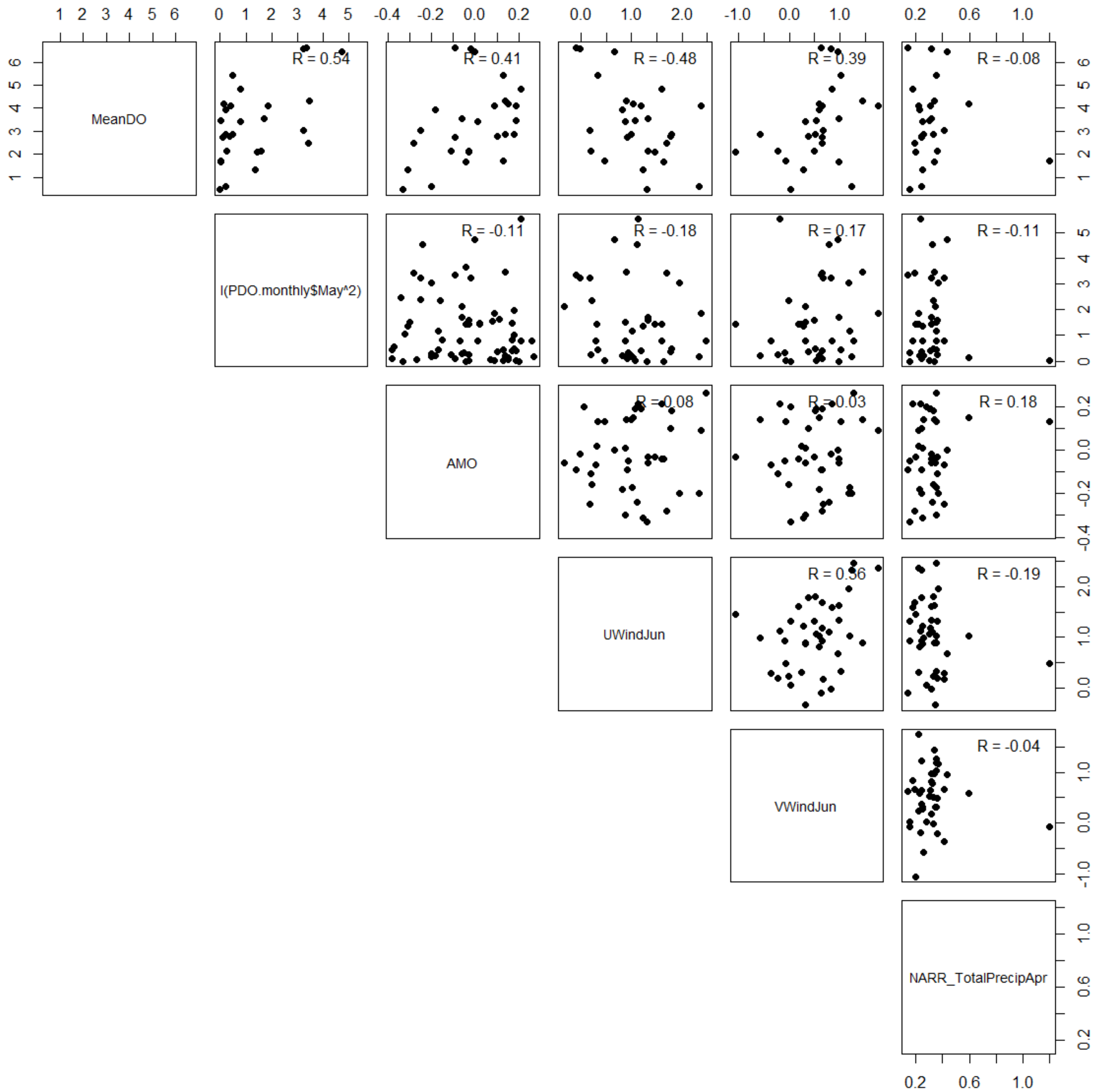

Figure 4-56. Linear correlations between mean dissolved oxygen, biological parameters, and physical forcings.

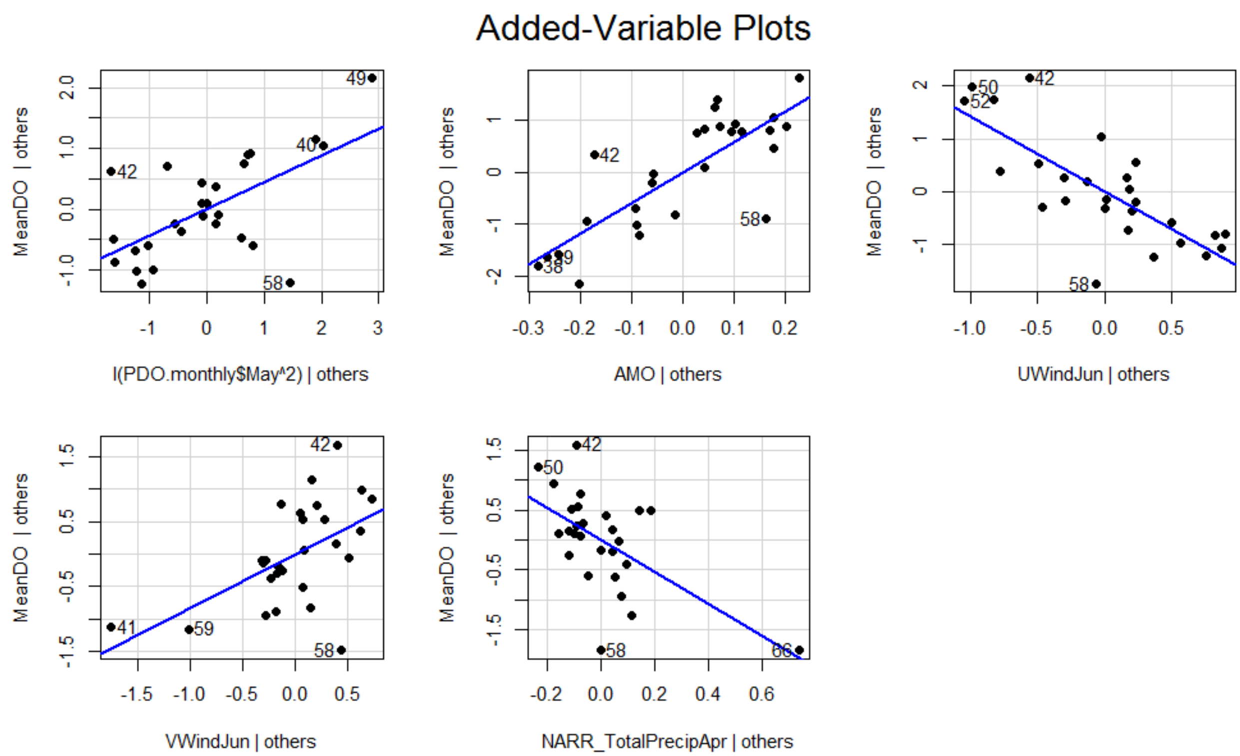

Figure 4-57. Added variable plot for mean dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

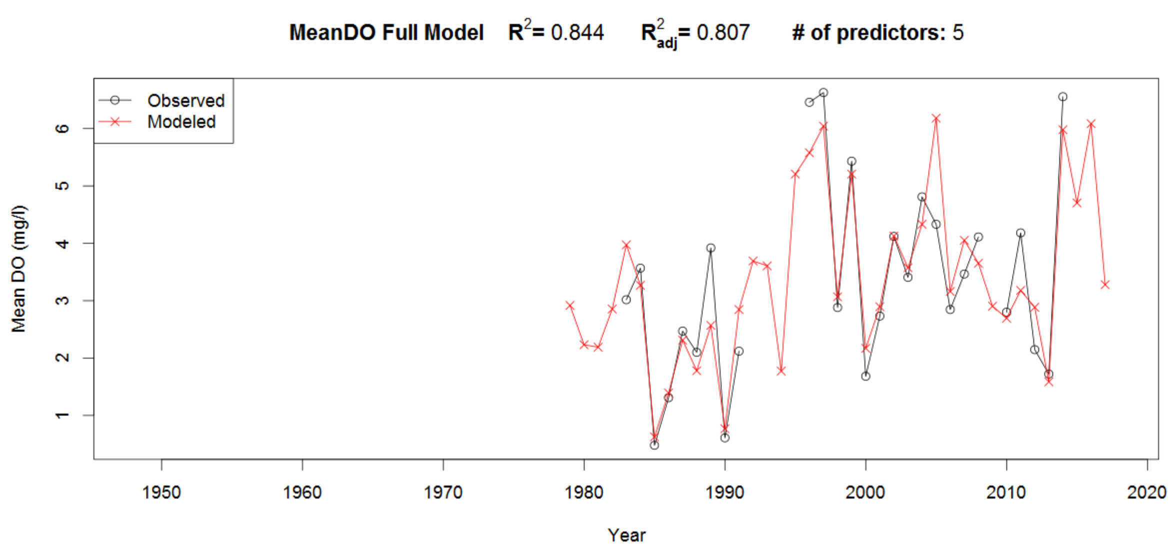

Figure 4-58. Time series plot of modeled vs. observed values (mean dissolved oxygen model).

4.7 MEDIAN DISSOLVED OXYGEN (DO) MODELS

Table 4.7-1. Correlations and p-values of median dissolved oxygen with teleconnection patterns. Significant correlations shown in boldface.

Index	r	p value	Significance (%)
ENSO	-0.171	0.395	60.5
$ENSO^2$	-0.052	0.798	20.2
NAO	-0.094	0.640	36.0
NAO^2	-0.100	0.620	38.0
AMO	0.392	0.043	95.7
AMO^2	-0.398	0.040	96.0
PDO	-0.033	0.868	13.2
PDO^2	-0.374	0.055	94.5

Winter Teleconnections Model

$$\text{MedianDO} = 6.62 + 0.02\text{ENSO} + -0.11\text{PDO}^2 + 4.36\text{AMO} + -0.43\text{WaterTempSpring} + 0.01\text{DurationDays} + -1.32\text{UWindJun} + 1.13\text{VWindJun} + -0.11\text{Diatom\_Apr} + e \qquad (4.7\text{-}1)$$

Table 4.7-2. Regression output for median dissolved oxygen model.

	MedianDO			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	6.62 *	1.80 – 11.44	2.69	0.016
ENSO	0.02	-0.56 – 0.59	0.06	0.956
PDO^2	-0.11	-0.58 – 0.35	-0.49	0.634
AMO	4.36 *	1.05 – 7.67	2.58	0.020
WaterTempSpring	-0.43	-0.91 – 0.04	-1.78	0.095
Duration..days.	0.01	-0.03 – 0.03	0.01	0.996
UWindJun	-1.32 **	-2.09 – -0.56	-3.39	0.004
VWindJun	1.13 *	0.23 – 2.03	2.47	0.025
Diatom_Apr	-0.11	-0.38 – 0.17	-0.76	0.461
Observations	25			
R^2 / adjusted R^2	0.737 / 0.606			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.7-3. Table summarizing the best subsets procedure for the median dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	UWindJun	0.202	0.170	0.029	28.627
2	UWindJun VWindJun	0.451	0.405	0.264	21.442
3	AMO UWindJun VWindJun	0.590	0.537	0.415	17.364
4	AMO WaterTempSpring UWindJun VWindJun	0.716	0.664	0.578	13.775
5	AMO WaterTempSpring UWindJun VWindJun Diatom_Apr	0.732	0.665	0.542	16.636
6	PDO^2 AMO WaterTempSpring UWindJun VWindJun Diatom_Apr	0.736	0.652	0.433	19.700
7	PDO^2 AMO WaterTempSpring Duration..days. UWindJun VWindJun Diatom_Apr	0.737	0.629	0.339	23.162
8	ENSO PDO^2 AMO WaterTempSpring Duration..days. UWindJun VWindJun Diatom_Apr	0.737	0.606	0.236	26.286

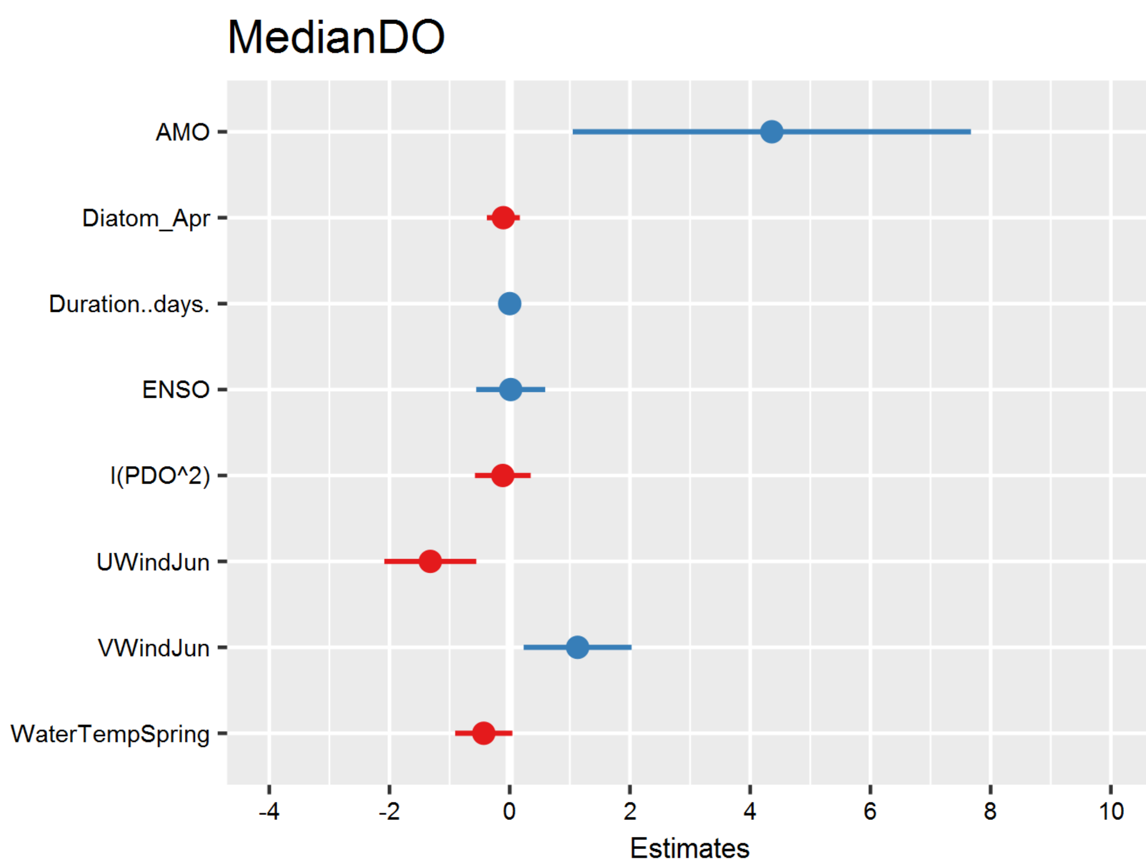

Figure 4-59. Regression coefficient plot (median dissolved oxygen model).

Figure 4-60. Linear correlations between median dissolved oxygen, biological parameters, and physical forcings.

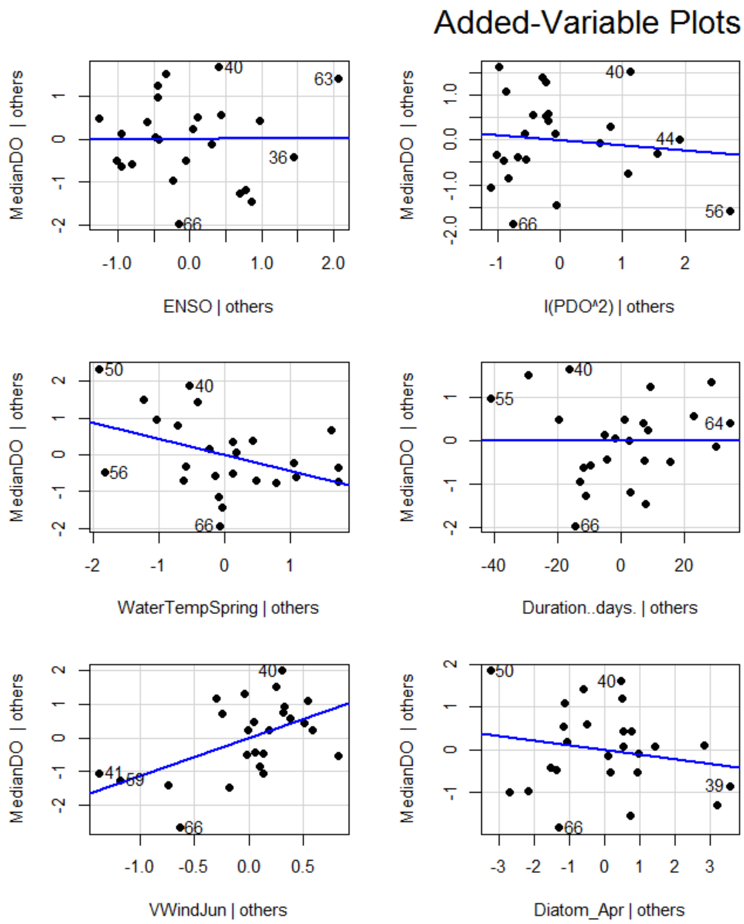

Figure 4-61. Added variable plot for median dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

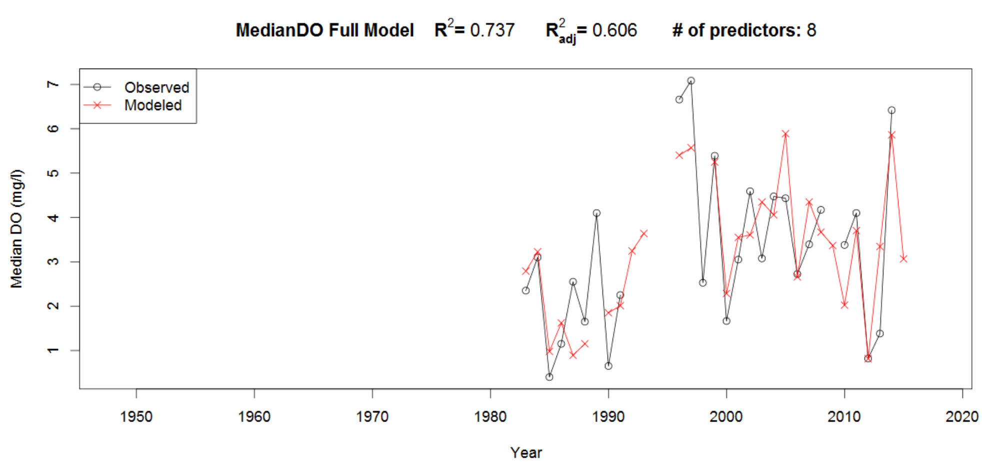

Figure 4-62. Time series plot of modeled vs. observed values (median dissolved oxygen model).

Monthly Teleconnections Model

$$\text{MedianDO} = 5.83 + 0.32\text{PDO.monthlyMay}^2 + 6.24\text{AMO} + -1.47\text{UWindJun} + 0.99\text{VWindJun} + -0.22\text{WaterTempSpring} + -2.63\text{NARR\_TotalPrecipApr} + e \quad (4.7\text{-}2)$$

Table 4.7-4. Regression output for median dissolved oxygen model.

	MedianDO			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	5.83 ***	4.38 – 7.28	7.89	<0.001
PDO.monthlyMay^2	0.32 *	0.02 – 0.61	2.11	0.048
AMO	6.24 ***	4.04 – 8.44	5.56	<0.001
UWindJun	-1.47 ***	-2.07 – -0.86	-4.76	<0.001
VWindJun	0.99 **	0.34 – 1.64	3.00	0.007
WaterTempSpring	-0.22	-0.44 – 0.00	-1.92	0.069
NARR_TotalPrecipApr	-2.63 *	-4.60 – -0.66	-2.61	0.017
Observations	27			
R^2 / adjusted R^2	0.822 / 0.768			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.7-5. Table summarizing the best subsets procedure for the median dissolved oxygen model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	PDO.monthlyMay^2	0.240	0.210	0.102	25.361
2	PDO.monthlyMay^2 AMO.monthlyJan	0.483	0.439	0.362	16.478
3	PDO.monthlyMay^2 AMO.monthlyJan Diatom_Apr	0.618	0.566	0.483	10.707
4	PDO.monthlyMay^2 AMO.monthlyJan monthlyErieWindsMAR Diatom_Apr	0.757	0.703	0.570	1.750
5	PDO.monthlyMay^2 AMO.monthlyJan Duration..days. monthlyErieWindsMAR Diatom_Apr	0.855	0.809	0.733	-0.659

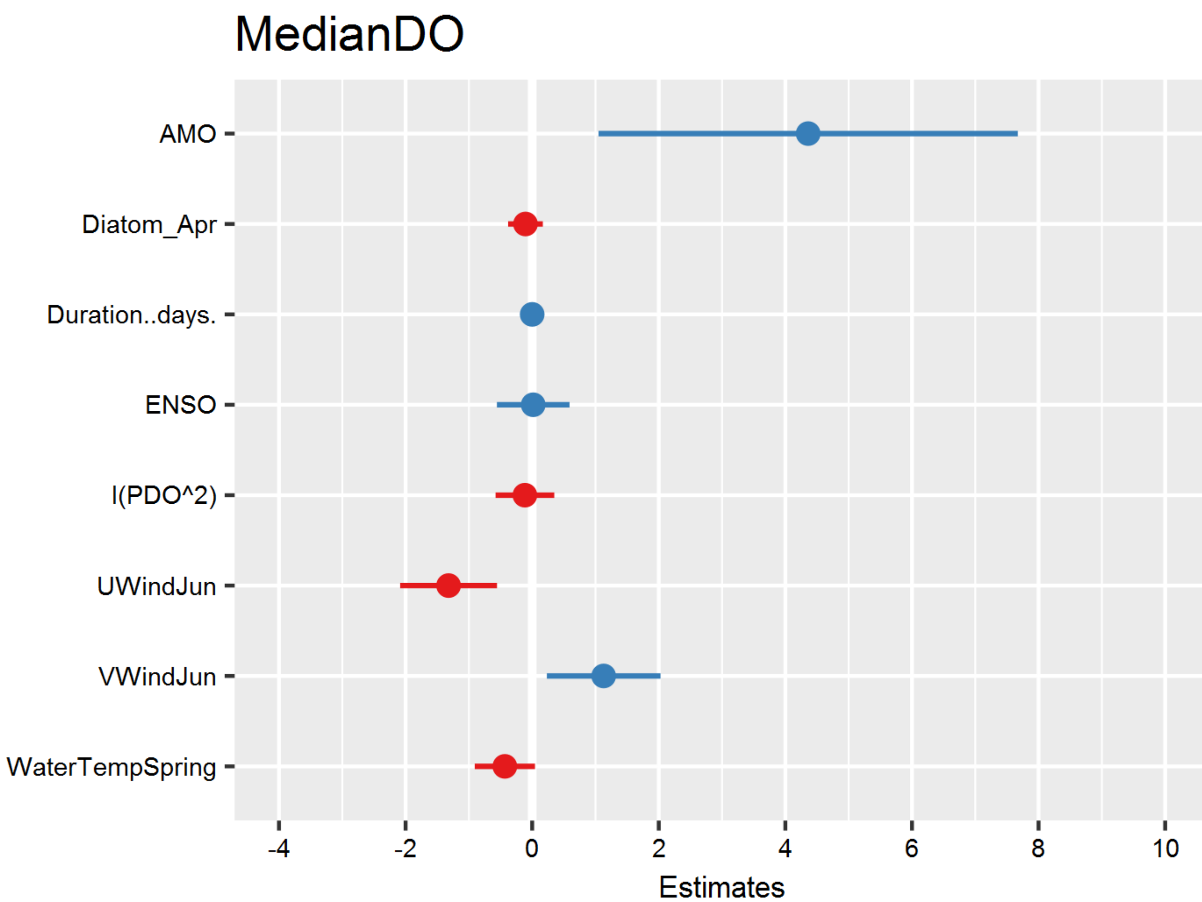

Figure 4-63. Regression coefficient plot (median dissolved oxygen model).

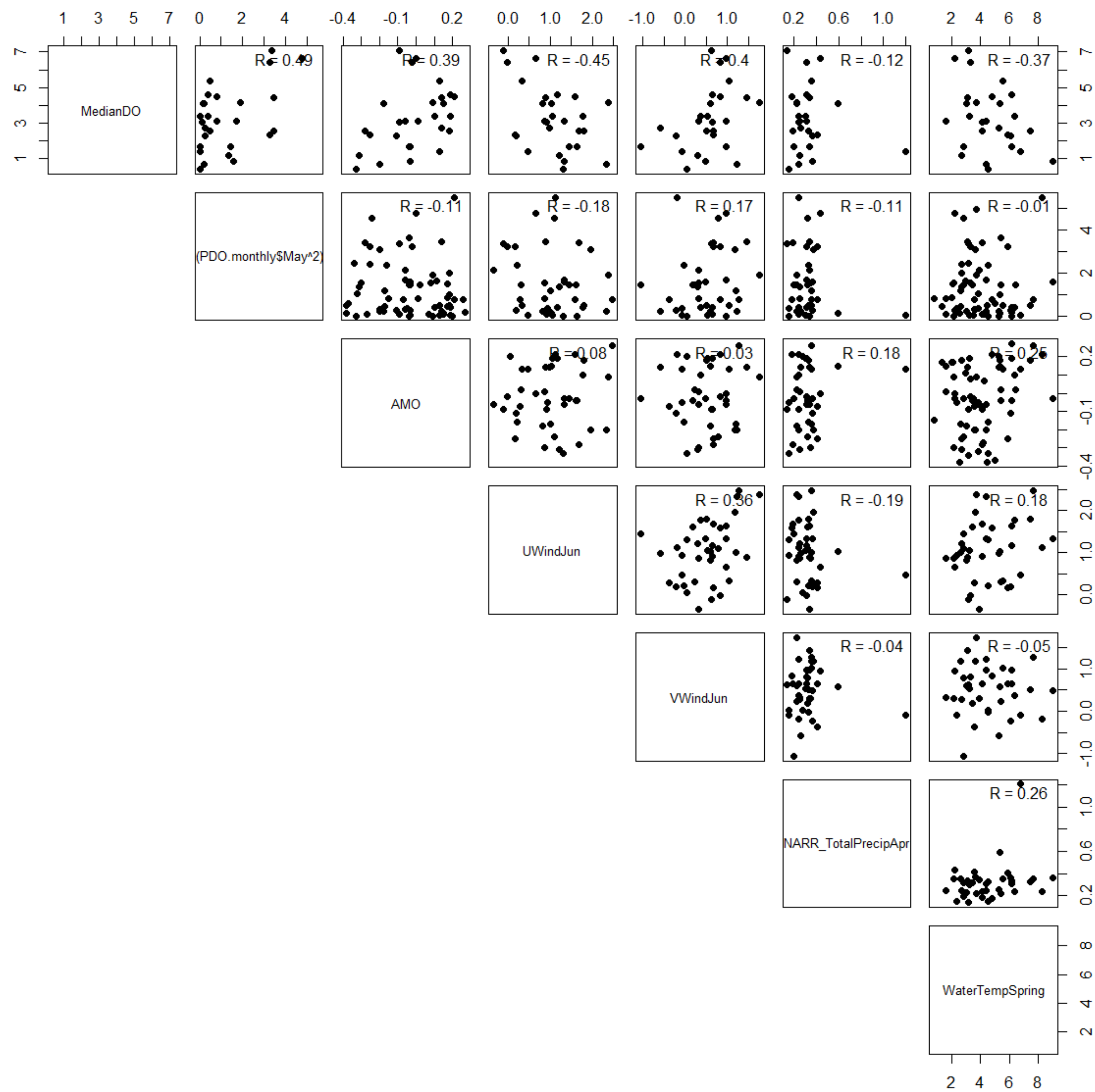

Figure 4-64. Linear correlations between median dissolved oxygen, biological parameters, and physical forcings.

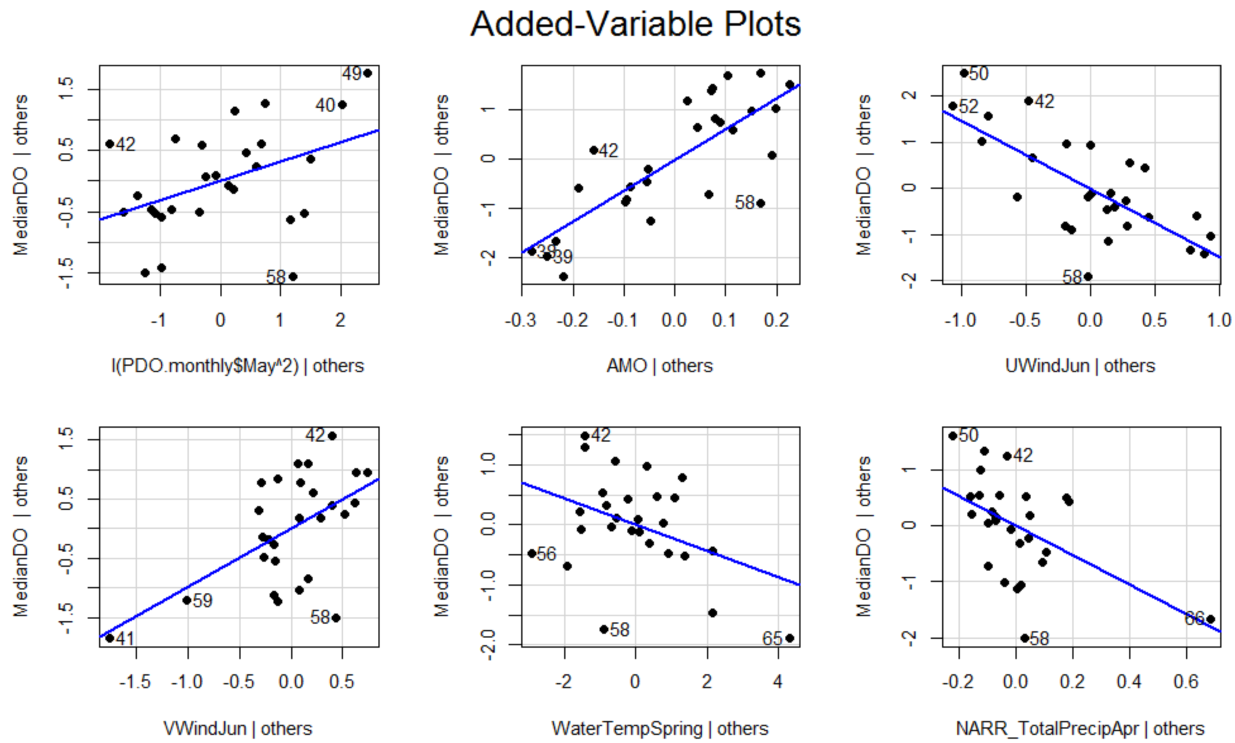

Figure 4-65. Added variable plot for median dissolved oxygen model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

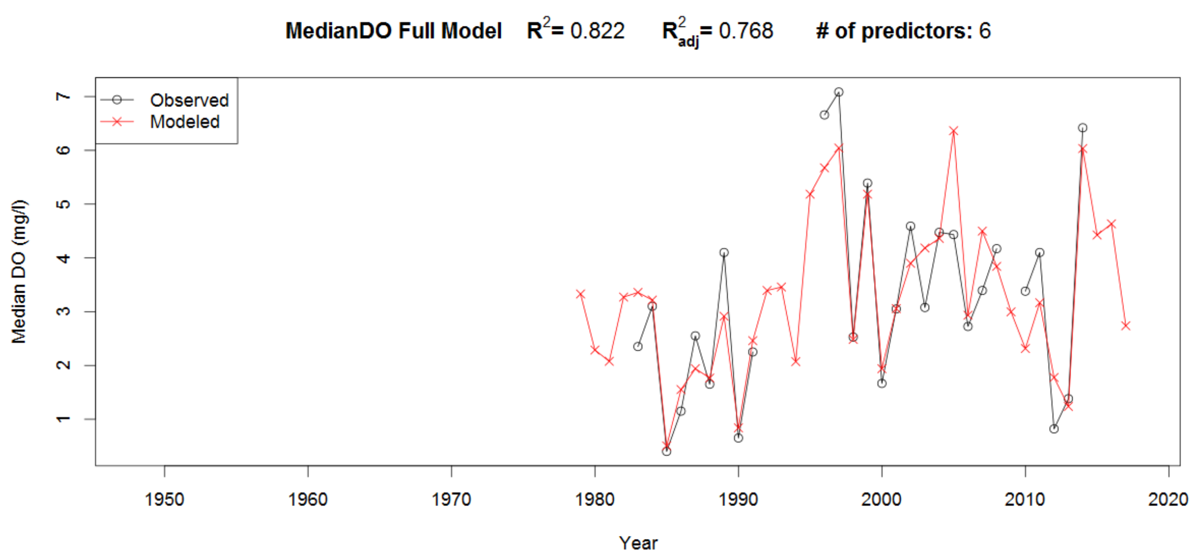

Figure 4-66. Time series plot of modeled vs. observed values (median dissolved oxygen model).

4.8 TOTAL PHOSPHATE LOAD MODELS

Based on previous studies on phosphate load (Dolan and Chapra 2012; Maxccoux et al., 2016) and Wang et al. (2018), we attempt to construct similar regression models on hindcast of phosphate load.

Table 4.8-1. Correlations and p-values of median dissolved oxygen with teleconnection patterns. Significant correlations shown in boldface.

Index	r	p value	Significance (%)
ENSO	-0.028	0.853	14.7
$ENSO^2$	-0.062	0.678	32.2
NAO	-0.362	0.013	98.7
NAO^2	-0.189	0.203	79.7
AMO	-0.261	0.076	92.4
AMO^2	0.191	0.200	80.0
PDO	-0.275	0.061	93.9
PDO^2	0.035	0.815	18.5

Winter Teleconnections Model

$$TPload = 69500.83 + -2075.52NAO + -2734.69PDO + -5208.75AMO$$
$$+ -1322.19WaterTempSpring + -59.25AMIC\_ERIE + -7211.18Wind + e \quad (4.8\text{-}1)$$

Table 4.8-2. Regression output for total phosphate model.

	Tpload			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	69500.83 ***	36647.21 – 102354.46	4.15	<0.001
NAO	-2075.52 *	-3993.52 – -157.51	-2.12	0.041
PDO	-2734.69 ***	-4116.38 – -1353.01	-3.88	<0.001
AMO	-5208.75	-13053.81 – 2636.32	-1.30	0.201
WaterTempSpring	-1322.19	-2607.79 – -36.59	-2.02	0.051
AMIC_ERIE	-59.25	-138.43 – 19.93	-1.47	0.151
Wind	-7211.18 **	-11970.33 – -2452.03	-2.97	0.005
Observations	45			
R^2 / adjusted R^2	0.493 / 0.413			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.8-3. Table summarizing the best subsets procedure for the total phosphate model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	Wind	0.176	0.156	0.080	773.025
2	PDO Wind	0.350	0.319	0.244	765.094
3	NAO PDO Wind	0.407	0.364	0.266	763.526
4	NAO PDO WaterTempSpring Wind	0.450	0.395	0.272	762.928
5	NAO PDO WaterTempSpring AMIC_ERIE Wind	0.471	0.403	0.272	763.879
6	NAO PDO AMO WaterTempSpring AMIC_ERIE Wind	0.493	0.413	0.285	764.860

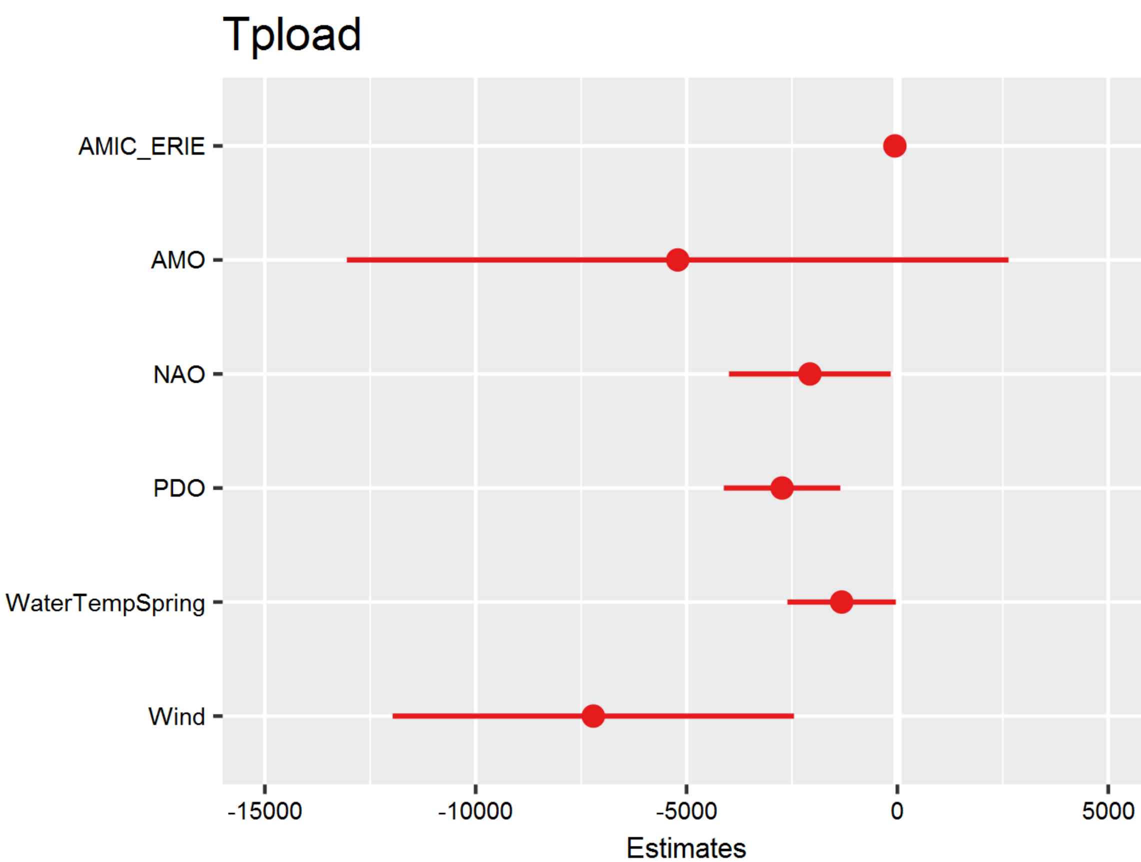

Figure 4-67. Regression coefficient plot (total phosphate model).

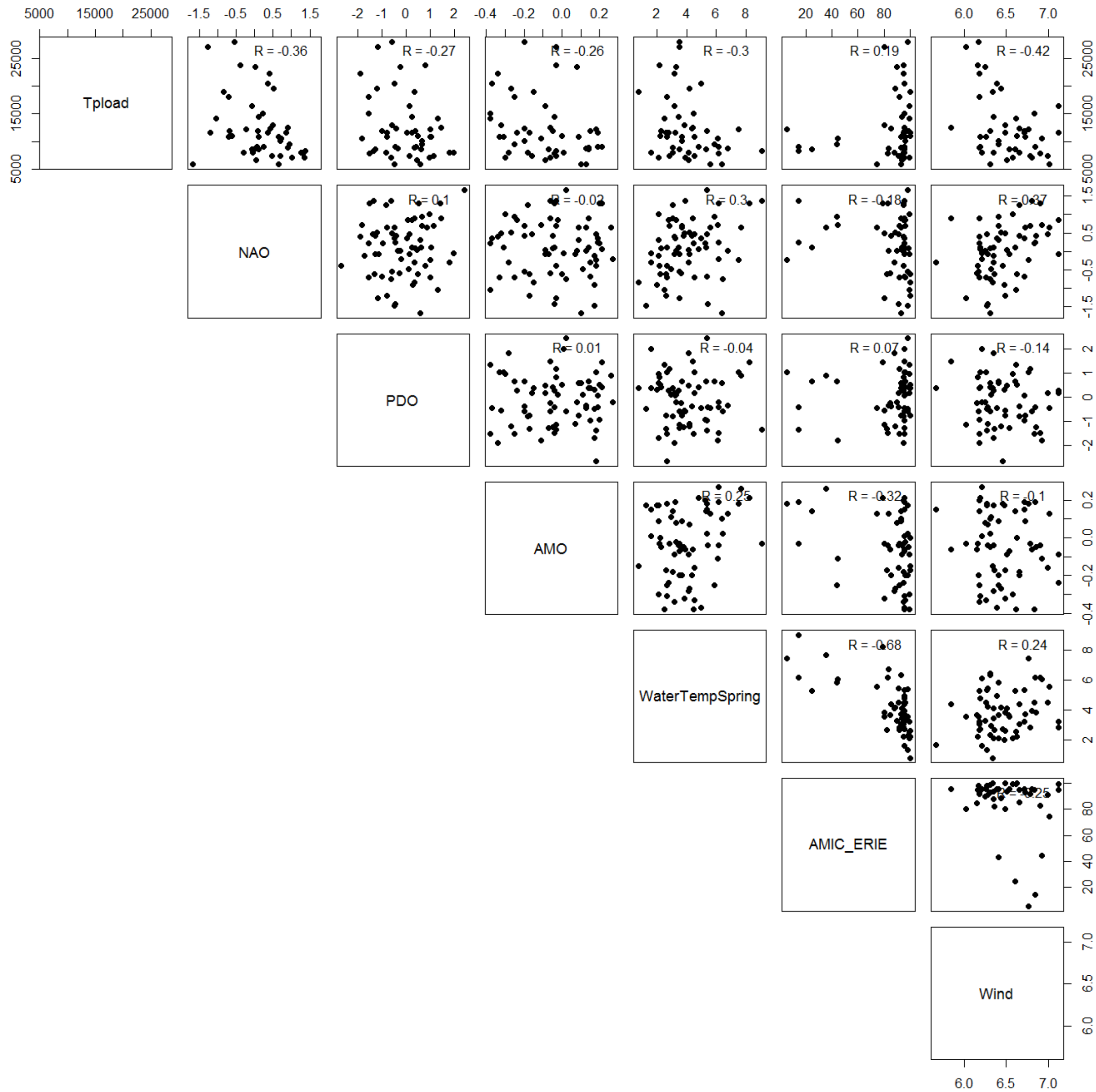

Figure 4-68. Linear correlations between total phosphate load, biological parameters, and physical forcings.

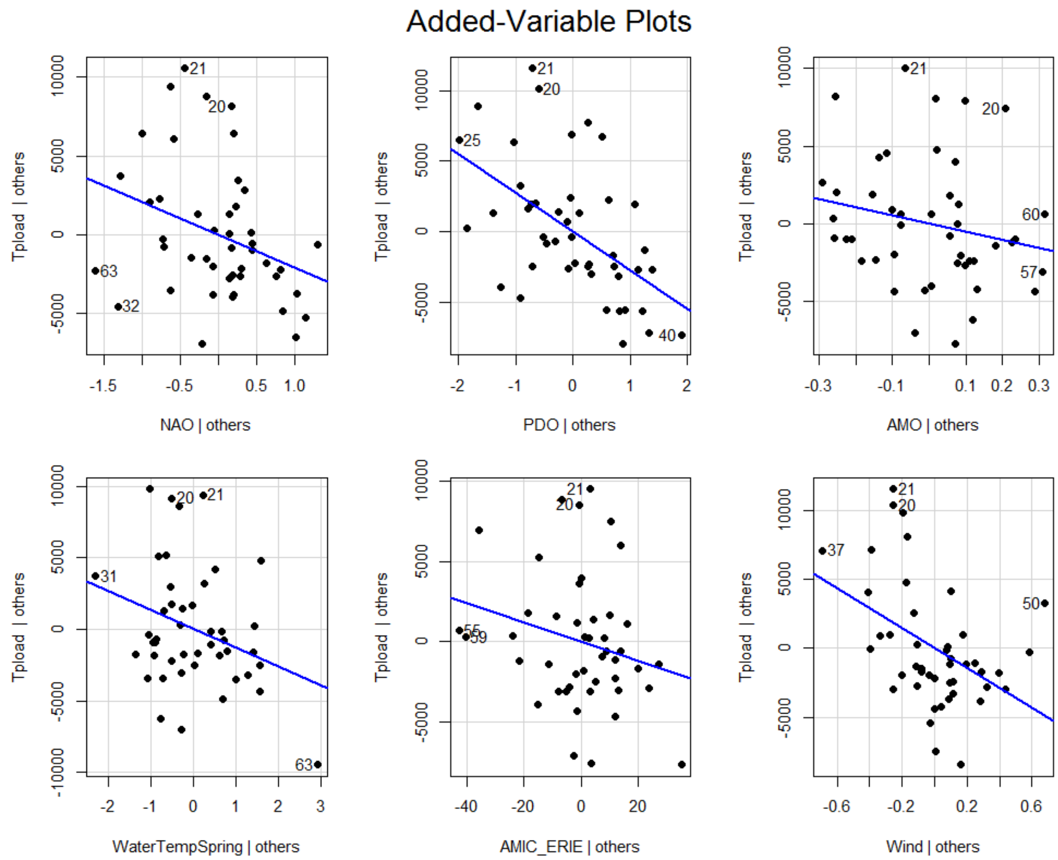

Figure 4-69. Added variable plot for total phosphate model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

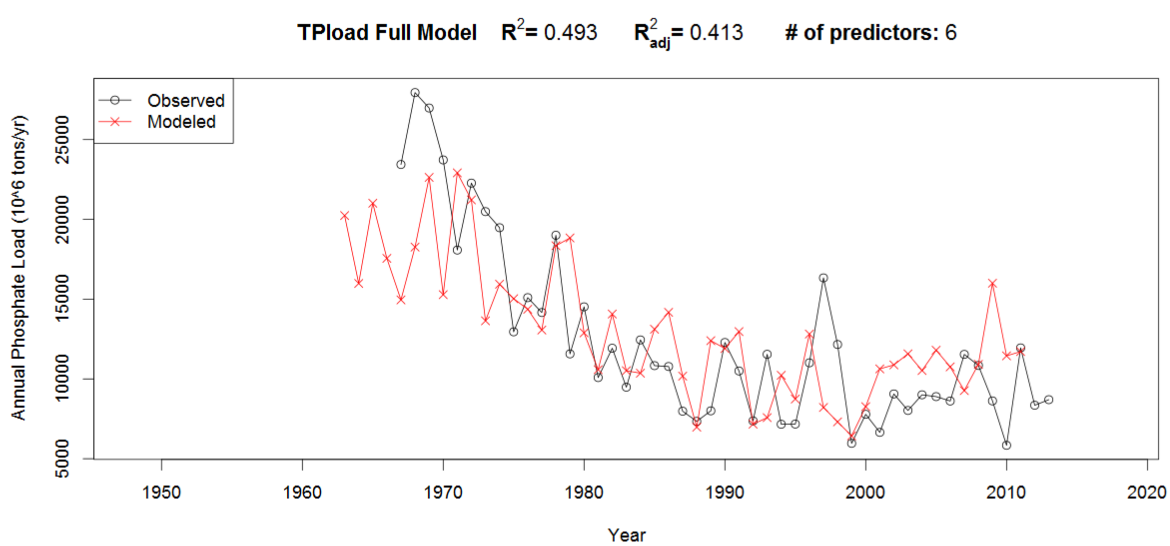

Figure 4-70. Time series plot of modeled vs. observed values (total phosphate load model).

4.9 HYPOXIC FACTOR MODELS

Hypoxic factor is a new metric used to describe hypoxia in the central basin of Lake Erie. It is calculated using the Mean Area metric from Zhou et al. (2015), and is virtually the same variable (hypoxic factor and mean area have correlation r = 0.99). However, there is less missing data for the hypoxic factor variable, making it useful for developing hypoxia prediction models.

Table 4.9-1. Correlations and p-values of hypoxic factor with teleconnection patterns. Significant correlations shown in boldface.

Index	**r**	**p value**	**Significance (%)**
ENSO	0.148	0.453	54.7
ENSO^2	0.114	0.564	43.6
NAO	-0.061	0.759	24.1
NAO^2	0.227	0.244	75.6
AMO	-0.316	0.102	89.8
AMO^2	0.134	0.496	50.4
PDO	0.106	0.592	40.8
PDO^2	0.554	0.002	99.8

Monthly Teleconnections Model (2 terms)

$$\text{HypoxicFactor} = 2.71 + 3.18\text{PDO}^2 + 10.45\text{eriePrecip\_PMarAprRatio} + e \quad (4.9\text{-}1)$$

Table 4.9-2. Regression output for hypoxic factor model.

	HypoxicFactor			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	2.71	-0.72 – 6.14	1.55	0.134
PDO^2	3.18 ***	1.69 – 4.66	4.20	<0.001
eriePrecip_PMarAprRatio	10.45 ***	7.17 – 13.74	6.24	<0.001
Observations	28			
R^2 / adjusted R^2	0.729 / 0.707			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.9-3. Table summarizing the best subsets procedure for the hypoxic factor model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	eriePrecip_PMarAprRatio	0.538	0.520	0.481	95.978
2	PDO^2 eriePrecip_PMarAprRatio	0.729	0.707	0.658	85.347

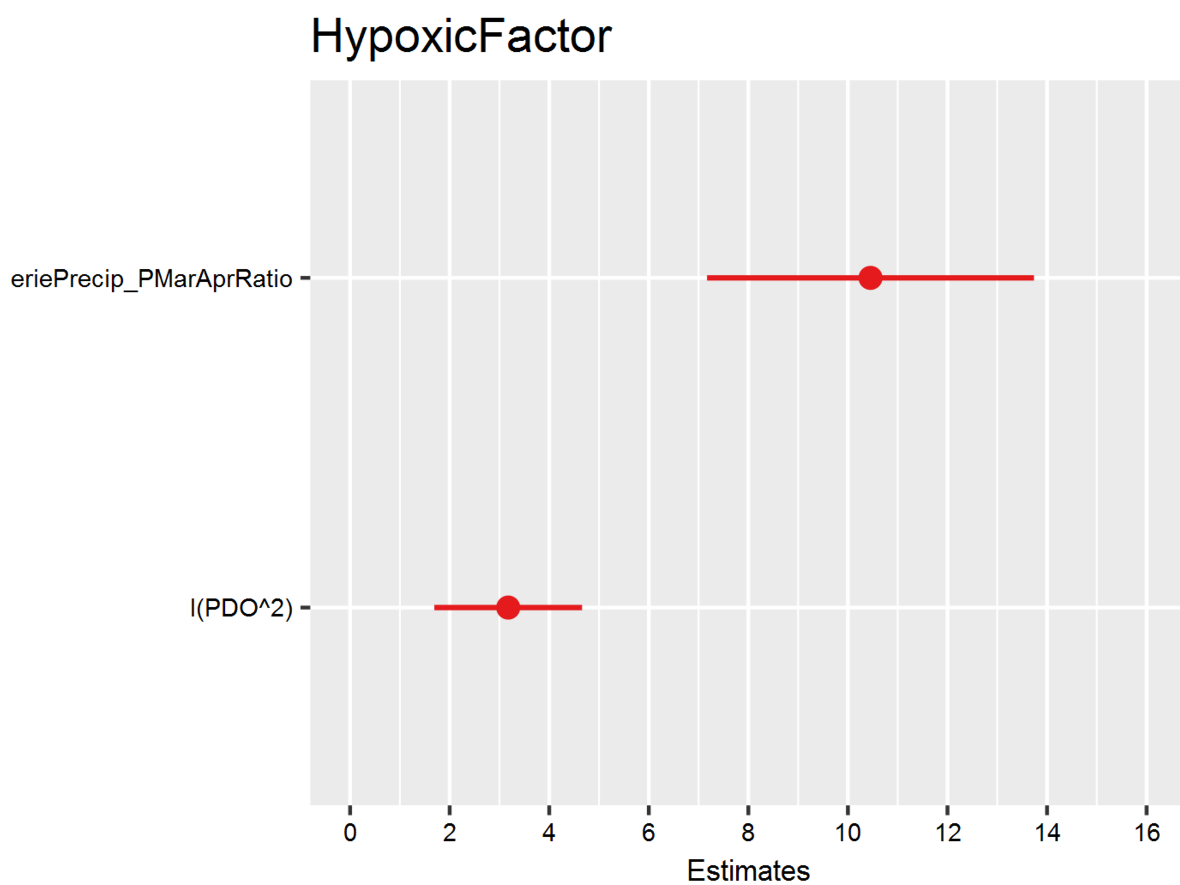

Figure 4-71. Regression coefficient plot (hypoxic factor model).

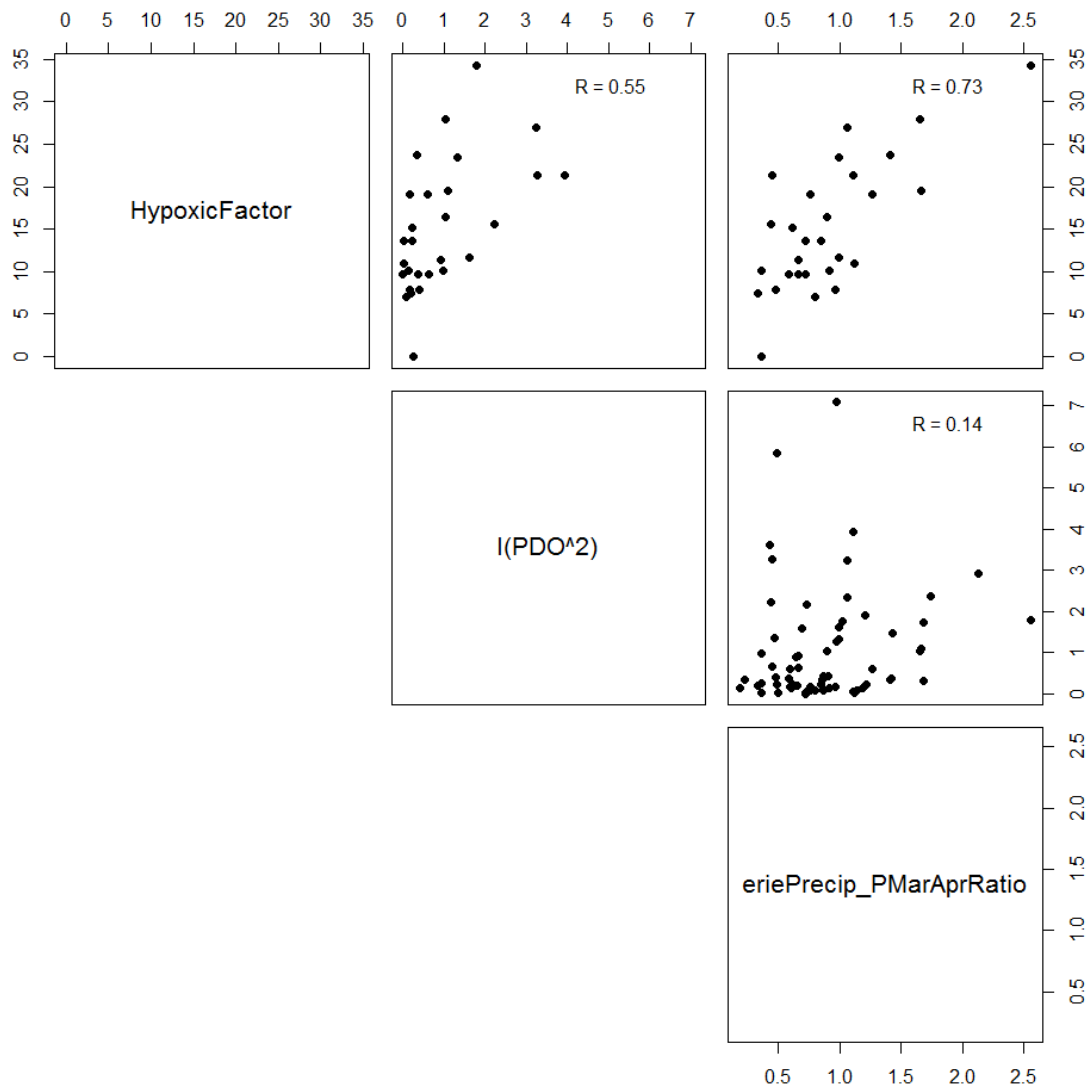

Figure 4-72. Linear correlations between hypoxic factor, biological parameters, and physical forcings.

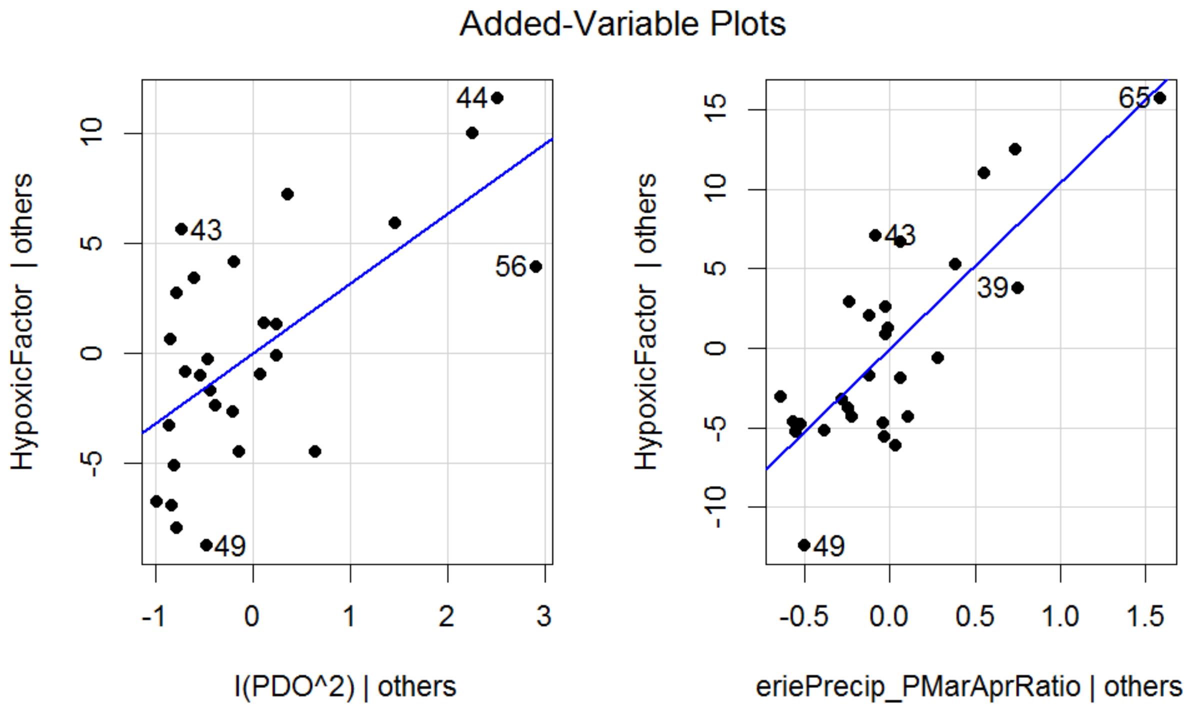

Figure 4-73. Added variable plot for hypoxic factor model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

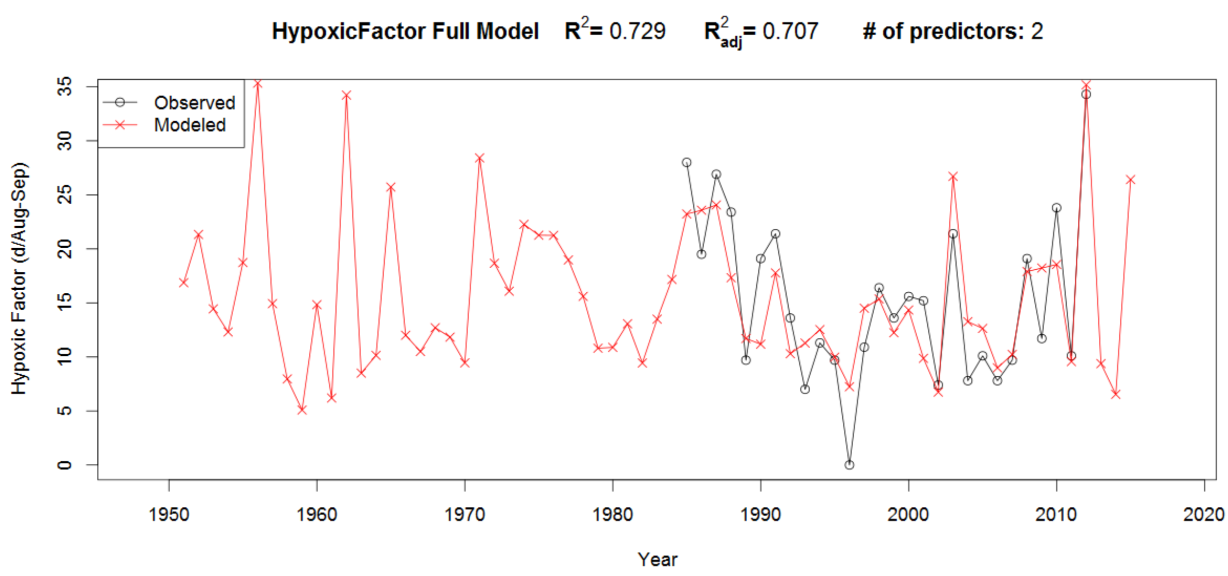

Figure 4-74. Time series plot of modeled vs. observed values (hypoxic factor model).

Monthly Teleconnections Model (2 terms, including ratio term)

$$\text{HypoxicFactor} = -13.11 + 13.29\text{erie7yrPrecip\_AprMayPrecip\_Ratio} + 3.25\text{erieAirTempMarApr} + e \quad (4.9\text{-}2)$$

Table 4.9-4. Regression output for hypoxic factor model.

	HypoxicFactor			
Predictors	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	-13.11 ***	-19.89 – -6.34	-3.79	0.001
erie7yrPrecip_AprMayPrecip_Ratio	13.29 ***	8.03 – 18.55	4.95	<0.001
erieAirTempMarApr	3.25 ***	1.99 – 4.51	5.06	<0.001
Observations	28			
R^2 / adjusted R^2	0.757 / 0.737			
**** p<0.05 ** p<0.01 *** p<0.001***				

Table 4.9-5. Table summarizing the best subsets procedure for the hypoxic factor model. The table shows the effect of removing one or more predictors on R^2, R^2_{adj}, R^2-predicted, and Bayesian information criterion (BIC).

N	Predictors	Rsquare	AdjRsq	PredRsq	BIC
1	erieAirTempMarApr	0.519	0.500	0.448	96.631
2	erie7yrPrecip_AprMayPrecip_Ratio erieAirTempMarApr	0.757	0.737	0.708	82.302

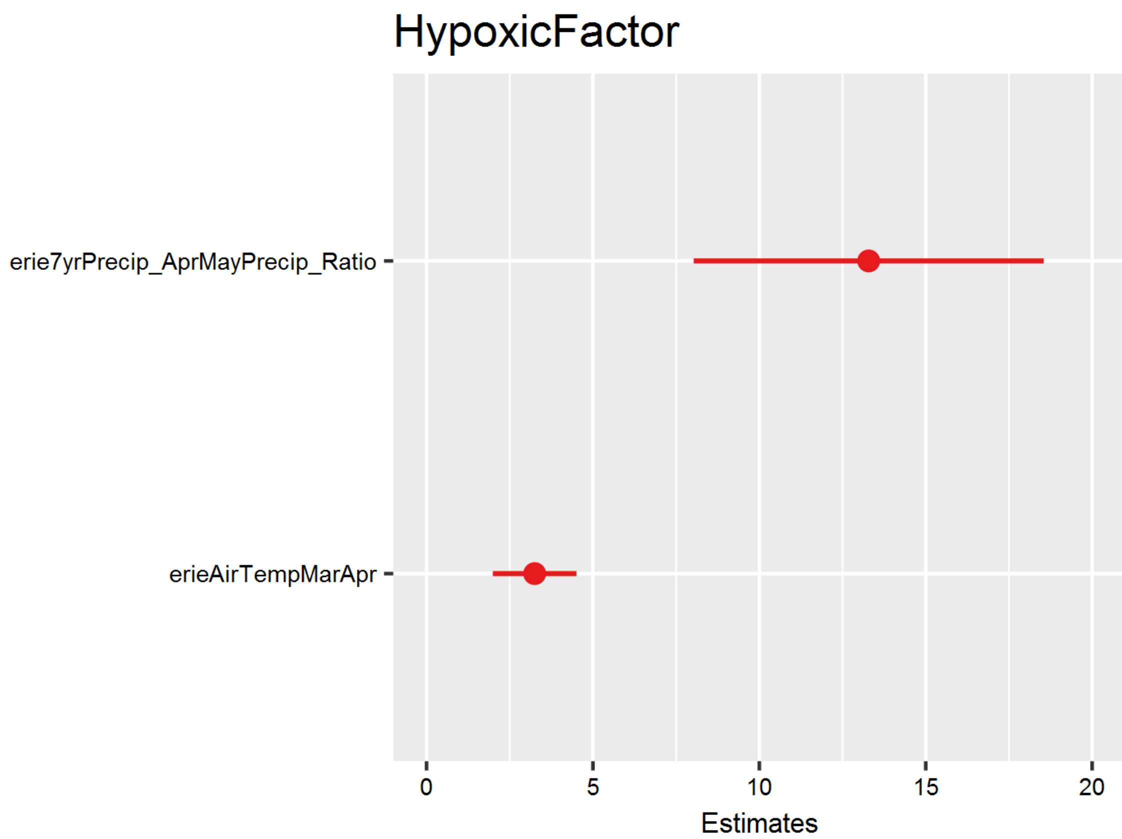

Figure 4-75. Regression coefficient plot (hypoxic factor model).

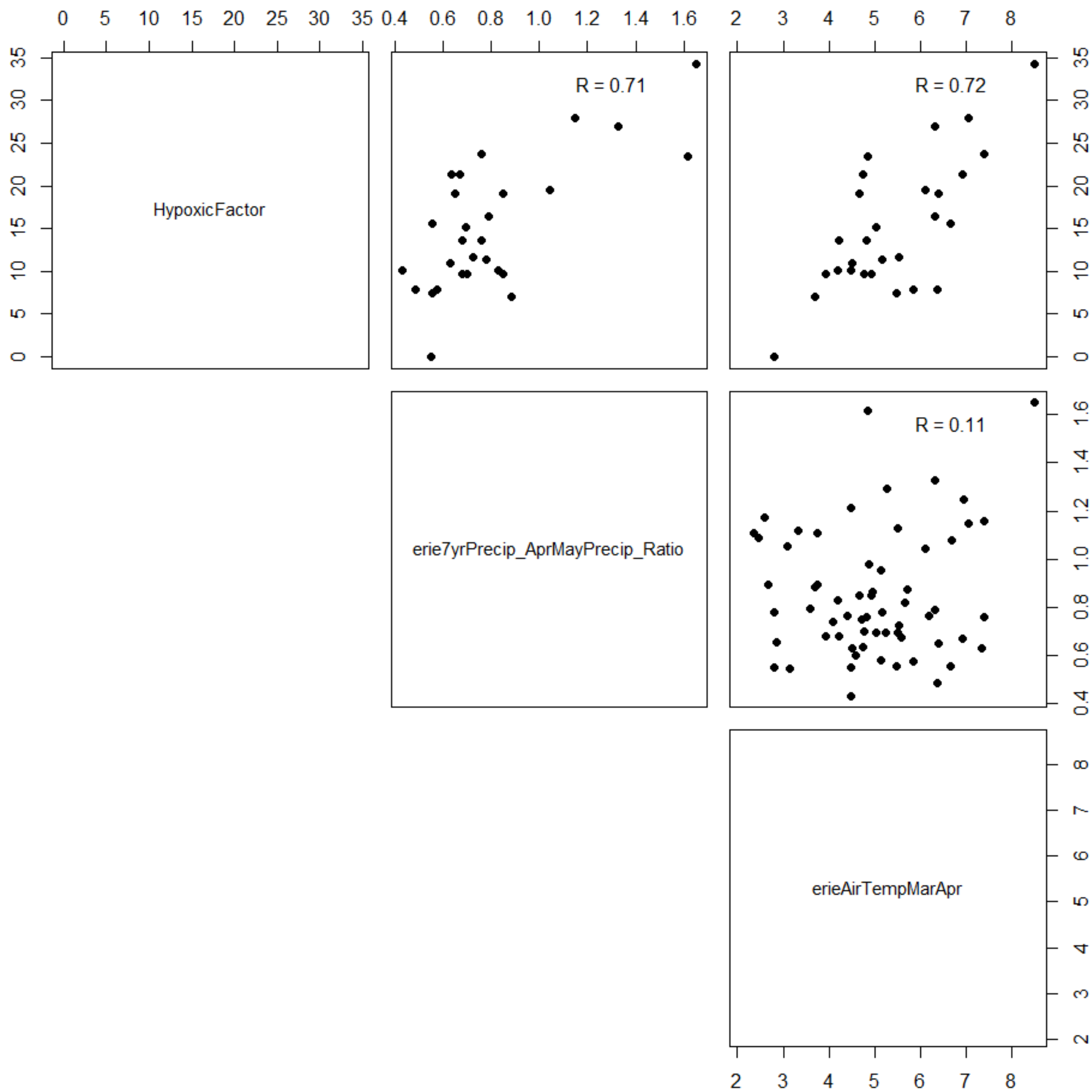

Figure 4-76. Linear correlations between hypoxic factor, biological parameters, and physical forcings.

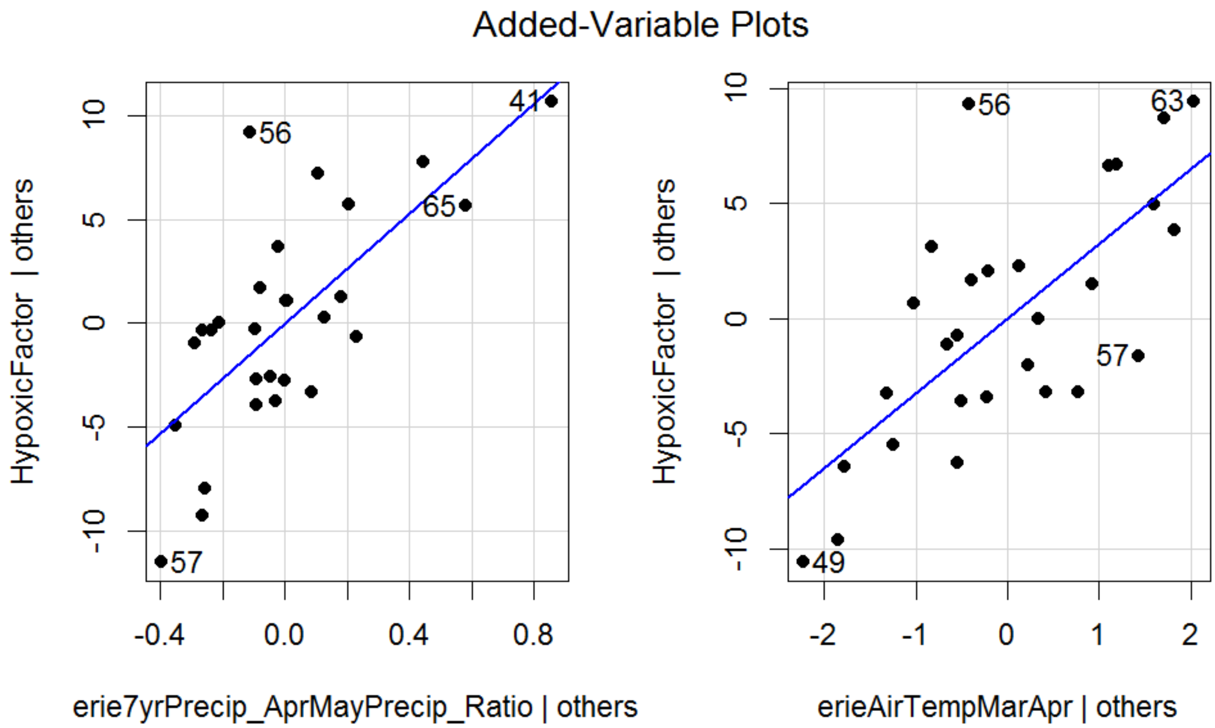

Figure 4-77. Added variable plot for hypoxic factor model. A strong linear relationship in the added variable plot indicates the increased importance of the contribution of X to the model already containing the other predictors. The numbers denote the data points in the data time series.

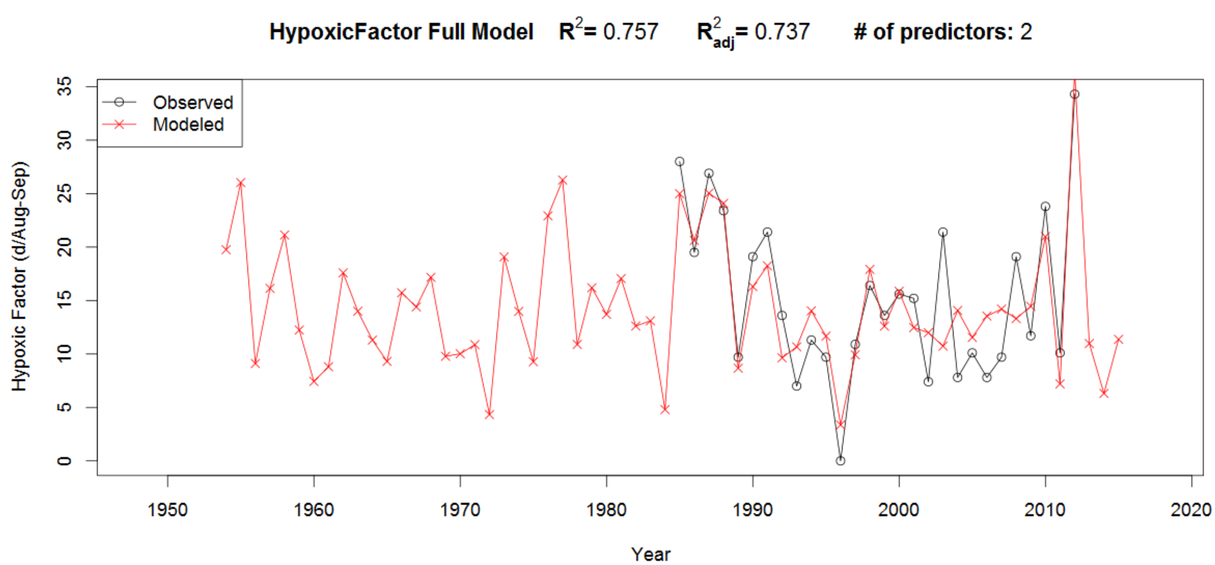

Figure 4-78. Time series plot of modeled vs. observed values (hypoxic factor model).

5.0 CORRELATIONS AND TIME SERIES OF TELECONNECTIONS AND WATER LEVEL AND PRECIPITATION

This section contains tables and time series plots of teleconnections and climate/hydrological variables (Hunter et al. 2015) that were not included in the technical memorandum (Zhang et al. 2018). Each time series plot includes two smooth lines to highlight trend in the data over time; the blue line is calculated with locally weighted smoothing and the red line is a 10-year running mean.

5.1 ERIE ANNUAL AVERAGE WATER LEVEL

Table 5.1-1. Correlations and p-values of Lake Erie water level with teleconnection patterns. Significant correlations shown in boldface.

Index	r	p value	Significance (%)
ENSO	-0.024	0.845	15.5
$ENSO^2$	0.133	0.280	72.0
NAO	0.278	0.022	97.8
NAO^2	-0.149	0.225	77.5
AMO	-0.427	0.000	100.0
AMO^2	0.514	0.000	100.0
PDO	0.197	0.107	89.3
PDO^2	0.052	0.671	32.9

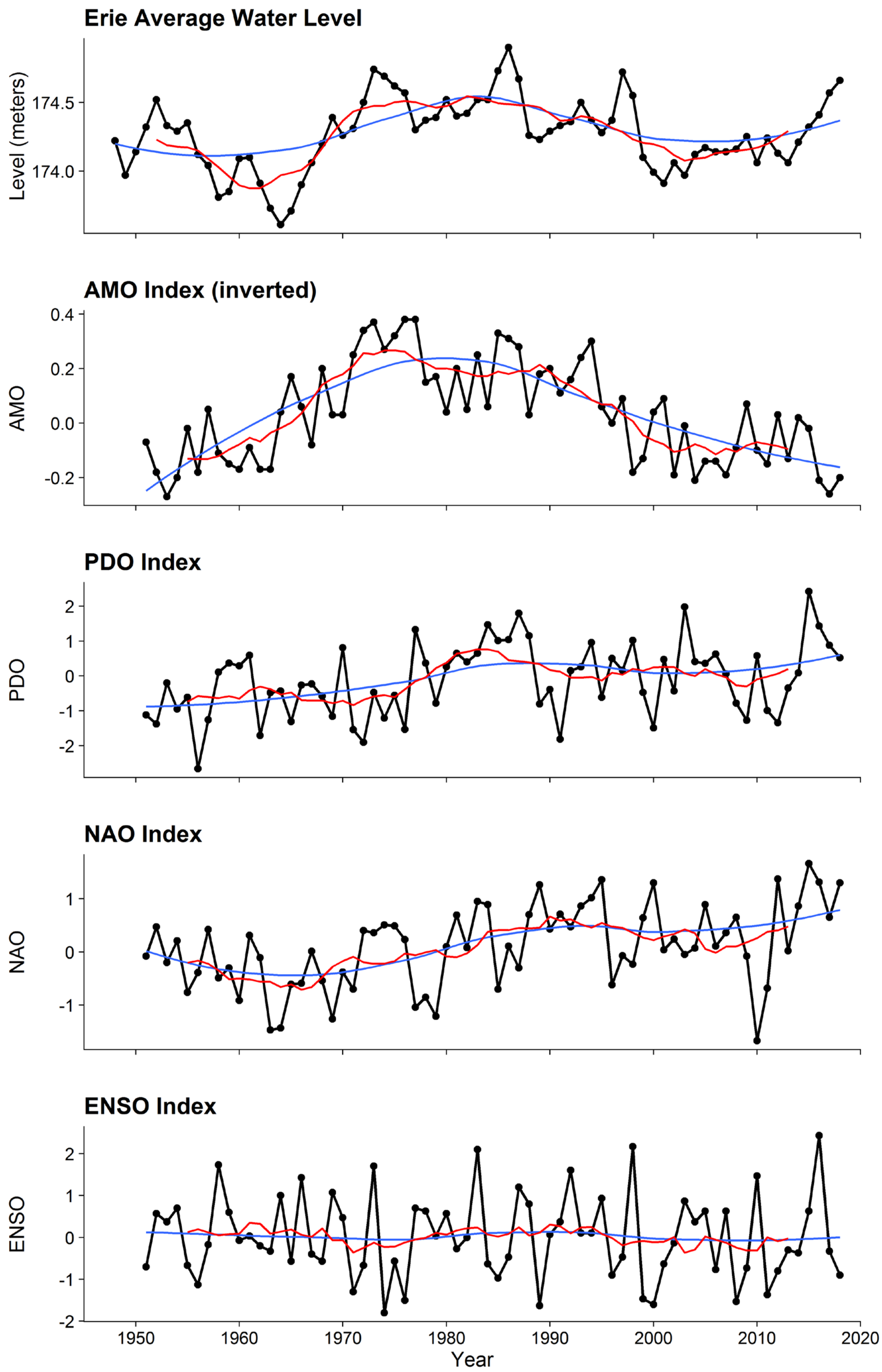

Figure 5-1. Erie annual average water level plotted with AMO, PDO, NAO, and ENSO.

5.2 GREAT LAKES BASIN ANNUAL PRECIPITATION

Table 5.2-1. Correlations and p-values of basin annual precipitation with teleconnection patterns. Significant correlations shown in boldface.

Index	**r**	**p value**	**Significance (%)**
ENSO	-0.147	0.261	73.9
ENSO^2	-0.094	0.477	52.3
NAO	0.113	0.389	61.1
NAO^2	-0.166	0.205	79.5
AMO	-0.357	0.005	99.5
AMO^2	0.160	0.222	77.8
PDO	0.183	0.163	83.7
PDO^2	-0.123	0.348	65.2

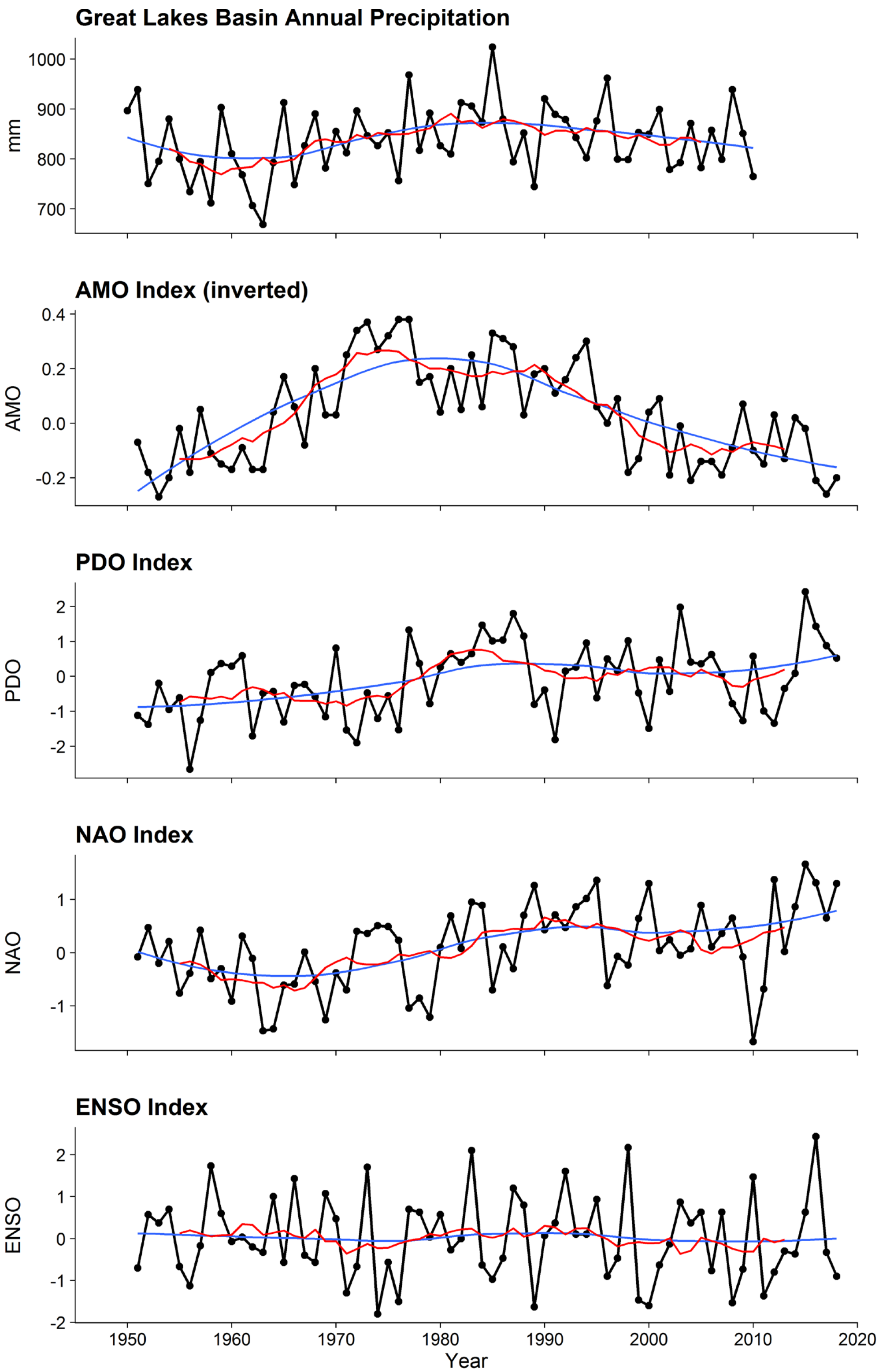

Figure 5-2. Great Lakes basin precipitation plotted with AMO, PDO, NAO, and ENSO.

6.0 SUMMARY

This report builds on the work of Wang et al. (2018) and Zhang et al. (2018) to further develop statistical prediction models for hypoxia and water quality in Lake Erie by linking water quality to teleconnection patterns. Observed relationships between biological parameters and environmental forcings were used to develop multi-variable regression equations to predict ecosystem changes in Lake Erie. Teleconnection patterns were added as linear or quadratic terms, depending on their relationship with water quality variables. The majority of statistical models used predictors from the winter or spring seasons, allowing for prediction of water quality variables later in the summer and fall.

Prediction models are accompanied by three tables and four figures that summarize model output and model skill. In addition, prediction models are shown as a “full model” and tables show the effect of removing one or more predictor on commonly used model skill statistics such as R^2. Thus, users may decide if any predictors should be excluded from the model based on prior knowledge or if model skill remains high with fewer predictors. This approach will facilitate future research into choosing the “best” models contained here that are the most physically realistic, have the highest explanatory power of water quality variables, and contain the fewest predictors.

Based on the above investigations, several summary points can be drawn:

1. In nearly all models, using only teleconnections as predictors was not enough to explain a large amount of the variance in water quality variables. Instead, the variables that had the strongest correlations with water quality variables were more local climate variables such as spring temperature and precipitation over the Lake Erie basin. In some models, two or three local climate variables were enough to explain over half of the variance in water quality variables (high R^2). Adding teleconnections to these simple models often improved the model fit, but these improvements were not always large.
2. Predicting hypolimnetic oxygen demand (HOD) was generally more complicated than predicting other water quality variables such as mean hypoxic area. This may be because HOD is primarily a function of nutrient (phosphorus) inputs and is less related to climate.
3. Annual mean ice cover (AMIC) was not strongly related to water quality variables, however, the duration of ice cover (Duration Days) was significantly correlated with several water quality variables. This may be because the duration of ice cover strongly related to the timing of thermal stratification in Lake Erie.
4. Models for mean hypoxic area suggest that a combination of wet prior-year springs (high nutrient load) followed by dry current-year springs (earlier stratification onset) may be a large driver of hypoxic extent. This was reflected in high correlations between water quality variables and ratio and difference terms that integrated wet prior springs and dry current springs into one variable.

5. The strongest evidence for a link between teleconnections and Lake Erie water quality seemed to be through the Atlantic Multidecadal Oscillation (AMO) affecting Great Lakes precipitation and Lake Erie annual water levels, which in turn would affect heating of the lake, the development and depth of the thermocline, and thus rates of dissolved oxygen depletion and hypoxia. There is evidence from previous studies that support this observation. First, Hanrahan et al. (2010) linked phases of the AMO to a 27-year periodicity in Michigan-Huron water levels, which are directly related to Lake Erie water levels (correlation r = 0.90) through their connection via the Detroit river. Further, this is consistent with Enfield et al. (2001) who found the AMO to have significant negative correlations with precipitation in the Great Lakes basin, also seen in this study (Table 4.3-1). Increases in Lake Erie hypoxia and harmful algal blooms in the 1990s was approximately coincident with a phase change in the AMO from negative to positive, though further research is needed to investigate this relationship and potential causal links.

7.0 ACKNOWLEDGMENTS

We appreciate support from the Cooperative Institute of Great Lakes Research (CIGLR) 2019 Summer Fellows program to Matthew Trumper and 2018 Summer Fellow program to Ting-yi Yang. This is GLERL Contribution No. 1958. Funding was awarded to Cooperative Institute for Great Lakes Research (CIGLR) through the NOAA Cooperative Agreement with the University of Michigan (NA17OAR4320152). This CIGLR contribution number is xxxx.

8.0 REFERENCES

Assel, R.A., and S. Rodionov. Atmospheric teleconnections for annual maximum ice cover on the Laurentian Great Lakes. *International Journal of Climatology* 18:425-442 (1998). http://www.glerl.noaa.gov/pubs/fulltext/1998/19980001.pdf (1998).

Bai, X., J. Wang, C.E. Sellinger, A.H. Clites, and R.A. Assel. Interannual variability of Great Lakes ice cover and its relationship to NAO and ENSO. *Journal of Geophysical Research* 117(C03002):25 pp. (DOI:10.1029/2010JC006932) (2012a).

Bai, X., and J. Wang. Atmospheric teleconnection patterns associated with severe and mild ice cover on the Great Lakes, 1963–2011. *Water Quality Research Journal Canada*, 47, 421–435, (DOI:10.2166/wqrjc.2012.009) (2012b).

Dolan, D.M. and S.C. Chapra. Great Lakes total phosphorus revisited: 1. Loading analysis and update (1994-2008). *Journal of Great Lakes* Research 38:730-740. (DOI: 10.1029/2010JC006932) (2012).

Hunter, T. S., A. H. Clites, K. B. Campbell, and, A. D. Gronewold. Development and application of a North American Great Lakes hydrometeorological database—Part I: Precipitation, evaporation, runoff, and air temperature. *Journal of Great Lakes Research*, 41(1): 65-77 pp. (DOI: 10.1016/j.jglr.2014.12.006) (2015).

Maccoux M.J., A. Dove, S.M. Backus, and D.M. Dolan. Total and soluble reactive phosphorus loadings to Lake Erie: A detailed accounting by year, basin, country, and tributary. *Journal of Great Lakes Research* 42:1151-1165. (DOI:10.1016/j.jglr.2016.08.005) (2016).

Enfield, D. B., A.M. Mestas-Nuñez and P.J. Trimble. The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters*, *28*(10), 2077–2080. (DOI:10.1029/2000GL012745) (2001).

Del Giudice, D., Y. Zhou, E. Sinha, and A.M. Michalak. Long-Term Phosphorus Loading and Springtime Temperatures Explain Interannual Variability of Hypoxia in a Large Temperate Lake. *Environmental Science & Technology*, 52(4), 2046–2054. (DOI:10.1021/acs.est.7b04730) (2018).

Hanrahan, J. L., S.V. Kravtsov, and P.J. Roebber. Connecting past and present climate variability to the water levels of Lakes Michigan and Huron. *Geophysical Research Letters*, **37**, L01701. (DOI:10.1029/2009GL041707) (2010).

Mishra, V., Cherkauer, K., Bowling, C., and Huber, M. Lake Ice phenology of small lakes: Impacts of climate variability in the Great Lakes region. *Global and Planetary Change*, (DOI:10.1016/j.gloplacha.2011.01.004) (2011).

Niimi, A.J. Economic and environmental issues of the proposed extension of the winter navigation season and improvements on the Great Lakes-St. Lawrence Seaway system, *Journal of Great Lakes Research*, 8(3), 532–549. (DOI:10.1016/S0380-1330) (1982).

Rowe, M.D., E.J. Anderson, D. Beletsky, C.A. Stow, S.D. Moegling, J.D. Chaffin, J.C. May, P.D. Collingsworth, A. Jabbari, and J.D. Ackerman. Coastal Upwelling Influences Hypoxia Spatial Patterns and Nearshore Dynamics in Lake Erie. *Journal of Geophysical Research: Oceans,* 124(8):6154-6175 (DOI:10.1029/2019JC015192). (2019)

Scavia, D., J. David Allan, K. Arend, S. Bartell, D. Beletsky, N.S. Bosch, and Y. Zhou. Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. *Journal of Great Lakes Research*, 40(2), 226–246. (DOI:10.1016/J.JGLR.2014.02.004) (2014).

Vanderploeg, H.A., S.J. Bolsenga, G.L. Fahnenstiel, J.R. Liebig, and W.S. Gardner. Plankton ecology in an ice-covered bay of Lake Michigan: utilization of a winter phytoplankton bloom by reproducing copepods. *Hydrobiologia,* 243–244, 175–183. (DOI:10.1007/BF00007033) (1992).

Wang, J., J. Kessler, X. Bai, A. Clites, B. Lofgren, A. Assuncao, J. Bratton, J, P. Chu, and G. Leshkevich. Decadal Variability of Great Lakes Ice Cover in Response to AMO and PDO, 1963–2017. *Journal of Climate*, *31*(18), 7249–7268. (DOI:10.1175/JCLI-D-17-0283.1) (2018).

Watson, S.B. et al. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. *Harmful Algae*, 56: 44-66. (DOI:10.1016/j.hal.2016.04.010) (2016).

Zhang, H., J. Wang, T.-y.. Yang, B. Lofgren, and P. Chu. Statistical relationships between biological parameters and environmental forcings in Lake Erie, 1970s–2010s. NOAA Technical Memorandum GLERL-173, 69pp. https://www.glerl.noaa.gov/pubs/tech_reports/glerl-173/tm-173.pdf (2018).

Zhou Y.T., A.M. Michalak, D. Beletsky, Y.R. Rao, and R.P. Richards. Record-Breaking Lake Erie Hypoxia during 2012 Drought. *Environmental Science and Technology*, 49:800 807. (DOI:10.1021/es503981n) (2015).