Assessing the Effectiveness of the Neuse Nitrogen TMDL Program and Its Impacts on Estuarine Chlorophyll Dynamics

by

Ibrahim Alameddine

Department of Environment
Duke University

Date: ______________________

Approved:

___________________________
Kenneth H. Reckhow, Supervisor

___________________________
Song S. Qian

___________________________
Dean L. Urban

___________________________
Hans W. Paerl

___________________________
Jonathan L. Goodall

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Environment in the Graduate School of Duke University

2011
ABSTRACT

Assessing the Effectiveness of the Neuse Nitrogen TMDL Program and Its Impacts on Estuarine Chlorophyll Dynamics

by

Ibrahim Alameddine

Department of Environment
Duke University

Date: ______________________
Approved:

___________________________
Kenneth H. Reckhow, Supervisor

___________________________
Song S. Qian

___________________________
Dean L. Urban

___________________________
Hans W. Paerl

___________________________
Jonathan L. Goodall

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Environment in the Graduate School of Duke University

2011
Abstract

Coastal eutrophication is a complex process that is caused largely by anthropogenic nutrient enrichment. Estuaries are particularly susceptible to nutrient impairment, owing to their intimate connection with the contributing watersheds. Estuaries experiencing accelerating eutrophication are subject to a loss of key ecological functions and services. This doctoral dissertation presents the development and implementation of an integrated approach toward assessing the water quality in the Neuse Estuary following the implementation of the total maximum daily load (TMDL) program in the Neuse River basin. In order to accomplish this task, I have developed a series of water quality models and modeling strategies that can be effectively used in assessing nutrient based eutrophication. Two watershed-level nutrient loading models that operate on a different temporal scale are developed and used to quantify nitrogen loading to the Neuse Estuary over time. The models are used to probabilistically assess the success of the adopted mitigation measures in achieving the 30% load reduction goal stipulated by the TMDL. Additionally, a novel structure learning approach is adopted to develop a Bayesian Network (BN) model that describes chlorophyll dynamics in the Upper Neuse Estuary. The developed BN model is compared to pre-TMDL models to assess any changes in the role that nutrient loading and physical forcings play in modulating chlorophyll levels in that section of the estuary. Finally, a set of empirical models are developed to assess the water quality monitoring program in the estuary,
while also exploring the possibility of incorporating remotely sensed satellite data in an effort to augment the existing in-situ monitoring programs.
Contents

Abstract............................................................................................................................... iv

List of Tables ....................................................................................................................... xi

List of Figures ...................................................................................................................... xiii

Acknowledgements ............................................................................................................ xx

1. Introduction ..................................................................................................................... 1

   1.1 Dissertation Objectives and Organization ................................................................. 4

2. Downscaling the SPARROW Model to the Watershed Level: the Need for a Bayesian Dynamic Model ................................................................. 8

   2.1 Introduction ............................................................................................................... 8

   2.2 Methods ................................................................................................................... 13

      2.2.1 Study Area and Data Sources .............................................................................. 13

       2.2.1.1 Stream Network, Sub-Watersheds, and Monitoring Stations .................. 14

       2.2.1.2 Total Nitrogen Loading ................................................................................. 15

       2.2.1.3 Nitrogen Sources ........................................................................................ 17

           2.2.1.3.1 Point Sources ....................................................................................... 18

           2.2.1.3.2 Non-Point Sources .............................................................................. 20

       2.2.1.4 Attenuation Data .......................................................................................... 26

           2.2.1.4.1 Landscape Attenuation ....................................................................... 26

           2.2.1.4.2 Aquatic Attenuation .......................................................................... 27

   2.2.2 The SPARROW Model ......................................................................................... 28
2.2.3 The Neuse Bayesian SPARROW ................................................................. 31
  2.2.3.1 Synthetic Data Generation and Model Fitting ................................. 32
  2.2.3.2 Model Fitting with Annual Data from the Neuse River Basin ............. 36
2.3 Results ........................................................................................................... 41
  2.3.1 Synthetic Data: the Value of Monitoring and the Curse of Dimensionality. 41
  2.3.2 Neuse Data: Dynamic Updating over Time .......................................... 53
2.4 Discussion ..................................................................................................... 64
  2.4.1 Bayesian Framework: Model Regionalizing ........................................ 64
  2.4.2 Bayesian Framework: Temporal Updating .......................................... 66
  2.4.3 Parameter Correlations and Model Inference ..................................... 67
  2.4.4 Increased Monitoring: What Do We Gain? ........................................ 68
3. Developing a Bayesian Changepoint-Threshold Flow Concentration Model .... 70
  3.1 Introduction .................................................................................................. 70
  3.2 Materials and methods .............................................................................. 74
    3.2.1 Study Area and Data ........................................................................... 74
    3.2.2 Model Development ......................................................................... 77
    3.2.3 Calculating TON Load Reductions ................................................... 83
  3.3 Results .......................................................................................................... 86
  3.4 Discussion and Conclusions ....................................................................... 94
    3.4.1 Dynamics of the Flow-Concentration Relationship .......................... 94
    3.4.2 The Effect of Water Temperature on TON Concentrations ................... 95
    3.4.3 Assessing Load Reductions and Their Management Implications .......... 96
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.4</td>
<td>Bayesian Framework, Adaptive Management, and Monitoring</td>
<td>100</td>
</tr>
<tr>
<td>4.</td>
<td>Evaluating Automated Structure Learning with Bayesian Networks: Chlorophyll Dynamics in the Upper Neuse Estuary</td>
<td>103</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>103</td>
</tr>
<tr>
<td>4.2</td>
<td>Materials and Methods</td>
<td>107</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Study Area</td>
<td>107</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Data Description</td>
<td>108</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Structure Learning Algorithms</td>
<td>110</td>
</tr>
<tr>
<td>4.2.3.1</td>
<td>PC-Algorithm</td>
<td>113</td>
</tr>
<tr>
<td>4.2.3.2</td>
<td>NPC-Algorithm</td>
<td>117</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Defined Structure Learning Scenarios</td>
<td>118</td>
</tr>
<tr>
<td>4.3</td>
<td>Results</td>
<td>119</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Generated Model Structures</td>
<td>119</td>
</tr>
<tr>
<td>4.3.2</td>
<td>BN Model Evaluation</td>
<td>124</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Model Interpretation: Post-TMDL Chlorophyll Dynamics</td>
<td>127</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion and Conclusions</td>
<td>129</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Structure Learning</td>
<td>129</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Chlorophyll Dynamics in the UNRE</td>
<td>132</td>
</tr>
<tr>
<td>5.</td>
<td>Optimizing an Estuarine Water Quality Monitoring Program Through an Entropy-Based Hierarchical Spatio-Temporal Bayesian Framework</td>
<td>138</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>138</td>
</tr>
<tr>
<td>5.2</td>
<td>Methods</td>
<td>142</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Study Area and Data Description</td>
<td>142</td>
</tr>
</tbody>
</table>
5.2.2 Bayesian Hierarchal Model ................................................................. 148
5.2.3 Entropy ......................................................................................... 150
  5.2.3.1 Total System Entropy .............................................................. 151
  5.2.3.2 Standard Violation Entropy .................................................... 152
5.2.4 The AHP Framework ...................................................................... 153
5.3 Results ............................................................................................. 155
  5.3.1 Optimizing Based on Total System Entropy .................................... 158
  5.3.2 Optimizing Based on Standard Violation Entropy ............................ 160
  5.3.3 AHP Based Results: A Compromise Solution ............................... 168
5.4 Discussion ......................................................................................... 171
  6.1 Introduction ..................................................................................... 173
  6.2 Methods ......................................................................................... 177
    6.2.1 Study Area and Data Acquisition ............................................. 177
      6.2.1.1 In-Situ Data ...................................................................... 179
      6.2.1.2 Satellite Processing ............................................................ 183
    6.2.2 Empirical Ocean Color Algorithms .......................................... 186
      6.2.2.1 Global Blue:Green Based Empirical Algorithms .................. 186
      6.2.2.2 Regional Blue:Green Based Empirical Algorithms ............. 188
      6.2.2.3 Near-Infrared:Red Based Empirical Algorithm .................... 190
      6.2.2.1 Red:Green Empirical Algorithm ....................................... 192
      6.2.2.1 MARS Model ................................................................. 192
List of Tables

Table 1: The nitrogen sources that were considered for the Bayesian Neuse SPARROW model ................................................................. 18

Table 2: SPARROW model parameters that were adopted to generate synthetic model data .............................................................................. 33

Table 3: Vague priors defined for the Neuse Bayesian SPARROW model that was fit with synthetic data .................................................................. 35

Table 4: Semi-informative priors defined for the Neuse Bayesian SPARROW model .................................................................................. 39

Table 5: Model fit as a function of the number of monitoring stations and their locations ............................................................................. 43

Table 6: Posterior model summary statistics for the TON flow-concentration changepoint-threshold Bayesian model ........................................ 88

Table 7: Summary of the main steps of the PC algorithm ......................................................................................................................... 115

Table 8: Defined scenarios to test the sensitivity of the constraint based structure learning algorithms to several user defined options ...................... 119

Table 9: Score results for the different models generated for the five defined scenarios in Table 8. Scores are calculated in terms of the AIC, BIC, and log-likelihood score ..... 125

Table 10: The marginal correlation matrix between the currently monitored stations in the ModMon program. Station 0 is located towards the beginning (river section) of the estuary, while Station 180 is the last station currently monitored before the Neuse Estuary opens to Pamlico Sound (Figure 30). Spatial correlations are evidently stronger in the lower part of the estuary. Grey shaded cells indicate correlations > then 0.5. .... 157

Table 11: The ferry routes that are used to collect the FerryMon water quality monitoring data Adapted from (Buzzelli et al. 2003) ......................................................... 180

Table 12: SeaWiFS spectral channels along with their typical uses .............................................................................................................. 184

Table 13: SeaWiFS global Blue:Green based empirical algorithms. The $R_{\text{RS}}(\lambda)$ is the reflectance at the specified wavelength, $\lambda$ ........................................................................... 188
Table 14: Regional Blue:Green based chlorophyll algorithms used with the SeaWiFS sensor

Table 15: Ordinary least squared regression based regionalization of the OC4 algorithm to the Pamlico Sound

Table 16: The linear regression model describing the relationship between the in situ fluorescence measurements from the FerryMon program and the Red:Green ratio from SeaWiFS
List of Figures

Figure 1: The 8 MRBs defined by the USGS NAWQA Program (Adapted from Seaber et al. (1987)) ........................................................................................................................................... 11

Figure 2: Map of the Neuse River basin and the 13 water quality monitoring stations. Also shown on the map are the major urban areas within the basin. The stream network is based on the ERF1.2 ........................................................................................................................................... 15

Figure 3: NLCD coverage in the Neuse River basin 1992 and 2001 ........................................................................................................... 21

Figure 4: Correlations between annual fertilizer application, manure application, and inorganic nitrogen deposition in the Neuse River basin. The data are for the year 2001. Similar patterns were observed across all years. ρ is the Spearman’s correlation coefficient. ........................................................................................................................................... 23

Figure 5: Agricultural density (in % landcover) at the HUC 12 watershed level for the Neuse River basin based on the 2001 NLCD LULC. Also showing on the map are the locations of the CAFO facilities in the basin ........................................................................................................................................... 25

Figure 6: The prior on the “agricultural” source term, which is the weighted sum of the prior distributions on β₁ (Fertilizer source coefficient), β₂ (Manure source coefficient), and β₃ (Atmospheric deposition source coefficient). Also shown are the prior distributions on each of the three terms ........................................................................................................... 38

Figure 7: Model fit for the SPARROW model using synthetic data as a function of the number of monitoring stations placed on first order streams within the Neuse River basin. Reported in each of the panels is the corresponding R² ........................................................................................................................................... 44

Figure 8: Model fit for the SPARROW model using synthetic data as a function of the number of monitoring stations placed on >1 order streams within the Neuse River basin. Reported in each of the panels is the corresponding R² ........................................................................................................................................... 45

Figure 9: Variations of the model coefficients with respect to the number and order of the monitoring stations used to fit the model. The blue lines are the originally set values that were used to generate the synthetic data. The unfilled circles represent the mean posterior values of the parameters. The thick horizontal segments correspond to the intervals associated with ±1 standard deviation. The thin segments show the intervals associated with ±2 standard deviation away from the mean. ........................................................................................................................................... 47
Figure 10: Bivariate plots showing the correlations between $\alpha$, $\beta_1$, $\beta_2$, and $\beta_{\text{point}}$ from the SPARROW model. Plotted values are based on the posterior distributions associated with the 13 existing monitoring stations. Spearman's rank correlation coefficients are reported in each of the panels. The grey shading in each the panel is proportional to the bivariate density. The red dot shows the location in 2-dimentional space of the “true” parameters in each panel. ................................................................. 50

Figure 11: Bivariate plots showing the correlations between $\alpha$, $\beta_1$, $\beta_2$, and $\beta_{\text{point}}$ from the SPARROW model. Plotted values are based on the posterior distributions associated with 48 monitoring stations (13 existing +35 1st order new stations). Spearman's rank correlation coefficients are reported in each of the panels. The grey shading in each the panel is proportional to the bivariate density. The red dot shows the location in 2-dimentional space of the “true” parameters in each panel. ................................................................. 51

Figure 12: Bivariate plots showing the correlations between $\alpha$, $\beta_1$, $\beta_2$, and $\beta_{\text{point}}$ from the SPARROW model. Plotted values are based on the posterior distributions associated with 48 monitoring stations (13 existing +35 higher order new stations). Spearman's rank correlation coefficients are reported in each of the panels. The grey shading in each the panel is proportional to the bivariate density. The red dot shows the location in 2-dimentional space of the “true” parameters in each panel. ................................................................. 52

Figure 13: Three-dimentional scatterplot showing the parameter space shared by $\alpha$, $\beta_1$, and $\beta_2$. In the right panel the values are based on the posterior distributions associated with data from the 13 existing monitoring stations. The panel on the right shows the results based on 48 stations (13 existing + 35 order 1 stations). The color gradient is based on the values of $\alpha$. Note that the scales are set to be the same in both plots. The blue asterisk shows the location in 3-dimntional space of the “true” parameters. ....... 53

Figure 14: The development of the posterior distributions of the source coefficients ($\beta_1$, $\beta_2$, and $\beta_{\text{point}}$) and the land-to water delivery factor ($\alpha$) over time. The black dots represent the posterior median values; the gray vertical segments are ± 1 standard deviations from the median; the black dashed horizontal line represents the mean value reported for the Southeast SPARROW model. ................................................................. 57

Figure 15: The development of the aquatic attenuation coefficients (Ks.u, Ks.b, Ks.m, and Kr) over time. The black dots represent the median values; the gray vertical segments are ± 1 standard deviations from the median; the black dashed horizontal line represents the mean value reported for the Southeast model. ................................................................. 59

xv
Figure 16: Predicted versus observed log TN loading in the Neuse as a function of time. The red solid line is the 1:1 line, which indicates a perfect fit. The solid points are the mean log TN loads; the dark grey vertical segments indicate the log load ± 1 standard deviation (68% credible interval); the lighter grey vertical segments represent the log load ± 2 standard deviation (95% credible interval).

Figure 17: The development of the posterior Gamma distribution on the model precision (τ = 1/σ²) over time. The dashed line represents the density of the semi-informative prior we initially set on the precision: Gamma(shape=2, scale=2).

Figure 18: Predicted loads versus observed TN loads in the Neuse based on the Southeast TN SPARROW model (Hoos and McMahon 2009). The red solid line is the 1:1 line, which indicates a perfect fit. The solid points are the mean log TN loads.

Figure 19: Spatio-temporal variability of the residuals from the Neuse SAPRROW model.

Figure 20: Neuse River Basin showing the major urban areas, the river network, as well as the town of New Bern, where the Neuse River opens to form the Neuse Estuary before discharging in Pamlico Sound. The image in the small frame shows the Neuse River Basin within the context of the State of North Carolina.

Figure 21: Boxplot of TON concentrations in natural logarithm scale from 1979 till 2008 for the Fort Barnwell station. Grey line indicates a locally-weighted polynomial regression LOWESS curve that traces TON concentrations across time. Sampling frequency for each year is also illustrated in the inverted histogram.

Figure 22: The adopted simulation methodology for estimating the probability distribution for % of TON load reduction goal for a given flow. In panel (a), a flow is drawn from the Fort Barnwell hydrograph. In this example the drawn flow corresponds to 231.48 m³/sec. Panel (b) shows the posterior predictive probability distribution of TON for both pre and post 1999 that correspond to a flow event of 231.48 m³/sec. Panel (c) presents the histogram of the % TON reduction given the pre-1999 and post-1999 distributions in panel (b).

Figure 23: Flow- TON relationship both for pre and post the 1999 system changepoint. Reported TON concentrations have been adjusted for the effect of water temperature. The relationship pre 1999 was linear and dilution dominated. Post 1999, the relationship is piecewise linear with a threshold in flow at 30.429 m³/sec.
Figure 24: Percent reduction in TON levels as a function of flow. Black solid points indicate the predicted mean change in TON levels between pre and post 1999 levels for the defined flow measurements. The dark grey bands correspond to ± 1 standard error credible interval, while the light grey bands indicate ± 2 standard error credible intervals. Dashed black line shows the TMDL stipulated reduction in nitrogen loading reaching the Neuse. Gaps towards the high flow values are a direct result of sampling from a right-tailed distribution. ................................................................. 93

Figure 25: The locations along the Neuse Estuary of Station 30 from the Modmon monitoring program as well as the USGS gauge station in relation to the city of New Bern, NC. ................................................................. 109

Figure 26: Matrix scatterplot of ModMon selected variables for Station # 30 with the corresponding histograms. The scatterplot plots each variable against the other and fits a locally weighted scatterplot smoothing (LOESS) line to help discern the underlying relationship between each pair. The histograms illustrate the distribution of each variable. Temperature (Temp) is plotted in °C; Nitrates and nitrites (NOx), dissolved inorganic nitrogen (DIN), Ortho-Phosphate (PO4) as well as Chlorophyll (Chl-a) are all plotted in (μg/l); Flow is plotted in (m^3/day)........................................................................ 110

Figure 27: Number of possible BNs increases at a rate that is faster than exponential with increased number of variables ........................................................................ 113

Figure 28: BN structure comparison as a function of the selected constraint based algorithm, the number of bins, the binning method, and the significance level. Note that all NPC generated models included an additional step that involved resolving ambiguous regions and selecting a directed path for the undirected arcs................................. 121

Figure 29: Temporal variations of (a) Dissolved Inorganic Nitrogen (DIN), (b) Ortho-phosphate (PO4), (c) DIN:PO4 ratio (molar), and (d) chlorophyll (Chl-a) in the Upper Neuse Estuary. The concentrations of DIN, PO4 and Chl-a are in μg/L and are log transformed. The grey shaded areas represent the summer season (from June 21 until Sep 21) of each year. The solid horizontal grey line in each graph represents the mean of the corresponding variable................................................................. 134

Figure 30: (a) Locations of the existing monitoring stations that form the ModMon monitoring program. Inter-station separation distances range between 1.5 km (between stations 60 and 70) and 9.4 km (between stations 160 and 180), (b) the 25 potential locations generated from the 5x5 km representation of the Neuse Estuary.................. 143
Figure 31: Variations in chlorophyll concentrations in the Neuse Estuary by months (a) and across the monitoring stations (b). Variations in Dissolved Oxygen (DO) levels in the Neuse Estuary by months (c) and across monitoring stations (d). The notch in each box represents the median. The solid grey area represents the interquartile range (IQR). The whiskers capture the outliers within 1.5 times the IQR. Values that are beyond the 1.5xIQR are shown as solid points in the graph. The dash-dotted lines represent the 40 µg/L and the 4mg/L standards for chlorophyll and bottom DO concentrations in North Carolina, respectively.

Figure 32: The locations of the proposed 25 locations are shown in the map. The relative size of each symbol is proportional to the computed entropy for that station. E1 is the station with the largest system entropy and as such is the most optimal site to augment the existing ModMon monitoring program. E2 and E3 are second and third in terms of entropy, respectively.

Figure 33: (a) Chlorophyll standard exceedance probability based on the 40 µg/L bottom standard at the 25 proposed locations. The size of each point is proportional to its exceedance probability; (b) standard violation entropy for the 25 proposed locations. The size of each point is proportional to its standard violation entropy value.

Figure 34: (a) DO standard violation probability based on the 4 mg/L bottom DO standard at the 25 proposed locations. The size of each point is proportional to its violation probability; (b) standard violation entropy for each of the 25 proposed locations. The size of each point is proportional to its standard violation entropy value.

Figure 35: The probabilities of violating the bottom DO standard of 4 mg/L at the two locations DO1P and DO1E across time. The former was associated with the largest violation probability, while the latter was associated with the largest DO violation entropy. The solid horizontal line represents the 50 % probability of standard violation. The dotted horizontal lines represent the 25 and 75 % chance standard violation. Note that x-axis shows the months and the years.

Figure 36: The relative importance of monitoring at the 25 proposed locations is reflected in the height of the associated vertical bars. The bar heights range from 1 to 25. The southern shoreline section that stretches between stations 160 and 180 has the highest values. The ranking is based on the three criteria we considered namely, total system entropy, Chl-a standard violation entropy, and DO standard violation entropy.
Figure 37: The Pamlico Sound in North Carolina showing the chlorophyll sampling locations by the FerryMon program and the samples that are part of the SeaBASS database. Also shown on the map are the total unique locations of the FerryMon samples, of which only a few matched with SeaWiFS scenes................................. 178

Figure 38: The distribution of chlorophyll concentrations for the combined data from FerryMon and SeaBASS. The top row shows the distributions for the entire dataset before satellite matchup. The second row shows the corresponding distributions following satellite matchup. The red vertical lines represent the mean chlorophyll value in each of the panels. Note the lognormal distribution for the observed chlorophyll concentrations in the Pamlico Sound......................................................... 182

Figure 39: SeaWiFS global and regional Blue:Green based OC chlorophyll predictions versus in-situ data. Validation statistics are reported in the lower left corner of each panel. For reference the 1:1 line is drawn in each panel. APD is the median absolute percent differences and the MR is the median ratio................................................................. 198

Figure 40: (a) The relationship between the MBR Blue:Green ratio and the OC4PSPoly predicted chlorophyll levels; the red line represents the median predicted concentrations from model and the dashed lines represent the 95 % confidence intervals; also shown is the model’s R² value. (b) The median OC4PSPoly predicted chlorophyll concentrations are plotted against the in-situ data; validation statistics are reported. (c) The relationship between the MBR Blue:Green ratio and the OC4PSPoly predicted chlorophyll levels; the red line represents the median predicted concentrations from the OC4PCLin model and the dashed lines represent the 95 % confidence intervals; also shown is the model’s R². (d) The medianOC4PSLin predicted chlorophyll concentrations are plotted against the in-situ data; validation statistics are also reported. APD is the median absolute percent differences and the MR is the median ratio. For reference the 1:1 line is also drawn in panels (b) and (d) ...................... 200

Figure 41: The NIR:Red based chlorophyll model fit to the combined data (left panel), the SeaBASS data (middle panel), and the FerryMon data (right panel). The red lines represent the median predicted chlorophyll values as a function of the NIR:Red ratio. The dashed lines show the 95% confidence intervals............................................................... 203

Figure 42: The Red:Green model based chlorophyll concentrations plotted against the SeaBASS data (left panel), and the FerryMon data (right panel). The red lines represent the median predicted chlorophyll values as a function of the Red:Green ratio. The dashed lines show the 95% confidence intervals............................................................... 204
Figure 43: The Red:Green chlorophyll model results for the combined data. In the left panel the red line shows the median predicted chlorophyll values as a function of the Red:Green ratio. The dashed lines show the 95% confidence intervals. In the right panel, the median predicted chlorophyll concentrations are plotted against the in-situ data. Validation statistics are also reported. APD is the median absolute percent differences and the MR is the median ratio. For reference the 1:1 line is also drawn in the panel to the right.

Figure 44: Relationship between in-situ fluoresce measurements collected from the FerryMon program and the Red:Green ratio from SeaWiFS (with the potential outliers)

Figure 45: The results from the MARS generated hybrid model. In the left panel, the line shows the predicted median chlorophyll concentrations as a function of the Blue:Green ratio while holding the Log_{10}(Red:Green) at its median value. In the right panel, the line shows the predicted median chlorophyll concentrations as a function of the Red:Green ratio while holding the Log_{10}(Blue:Green) ratio at its median value. In both panels rugs are added above the x-axis in order to shows the distribution of the data with respect to the predictor.

Figure 46: The MARS based model median chlorophyll concentrations plotted against the in-situ data; validation statistics are reported for the model fit based on the SeaBASS and FerryMon datasets. Also shown is the $R^2$. A 1:1 line is added for reference.
Acknowledgements

I would like to express my deepest gratitude and appreciation to my committee members: Ken Reckhow (advisor), Song Qian, Dean Urban, Hans Paerl, and Jon Goodall. Thank you, Ken, for your trust, insight, and guidance; without them much of this work would not have been possible. You were always there to make sure that I did not lose sight of the forest for the trees and to encourage me to explore new research paths and seek new collaborations. I’m very grateful to Song for working closely with me to solve what often appeared as unsurpassable roadblocks and for providing indispensable suggestions throughout my six years at the Nicholas School. Song your door was always open, and I will always cherish our long insightful discussions. Thank you, Dean, for your support and encouragement; it has been a great pleasure having you on my committee. Thank you Hans for sharing your extensive knowledge of the study area, which helped me better appreciate the complexity of coastal eutrophication and to venture beyond the modeling confines of my work. Your commitment towards maintaining an extensive and unique monitoring system in the Neuse was crucial for the success of this work. Finally, thank you, Jon, for working closely with me on the SPARROW model and for teaching me how to effectively use GIS in my research. Working with you and with Jahangir Alam was a very productive partnership that I will always cherish.
Many thanks to the United States Environmental Protection Agency (USEPA), Office of Research and Development’s Advanced Monitoring Initiative (AMI) Pilot Projects, for supporting and funding this work. I am also very grateful to the scholarship support provided by the Quantitative Environmental Analysis (QEA), LLC and for the summer funding and travel awards provided by the Nicholas School of the Environment. I am also thankful to Hans Paerl and the University of North Carolina-Institute of Marine Sciences’ ModMon Program for providing me with the monitoring data for the Neuse and Pamlico Sound.

I would also like to thank my colleagues and friends in the Reckhow Lab (both current and former) for many fruitful discussions, collaboration, and many happy memories. I will always treasure the time I spent with you all. Thank you YoonKyung Cha, Andrew Gronewold, Yun Jian, Melissa Kenney, Roxolana Kashuba, Conrad Lamon, Boknam Lee, Michaela Margida, Eric Money, Farnaz Nojavan, and Kriss Voss. I would also like to thank my non-Bayesian friends at the Nicholas School in particular Muhan Cheng, NamHee Kim, Soumya Hassan, Rob Schick, Sathya Gopalakrishnan, Shaleyla Kelez, John Kim, Joe Sexton, Akiko Haruna, Alex Konings, and Ben Best. Your friendship added much flavor to these past six years.

I would also like to extend my sincerest appreciation to all my collaborators. Thank you Reckhow Lab members, Ben Best, Tom Cuffney, Mark Borsuk, Russell Gosnell, Subhankar Karmakar, Jahangir Alam, Jerry McMahon, Ryan O’Banion, Conrad Lamon,
and Craig Stow. Working with you all has contributed to my growth as a scientist and expanded my research interests beyond the confines of my dissertation.

I deeply appreciate the efforts of the Duke University Graduate School and the Nicholas School of the Environment. Thank you, Meg Stephans, you were always there to help me navigate through the administrative bureaucracy. Many thanks go to Gaby Katul, the Director of Graduate Studies for the Environmental Science and Policy Program at the Nicholas School. His friendship, help, and guidance were instrumental throughout my six years at the school. Thank you Nicholas OIT for your valuable technical support across the years.

Last but not least, I am forever indebted to my family and friends that have supported me throughout my dissertation. Several deserve special mention. Thank you mom (Sana) and dad (Adnan) for believing in me and for your endless support and boundless love. I would have never been able to complete my journey without your help, guidance, and encouragement. Many thanks to my loving sister, Dina, who challenged me to graduate before my nephew, Karim, started his own academic journey. I hope his will be as enjoyable and productive as mine. Many thanks to my uncles and aunties both in Lebanon and the US; you were always there for me and never hesitated to go out of your way to support me. Thank you to my fellow Alumni in the American University of Beirut-North Carolina Chapter for your invaluable support and sense of community. Special thanks go out to Nuhad Abou Zeid, Rami Baderddine, Gabi
Hanna, Shireen Wehbe, Raghid Bitar, Rania Kazan, Wissam Kourani, Louisa Baroudi, Fares and Zeina Hanna, and Walid and Ghada Ramadan for being my family away from home.
1. Introduction

Coastal eutrophication is a complex process that is primarily caused by anthropogenic nutrient enrichment. Currently, excessive nutrient loading is one of the main sources of water body impairment in the United States (US) (Environmental Protection Agency 2009). While our ability to quantify changes in nutrient delivery at the global and continental scale has improved significantly over the years, our understanding of these changes at the river basin scale remains a challenge (Alexander et al. 2002a, Galloway et al. 2003, Schlesinger 2007, Galloway et al. 2008, Canfield et al. 2010).

Estuaries are susceptible to nutrient impairment, which often results in a loss of important ecological functions and services. Eutrophic estuaries have been known to experience fishkills, harmful algal blooms, changes in the water’s biogeochemistry, loss of biodiversity, demise of seagrasses, and the spread of dead zones (hypoxic) (Nixon 1995, Jørgensen and Richardson 1996, Bricker et al. 1999a, Borsuk et al. 2001b, Cloern 2001, Livingston 2001, Environmental Protection Agency 2007, Paerl et al. 2007, Scavia and Donnelly 2007, Boesch et al. 2009, Conley et al. 2009, Liu et al. 2010). In an effort to protect these sensitive systems from these adverse impacts, the National Estuary Program (NEP) was established as part of the 1987 amendments to Section 320 of the Clean Water Act (Environmental Protection Agency 2007). The program aimed at developing plans for improving or maintaining water quality in estuarine waters,
conducting habitat restoration activities when deemed necessary, as well as supporting
data collection as part of the United States Environmental Protection Agency’s (USEPA's)
National Coastal Assessment program. Since 1987, 28 different estuaries have been
added to the NEP. One of the first estuaries to be designated as “an estuary of national
significance” in the NEP was the Albemarle-Pamlico Estuarine Complex, which includes
the Neuse Estuary in North Carolina.

The Neuse Estuary experienced the classic symptoms associated with
eutrophication in the 1990’s as a direct result of nutrient overloading from upstream
activities, including algal blooms, extensive bottom water hypoxia, fish kills and habitat
loss (Stanley 1988, Paerl et al. 1995, Pinckney et al. 1997). While the process of
eutrophication in the estuary goes beyond just excessive nutrient loading, it has been
well established that the eutrophication in the Neuse was primarily nitrogen driven
(Rudek et al. 1991, Paerl et al. 1995). Over the past two decades the system has
experienced large scale changes that included the implementation of basin-level
management actions, rapid changes in landuse and landcover, and extreme climatic
2006b). Possibly the most significant of the basin-scale management measures in the
Neuse has been the adoption of the mandatory nutrient delivery control plan by the
state of North Carolina in 1997, which stipulated a 30 % reduction in the amount of
nitrogen delivered to the Neuse Estuary based on mean 1991-1995 loads (NC
Department of Environment and Natural Resources 1999). While this basin level
management action has been credited with decreasing the frequency of excessive algal growth and fishkills in the Neuse, there has been little consensus as to whether the adopted mitigation measures have achieved the Total Maximum Daily Load (TMDL) set goals (Deamer 2009).

In light of these changes and the spatio-temporal variability of water quality in a dynamic system like the Neuse basin and estuary, there is a need to support extensive monitoring activities. Yet, in-situ monitoring networks are expensive to maintain and can only capture what is happening at a given point in space. Augmenting the existing monitoring programs with remotely-sensed data may help us gain a better understanding of these systems without incurring additional costs. The National Aeronautics and Space Administration’s (NASA’s) ocean color sensor, the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), provides a rich time series of chlorophyll data at the global scale starting with the year 1997. The integration of in-situ and remotely sensed data can allow us to better quantify the interaction of anthropogenic and physical forcings and assess how these interactions alter the natural dynamics in the Neuse Estuary. Nevertheless, it is also important to recognize that the use of remotely sensed data in optically complex waters has proven to be difficult, uncertain, and lacking robustness. Therefore, it is important to continue with in-situ based monitoring, whose role and informational content need to be reexamined over time in order to identify redundancies and locations with high levels of uncertainties.
1.1 Dissertation Objectives and Organization

My dissertation is divided into six chapters. The first chapter describes the main impetus behind my doctorate research and presents the overall research objectives. In Chapter 2, I present a dynamic regionalized Bayesian SPAtially Referenced Regression On Watershed Attributes (SPARROW) nutrient loading model that I developed for the Neuse River basin using input data spanning more than a decade. The model is capable of accounting for basin-scale changes over time and implements a novel regionalization approach, which nests the model within a regional large-scale SPARROW model that has been calibrated for the entire Southeastern (SE) United States (US). I also made use of the Bayesian features of the model to sequentially update it, thus overcoming the static nature of conventional SPARROW models. The initial findings in Chapter 2 highlighted the differences between the regional SE model and the Neuse model, which indicates a growing need towards the development of better downscaling approaches. This work also provided important insights towards the inner workings of the SPARROW model through an in-depth assessment of parameter correlations that was conducted on synthetically simulated data. The developed framework in this chapter will assist the SPARROW model to fully realize its potential as an invaluable tool within the TMDL program. Much of the work was developed in collaboration with Dr. Song Qian, Dr. Jonathan Goodall, Mr. Jahangir Alam, and Ms. Yun Jian.

In Chapter 3, I developed a Bayesian multi-level flow-concentration model for the Neuse basin that incorporated a changepoint in time and did not assume a static
relationship between river flow and nutrient loading rates. The results identified a
temporal shift in the flow-concentration relationship in the Neuse River around the year
1999, which was highly correlated with the TMDL implementation. Additionally, the
modeling effort was used to generate a distribution on the achieved nitrogen load
reductions. The results suggested that there is strong evidence to indicate that the
stipulated 30% reduction in nitrogen loading has been achieved. Additionally, the
analysis highlighted the growing impact of non-point sources in nitrogen delivery to the
system, as the regulations on point sources have been strengthened. Much of the work
(and text) in Chapter 3 was developed in collaboration with Dr. Song Qian, and Dr.
Kenneth Reckhow and was recently published in Water Research (Alameddine et al.
2011b).

In Chapter 4, I used automated constraint based structure learning algorithms to
build a post-TMDL chlorophyll Bayesian Network (BN) model for the Upper Neuse
Estuary. Different model structures were generated based on different assumptions
concerning the adopted structure learning framework. I used these differences in model
structures to examine the functionality and usefulness of structure learning as a means
towards building model topology. Furthermore, I proposed the use of a hybrid approach
towards structure learning whereby expert defined constraints are incorporated in
conjunction with the automated structure learning algorithms. This proposed approach
would generate plausible model structures and minimize the sensitivity of the learning
algorithms. This analysis also showed that the effects of the physical environment
remained the primary driver of chlorophyll dynamics in the study area. Yet, the response of chlorophyll levels to nutrient concentrations was found to have been slightly altered. Much of the work (and text) in Chapter 4 was developed in collaboration with Yoonkyung Cha, and Dr. Kenneth Reckhow. This work was recently published in Environmental Modelling and Software (Alameddine et al. 2011a).

In Chapter 5, I developed a Bayesian hierarchical spatio-temporal model that uses entropy to quantify the informational gain/loss that is associated with the redesign of a water quality network. The model was assessed using the existing water quality monitoring network operating in the Neuse Estuary. Multiple design criteria were defined in order to identify the locations with maximum uncertainty. A Multiple Attribute Decision Making (MADM) framework was implemented to compare different ungauged sites and to incorporate the information elicited from experts concerning the relative weights that need to be assigned to each criterion. The optimization results emphasized the difference between a probability and an entropy based design, while highlighting the advantages of adopting the latter as a monitoring design criterion.

Finally, in Chapter 6 I focus on exploring the potential use of NASA’s SeaWiFS remotely sensed ocean color time series for predicating chlorophyll levels in the Neuse Estuary and Pamlico Sound. The results indicated that the sensor was unusable for estimating chlorophyll levels in the Neuse estuary primarily due to the coarse spatial resolution of the SeaWiFS sensor. Yet, the analysis showed promise for its use in the Pamlico Sound. Different empirical ocean color models were explored and assessed. The
results indicated that predictions based on globally calibrated models were biased, lacked robustness, and failed to recreate the chlorophyll dynamics observed in the collected in-situ data. Moreover, regionally calibrated models showed more promise. Furthermore, the research in Chapter 6 highlighted the limitations of the algorithms that were based on the Blue:Green reflectance ratios as compared to models that made use of the Red:Green ratio. Our findings also indicated a poor performance for the Near-Infrared:Red optically based algorithm for predicting chlorophyll levels in the Pamlico Sound.
2. Downscaling the SPARROW Model to the Watershed Level: the Need for a Bayesian Dynamic Model

2.1 Introduction

Changes in landuse and landcover (LULC), technological innovations, human behavior, as well as the implementation of environmental management programs have all resulted in significant alterations to the water quality of receiving river basins, lakes, and coastal waters. This is particularly true for nutrient related pollution, which is one of the main sources of water body impairment in the United States (US). Currently nutrient impairment ranks fourth on the national impairment list, of which more than 17 % have nitrogen explicitly listed as the cause of impairment (Environmental Protection Agency 2009). Estuaries are particularly susceptible for nitrogen impairments; out of the 44 estuaries with accelerating eutrophication 63% had high to moderate nitrogen inputs (Bricker et al. 1999b).

Anthropogenic increases in reactive nitrogen loads reaching the aquatic environment have been caused by changes in land management, growing fossil fuel use, production of fertilizer, and human development across time (Galloway et al. 2003, Smith et al. 2003, Schlesinger 2007, Galloway et al. 2008, Taylor and Townsend 2010). Rivers and streams play an important role in delivering nitrogen from the land to the coastal waters. In the year 2000, the world rivers were estimated to have exported 43 Teragrams of nitrogen to the coastal oceans (Seitzinger et al. 2010). This nitrogen load has resulted in negative changes to the water quality of the receiving water bodies,
leading to pronounced coastal eutrophication events. Our ability to understand and quantify the effects of nitrogen pollution on the global and continental scales has improved significantly over the years (Galloway et al. 2003, Schlesinger 2007, Galloway et al. 2008, Canfield et al. 2010). Yet, understanding these changes at the river basin scale remains challenging (Alexander et al. 2002a). This is particularly true for the Neuse River Basin, which has experiences in the past two decades significant changes in its LULC, increased economic development, as well in the enactment of progressive environmental regulations that aim at limiting nutrient related pollution.

The need for a model that can link the water quality of our rivers to changes in their contributing watersheds and to enacted basin scale management actions is crucial. In this study, we explore the use of the semi-empirical SPARROW model (Smith et al. 1997) to characterize nitrogen loading in the Neuse River basin. SPAROW is an annual nutrient loading model that has been widely used both nationally and internationally (Preston and Brakebill 1999, Alexander et al. 2000, Alexander et al. 2002b, McMahon et al. 2003, Elliott et al. 2005, Alexander et al. 2008a). SPARROW relates observed in-stream water quality measurements to corresponding watershed attributes; a feature that is particularly important from a management perspective. The model also takes into account the origin, fate, and transport of the pollutants as they travel over land and within the river network. The model has mass-balance constraints and is capable of accounting for non-conservative transport processes (Alexander et al. 2008a). These
features make SPARROW a very useful management tool with great relevance to the TMDL program (National Research Council 2001).

SPARROW has shown great promise in linking watershed data to water quality conditions in rivers and streams across the US. It has been successfully used to simulate load reductions in support of the TMDL program and to identify locations in need of future monitoring (McMahon and Roessler 2002, McMahon et al. 2003, Robertson et al. 2009).

SPARROW was initially developed to be used on the national scale; yet recent versions have focused on regionalizing the model to specific contiguous areas. While the national SPARROW models are informative when it comes to estimating continental loading rates, they often have limited use when it comes to assessing water quality at the watershed level. This is due to the fact that the national model parameterization does not account for local characteristics. This limitation can lead to potentially uncertainties in model predictions at the local scale.

Recognizing this shortcoming, the Unites States Geological Survey (USGS) National Water Quality Assessment (NAWQA) program has embarked on developing regional scale nitrogen and phosphorus SPARROW models for the 8 Major River Basins (MRB) in the US (Figure 1). Currently, several regionalized SPARROW models have been developed. These include the Southeast nitrogen SPARROW (Hoos and McMahon 2009), the Chesapeake Bay suspended sediment SPARROW (Brakebill et al. 2010), and the Southeast phosphorus SPARROW (García et al. In Press). These models are better
capable of capturing regional characteristics that may affect the delivery of pollutant loads to the receiving water bodies.

Figure 1: The 8 MRBs defined by the USGS NAWQA Program (Adapted from Seaber et al. (1987))

Even with this push towards regional-scale models, the adopted spatial extent remains relatively coarse as data is pooled from multiple watersheds. This procedure assumes that the underlying pollutant transport processes remain unchanged across
space (McMahon et al. 2003). However, watersheds in a given region are often different and thus nonexchangable when it comes to nutrient delivery. River basins have different local factors including climate, physiography, ecology, geology, urbanization patterns, and agricultural practices (e.g. crop types, best management practices, irrigation schemes). Therefore, pooling results from different watersheds may result in inaccurate assessments and the implementation of ill-defined management practices within a given study area. Some implementations of SPARROW have tried to resolve this issue by the inclusion of a set of sub-region indicator coefficients (Hoos and McMahon 2009). While such an approach may be adequate to better characterize regional-scale nutrient exports, these models remain inefficient for setting and/or assessing environmental management policies at the river basin scale (e.g. TMDL program). The only way to overcome the limitation brought about by the spatial mismatch is through fitting a model at the river basin level.

For SPARROW to fully realize its full potential within the TMDL program, regionalized basin-specific models similar to the Waikato SPARROW in New Zealand (Alexander et al. 2002a) and the coastal North Carolina SPARROW models (McMahon et al. 2003, Qian et al. 2005) have to be developed. Additionally, SPARROW needs to have the capacity to account for basin-scale changes across time and to reflect these changes within the model structure itself. Unfortunately, the existing model structure is static over time.
The purpose of this study is to explore the feasibility of developing and adequately calibrating a Bayesian SPARROW model that operates at the Neuse River basin scale. Furthermore, we capitalize on the Bayesian features of the model to sequentially update it, thus overcoming the static nature of conventional SPARROW models. In this chapter, we set about accomplishing five main tasks, namely: 1) to develop the Bayesian SPARROW modeling framework and to test its accuracy through the use of synthetic data; 2) to assess the inner dynamics of the model and gain a better understanding of the effects of increasing monitoring efforts on model fit; 3) to calibrate the model based on the total nitrogen (TN) loads in the Neuse River basin using data collected from a 12 year monitoring interval across 13 monitoring stations within the basin; 4) to track and explain any changes across time with respect to the nitrogen loading and attenuation; and finally 5) to compare our results with those generated from the nitrogen Southeast model (Hoos and McMahon 2009).

2.2 Methods

2.2.1 Study Area and Data Sources

The Neuse River basin is the third largest river basin in North Carolina, draining an area approximately 16,000 km². The Neuse is also one of only three major river basins whose boundaries are entirely located within the state. The basin has a diverse LULC. Just east of its head waters, an urbanized area -that includes the cities of Raleigh, Durham, and Cary- dominates the basin. Intensive agricultural areas and Concentrated
Animal Feed Operations (CAFOs) become more prominent towards the lower portions of the basin, as the river traverses the North Carolina coastal plain. Over the years, continued landuse changes in the basin have resulted in an increase in the relative importance of non-point nutrient sources to the overall Neuse nitrogen budget. Changes in the overall LULC pattern within the Neuse River basin indicate an increase in the urban areas at the expense of a decline in the acreage of cultivated and forested lands. Fortunately, the increase in urbanization has also coincided with a push by the state to upgrade the wastewater treatment plants within the basin in order to achieve higher TN load removal rates. The major nitrogen point source emitters in the basin are the 20 wastewater treatment plants that service the cities and townships in the 19 counties that fall within the Neuse basin.

2.2.1.1 Stream Network, Sub-Watersheds, and Monitoring Stations

The drainage network for the Neuse River was constructed from the Enhanced River Reach File 2.0 (ERF1.2) dataset (Nolan et al. 2002). ERF1.2 remains the foundation for most SPARROW model applications, although there has been a recent push towards adopting the NHDPlus stream networks (Environmental Protection Agency 2006). Given the limited number of monitoring stations within the Neuse basin, the use of NHDplus was deemed not appropriate. The ERF1.2 represents the Neuse River network by 123 stream segments (Figure 2). The mean velocity in each of these segments is reported.
This allows us to estimate the mean travel time by dividing the length of the reach by its corresponding average velocity.

The size of the sub-catchments in the Neuse ranged between 1 km$^2$ and 618 km$^2$ with a median value of 57 km$^2$. The basin had 13 long-term water quality monitoring stations which are shown in Figure 2.

**Figure 2**: Map of the Neuse River basin and the 13 water quality monitoring stations. Also shown on the map are the major urban areas within the basin. The stream network is based on the ERF1.2

### 2.2.1.2 Total Nitrogen Loading

Annual TN loading rates were estimated through the use of the flow weighted load estimation method (Littlewood 1995, Moatar et al. 2006, Quilbé et al. 2006, Salles
et al. 2008). TN concentrations in the Neuse watershed were downloaded from both the STOrage and RETrieveal (STORET) system (Environmental Protection Agency 2010) and the National Water Information System (NWIS) (United States Geological Survey 2010), while the average daily stream flow data were collected from nearby USGS flow monitoring gauges (United States Geological Survey 2010). Note that we selected to only use data from stations that had an excess of four TN concentrations measurements per year. This decision was made in order to reduce loading estimation uncertainties.

Mean annual loads (in Kg/year) were computed, over the 12 year monitoring record (1990 through 2001) for each of the 13 Neuse monitoring stations, based on flow weighted concentration as shown in Equation 1 and 2. The number of valid stations per year varied between 11 and 13. In conventional SPARROW application, the number of stations has to be consistent over time; however, the Bayesian model structure that we developed is capable of handling discontinuous monitoring records over time. Missing data are simply treated as model parameters and assigned prior distributions; posterior distributions are then generated for each missing observation.

\[
\text{Loading}_{j,k} = 365 \times \overline{\text{Flow}}_{j,k} \times \text{FWC}_{j,k} \quad \text{Equation 1}
\]

\[
\text{FWC}_{j,k} = \frac{\sum_{i}^{n} C_{i,k,j} \times \text{Flow}_{i,k,j}}{\sum_{i}^{n} \text{Flow}_{i,k,j}} \quad \text{Equation 2}
\]

Where \( \overline{\text{Flow}}_{j,k} \) is the average annual flow recorded for year \( j \) at station \( k \); \( \text{FWC}_{j,k} \) is the flow weighted concentration for year \( j \) at station \( k \); \( C_{i,k,j} \) is the \( i^{th} \) nitrogen
concentrations at station $k$ for year $j$; and $Flow_{i,k,j}$ is the $i^{th}$ recorded daily flow at gauging station $k$ for year $j$ for which a concurrent water quality measurements was available. The total number of concentration measurements at station $k$ in year $j$ is $n$.

### 2.2.1.3 Nitrogen Sources

The availability of only 13 monitoring stations within the Neuse basin, compelled us to build a relatively simple, yet informative SPARROW model. More complex model structures, with a larger number of explanatory variables, are only possible with a larger number of stations. Typical regional SPARROW models often have between 5 and 8 explanatory variables, some of which are often found to be not statistically significant ($p > 0.05$) (Alexander et al. 2002a, Hoos and McMahon 2009). On the other hand, national SPARROW models are usually constructed with up to 18 variables (Smith et al. 1997, Schwarz et al. 2006).

In the case of the Neuse Bayesian SPARROW model, we accounted for TN loading from atmospheric deposition, fertilizer application, manure application, point source discharge, as well as non-point urban runoff. A summary of the TN sources that were considered for the Neuse Bayesian SPARROW model is shown in Table 1.
Table 1: The nitrogen sources that were considered for the Bayesian Neuse SPARROW model

<table>
<thead>
<tr>
<th>Annual TN source variable</th>
<th>Units</th>
<th>Reference/data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Manure application rate</td>
<td>Kg/year</td>
<td>– Animal waste from farm and nonfarm application rates (Ruddy et al. 2006)</td>
</tr>
<tr>
<td>• Fertilizer application</td>
<td>Kg/year</td>
<td>– Farm fertilizer and nonfarm fertilizer application rates (Ruddy et al. 2006)</td>
</tr>
<tr>
<td>Atmospheric deposition</td>
<td>Kg/year</td>
<td>National Atmospheric Deposition Program (National Atmospheric Deposition Program 2007)</td>
</tr>
<tr>
<td>Urban non-point</td>
<td>km²</td>
<td>Urban areas estimated from the 1992 and 2001 NLCD datasets (Homer et al. 2004, Fry et al. 2008)</td>
</tr>
<tr>
<td>Point-source</td>
<td>Kg/year</td>
<td>2002 survey on municipal and industrial wastewater emitters in the Southeastern US (McMahon et al. 2007). Annual population data from the US Census Bureau.</td>
</tr>
</tbody>
</table>

2.2.1.3.1 Point Sources

TN loads from point-source dischargers of municipal and industrial wastewater in the Neuse River basin were compiled based on the 2002 McMahon et al. (2007) estimates for annual wastewater nutrient loads in the SE US. Given that these estimates are specific for the year 2002, we accounted for the changes in loading over time by linking these estimates to the annual population data in the Neuse. Therefore, we adjusted the 2002 loads over time by taking into account changes in urban population as shown in Equation 3. This approach assumes that the annual point-source TN load is linearly related to the urban population in each sub-catchment. Such an assumption was
Deemed reasonable as most urban areas are serviced by wastewater treatment facilities.

Annual population data were compiled from county level population estimates from the United States Census Bureau. Note that Alexander et al. (2008a) had previously used spatially detailed population statistics as a surrogate for point source loading in their Mississippi River Basin SPARROW model. The decision not to use population directly as a surrogate for point source loading was made in order to account for the fact that the wastewater treatment plants may serve areas that are larger than the sub-catchments in which they are located. While the method we adopted does not totally resolve this discrepancy, the assumption that we make is more defensible and less constraining as compared to using the 2002 nitrogen point source loadings across the 12 years or to using the county based population data alone.

\[
LoadPoint_{year_{i,j}} = LoadPoint_{year_{2002,j}} \times \frac{(Population_{year_{i,j}} + 1)}{(Population_{year_{2002,j}} + 1)}
\]

Equation 3

Where \( LoadPoint_{year_{i,j}} \) and \( Population_{year_{i,j}} \) represent the point-source loads and population values for year \( i \) in sub-catchment \( j \), respectively. \( Load_{year_{2002,j}} \) are the 2002 point source loads estimated by McMahon et al. (2007) for the SE US. Note that we added 1 to the population data in order to avoid dividing/multiplying by zero, as the population data showed that some sub-catchments were uninhabited.
In all SPARROW models, the TN load from point sources are assumed to enter directly into the river segment without losses; thus it is not associated with a land-to-water delivery factor. The model coefficient on the point sources is dimensionless.

2.2.1.3.2 Non-Point Sources

2.2.1.3.2.1 Urban Nitrogen Loading

Nitrogen loading from the urban areas in the Neuse were estimated based on the urban coverage in each sub-catchment. As such, LULC data were obtained from the 1992 and 2001 National Land Cover Dataset (NLCD) coverage (Homer et al. 2004, Fry et al. 2008). The 1992 coverage was assumed to be valid for the years between 1990 and 1996, while the 2001 LULC data were adopted for the period between 1997 and 2001. Figure 3 shows the LULC in the Neuse River basin in 1992 and 2001. The LULC classes were reclassified into the major land use types (e.g., urban, forest, cropland, grassland, wetlands, and water). We used the area of urban land in each watershed as a surrogate for non-point source pollution from cities and towns.
Figure 3: NLCD coverage in the Neuse River basin 1992 and 2001
2.2.1.3.2.2 Atmospheric Deposition

Atmospheric deposition is an important source of nitrogen reaching streams and rivers (Jones et al. 2001). We accounted for deposition by considering the wet deposition of inorganic nitrogen (nitrate and ammonia) (in kg/yr) from the set of 9 National Trends Network (NTN) sites operated in North Carolina (National Atmospheric Deposition Program 2007). Annual deposition data were retrieved from the NTN stations and were used to interpolate the deposition values on a 1 km raster grid of the Neuse River basin. Interpolation was conducted through the use of the inverse-distance weighting method. The generated deposition grid was then used to generate the atmospheric deposition loading rates at the sub-catchment level. While NADP deposition sites are selected to avoid local influences resulting from urbanization, vegetation, and agriculture (National Atmospheric Deposition Program 1991), the deposition data for the Neuse showed strong correlations with fertilizer and manure application rates as seen in Figure 4.
2.2.1.3.2.3 Fertilizer and Manure

Fertilizer application is a major source of in-stream nitrogen input to the Neuse River. TN fertilizer application rates were compiled from the Association of American Plant Food Control Officials (AAPFCO), who reports annual state level fertilizer sales for the entire US. The state level annual fertilizer sales were supplemented with county level fertilizer expenditure data collected from the Census of Agriculture along with county based population estimates (Ruddy et al. 2006). This allowed us to estimate the county level farm fertilizer and nonfarm fertilizer application rates. These estimates were then used to compute fertilizer application at the sub-catchment level by

\[ \rho = \text{Spearman's correlation coefficient} \]

Figure 4: Correlations between annual fertilizer application, manure application, and inorganic nitrogen deposition in the Neuse River basin. The data are for the year 2001. Similar patterns were observed across all years.
distributing the fertilizer loads based on the total agricultural acreage and population density in each of the sub-catchments.

Similarly, manure application data were compiled from the county level livestock database maintained by the Census of Agriculture. Livestock numbers were converted to nitrogen loadings following the methodology proposed by Ruddy et al. (2006). The method accounts for both recoverable manure from confined animals and unrecoverable manure from confined and unconfined animals. The data was downscaled to the sub-catchment level by distributing the county level loads according to the agricultural acreage in each sub-catchment.

2.2.1.3.2.4 Combined Agricultural Sources

As noted previously, we found strong multicollinearity between annual fertilizer application, wet atmospheric deposition, and manure application as shown in Figure 4. The strong correlations between manure and fertilizer application is expected in the Neuse, as agricultural areas in the basin are also associated with the highest densities of CAFOs (Figure 5).
Figure 5: Agricultural density (in % landcover) at the HUC 12 watershed level for the Neuse River basin based on the 2001 NLCD LULC. Also showing on the map are the locations of the CAFO facilities in the basin.

Strong correlation between input variables can impose numerical difficulties when estimating model parameters (Stow and Scavia 2009). Under these conditions, the estimated model coefficients often turn out to be very sensitive to small changes in input data and their interpretability becomes difficult (Qian 2010). The problem of collinearity between different pollutant sources is a known concern in SPARROW models (Schwarz et al. 2006, García et al. In Press). Detailed information on the problem of multicollinearity in SPARROW as well as some of the common metrics that are used to assess its severity are discussed by in Schwarz et al. (2006).
One way to deal with strong collinearities is to combine the loads from these multiple sources into a single term and then statistically estimate a single model coefficient. We adopted this approach and combined the three sources into a single source term that we refer to as the “agricultural” source. This source is the sum of the annual fertilizer, manure, and atmospheric application/deposition rates at each sub-watershed. The coefficient on the “agriculture” term is dimensionless and can be interpreted as the mean annual fraction of TN delivered to the Neuse reaches from these sources.

2.2.1.4 Attenuation Data

2.2.1.4.1 Landscape Attenuation

Average depth-weighted soil permeability data were estimated for each sub-catchment based on the values derived from the State Soil Geographic (STATSGO) database (Schwarz and Alexander 1995, Natural Resources Conservation Service 2011). In general, soils with higher permeability tend to have less runoff and favor denitrification by increasing the residence time, which allows more contact with denitrifiers (Boyer et al. 2006). Moreover, higher soil porosity enhances the wetness of the soils, which in turn supports the formation of anoxic conditions that are favorable for denitrification.

The use of the natural logarithm of permeability in SPARROW has been shown to improve model fit as compared to the use of untransformed values (Schwarz et al. 2006). Under logarithmic transformation, the model coefficient on soil permeability
signifies the percent change in the land-to-water delivery associated with a 1 % change in soil permeability. While Schwarz et al. (2006) and Hoos and McMahon (2009) recommend centering soil permeability to improve model fit, we opted not to center permeability for the Neuse SPARROW model. Centering would have resulted in negative log permeability values for the catchments that had their permeabilities (in log units) lower than the median for the entire Neuse basin. Negative values would have complicated the interpretation of the land-to-water coefficient. Note that we did not incorporate other commonly used physical landscape variables such as precipitation and bedrock depth, as these variables did not show significant spatial variability at the basin scale. Annual variations in precipitation were accounted for through changes in river flows.

2.2.1.4.2 Aquatic Attenuation

Areal hydraulic loading for the four reservoirs in the Neuse basin were derived from the data compiled for the 2.8 version of the national SPARROW model (Schwarz et al. 2006). Lakes and reservoirs act as sinks for the incoming nitrogen loads, mainly through settling, increased residence time, and biotic assimilation.

Another source of aquatic attenuation is in-stream nitrogen removal through the process of denitrification. The rates of denitrification have been shown to be affected by stream size and flow levels (Alexander et al. 2000, Schwarz et al. 2006, Hoos and McMahon 2009). As such, we allowed the in-stream denitrification rates to vary by flow.
Three flow intervals were defined (mean \( Q < 2.8 \text{ m}^3/\text{day} \); \( 2.8 \text{ m}^3/\text{day} \) \( \leq Q \leq 28 \text{ m}^3/\text{day} \); \( Q > 28 \text{ m}^3/\text{day} \)) and a separate attenuation coefficient was assigned for each interval.

Accounting for different flow-dependant in-stream decay terms has become a common feature in most nitrogen SPARROW applications following the findings of Alexander et al. (2000), who showed that nitrogen delivery was a function of the stream channel size.

### 2.2.2 The SPARROW Model

The SPARROW model is a non-linear regression model with a spatially explicit component that accounts for the hierarchical and nested structures associated with the hydrological stream network. The river reach is the basic spatial unit in SPARROW and it is defined as the stream channel that connects two tributary junctions (Schwarz et al. 2006). The model predicts water quality loads at given monitoring locations from a set of watershed and river attributes. The SPARROW model also accounts for load attenuation from landscape and aquatic transport processes (Smith et al. 1997, Qian et al. 2005).

The model is capable of handling a variety of sources including point (primarily wastewater and industrial discharges) and non-point pollution sources (e.g. agricultural runoff, atmospheric deposition, animal wastes, urban runoff). The SPARROW model formulation is presented in Equation 4.
\[ \log(L_i) = \log \left( \sum_{j \in J(i)} \left( (\beta_1 S_{1,j} + \beta_2 S_{2,j}) e^{-\alpha Z_j} + \beta_{\text{point}} S_{3,j} + L_{\text{upstream of } i} \right) H_{i,j}^S H_{i,j}^R \right) + \epsilon_i \]  

Equation 4

\( L_i \) is the nitrogen load at reach \( i \) in Kg/year; \( J(i) \) is the set of upstream reaches from reach \( i \), exclusive of the reaches at or above all upstream monitoring stations. \( S_{1,j} \) is the agricultural load (sum of the atmospheric, fertilizer and manure loads) at sub-catchment \( j \); \( \beta_1 \) is the agricultural source coefficient. \( S_{2,j} \) are the non-point TN loads generated from the urban areas in sub-catchemnt \( j \); \( \beta_2 \) is the urban model coefficient, which is interpreted as a land-use export coefficient. \( S_{3,j} \) is the sum of the TN point sources discharging in stream reach \( j \); the source coefficient on the point sources is \( \beta_{\text{point}} \cdot L_{\text{upstream of } i} \) are the observed loads from the monitoring stations upstream of station \( i \).

Note that similar to the approach adopted by Qian et al. (2005), we separated the upstream loads from the within reach catchment point sources. Therefore, \( \beta_{\text{point}} \) applies exclusively to the point source loadings within the reach sub-catchments. This separation has been shown to remove systematic spatial bias that often accrues as we move from the head waters down to the river receiving water body (Qian et al. 2005).

The land-to-water delivery coefficient, \( \alpha \), takes into account the over-land attenuation associated with the landscape characteristic \( Z_j \) for stream segment \( j \). The land-delivery attenuation is expressed in the form of a first order decay process \( (e^{-\alpha Z_j}) \). The model also accounts for aquatic transport through two nitrogen processing terms.
namely, $H_{i,j}^S$ and $H_{i,j}^R$. These two terms account for the nitrogen removal in streams and reservoirs, respectively. $H_{i,j}^S$ is a first order decay term as shown in Equation 5. As noted previously, the in-stream decay coefficients ($k_s$) in the Neuse model were allowed to vary by the mean flow rate at each river segment. Three $k_s$ ($k_u$, $k_m$, and $k_b$) variables were incorporated in the model, with discrete flow thresholds as defined by Hoos and McMahon (2009). $k_u$ was the assigned as the in-stream decay coefficient for average flows $> 28 \text{ m}^3/\text{day}$, $k_m$ for $2.8 \text{ m}^3/\text{day} \geq \text{flow} \leq 28 \text{ m}^3/\text{day}$, and $k_b$ for flows $< 2.8 \text{ m}^3/\text{day}$.

On the other hand, nitrogen removal in reservoirs is accounted for by $H_{i,j}^R$ as shown in Equation 6, where $q_i$ is the areal hydraulic loading rate and $k_r$ can be interpreted as the TN settling velocity coefficient in m/yr.

\[
H_{i,j}^S = \prod_m e^{-k_{sm} \times t_{i,j,m}} \quad \text{Equation 5}
\]

\[
H_{i,j}^R = \frac{1}{1+k_r(q_i)^{-1}} \quad \text{Equation 6}
\]

$t_{i,j,m}$ is the travel time between reach $j$ and reach $i$ in flow class $m$; $m$ is the number of discrete flow classes defined; and $q_i$ is the areal hydraulic loading rate for the reservoir associated with reach $i$. The error term, $\varepsilon_i$, is multiplicative in nature but additive in natural logarithm space. Under log-scale the errors are assumed to be independent and identically distributed ($\varepsilon_i \sim \text{Norm}(0, \sigma^2)$). Moreover, the residuals should not show any spatial correlations. Readers interested in an in-depth discussion of
the SPARROW model structure, assumptions, and supporting equations are encouraged to refer to Smith et al. (1997), Schwarz et al. (2006), and Alexander et al (2008a).

2.2.3 The Neuse Bayesian SPARROW

We developed a new updated version of the Bayesian SPARROW model, which was originally proposed by Qian et al. (2005). The new version takes into account the new SPARROW formulation for the reservoir loss term $H_{i,j}^R$. Additionally, the model uses travel time instead of stream distance to estimate first order in-stream decay, while allowing the in-stream aquatic decay coefficients to vary with the mean flow values recorded at each river segment. We also updated the model data files and allowed the model to move from being static in nature to becoming dynamic over time. This new feature allowed us to fit the model using annual loading data rather than the conventional long-term multi-year average loading estimates used in other SPARROW models.

We also improved the modeling interface through the use of the R2WinBUGS package (Sturtz et al. 2005) in R (Ihaka and Gentleman 1996, R Development Core Team 2010). This permits future users of the model to run it directly from R, which is connected seamlessly to the Bayesian software package WinBUGS (Lunn et al. 2000, Spiegelhalter et al. 2003), where the actual model fitting process is conducted through a
Markov Chain Monte Carlo (MCMC) procedure. Note that both R and WinBUGS are open source software.

All results in this chapter are based on three chains that were initiated at different initial values. Posterior model results are based on 13,000 MCMC samples that were retained after a burn-in period of 7,000 iterations. Convergence was assured by monitoring the potential scale reduction factor, $\hat{R}$, for each parameter and making sure it was equal to 1.0 (Gelman et al. 2004, Gelman and Hill 2007). The WinBUGS code for the proposed Bayesian SPARROW model is included in Appendix A.

### 2.2.3.1 Synthetic Data Generation and Model Fitting

In an effort to assess the accuracy of our developed modeling framework and to better capture the effects of adding extra stations on the model fit, we first generated synthetic TN loading data by fixing the model coefficients before hand to values within the ranges previously reported in other SPARROW models (Preston and Brakebill 1999, Alexander et al. 2008a, Hoos and McMahon 2009). The predefined model parameters values are summarized in Table 2. Using these coefficients along with the sub-catchment data for the Neuse River basin and the spatial information from the river network, we calculated synthetic loading data for each stream segment. Incremental loading rates were then accumulated as we moved from the upstream reaches down towards downstream segments. This was accomplished by developing a recursive algorithm that first determines the headwater reaches (i.e. the reaches with no streams
connecting to their upstream node), then moves on to identify 2\textsuperscript{nd} order stream reaches that exclusively have head water segments connecting to their upstream nodes, and then it identifies the 3\textsuperscript{rd} order stream segments that exclusively have either head water segments or 2\textsuperscript{nd} order segments connecting to their upstream nodes, and so forth until we reach the last segment in the Neuse River that connects to the estuary. The developed recursive algorithm is similar to a certain extent to the Shreve stream order assignment (Shreve 1967). We used the developed algorithm to successively add loads and to account for nitrogen attenuation. This process ensures that we appropriately accumulate TN loading from one river junction to another as we move downstream.

### Table 2: SPARROW model parameters that were adopted to generate synthetic model data

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Value</th>
<th>Model Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.42</td>
<td>$k_b$</td>
<td>0.36</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.13</td>
<td>$k_m$</td>
<td>0.10</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>5.99</td>
<td>$k_u$</td>
<td>0.05</td>
</tr>
<tr>
<td>$\beta_{point}$</td>
<td>1.00</td>
<td>$k_r$</td>
<td>13.10</td>
</tr>
<tr>
<td>Precision ($\tau = 1/\sigma^2$)</td>
<td>16.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In an effort to gauge the effects of adding new monitoring stations, we randomly sampled a set of possible monitoring stations from the set of Neuse river junctions. Once a potential station was selected, its corresponding TN load was calculated. Note that we added some random noise to the loading estimates by assuming that the
calculated loads came from a lognormal distribution with a predefined variance. We believe that adding noise to the calculated loads makes the process more realistic. The loads that were drawn from these distributions were assumed to be the “observed loads”, on which the Bayesian model was calibrated on. Note that a perfect model fit would regenerate the exact values that we used to simulate the synthetic data.

The synthetic data allowed us to assess in more detail the behavior of the SPARROW model, especially when it came to quantifying correlations between model coefficients, evaluating the equifinality of the model, and judging the effectiveness of the commonly used model assessment metrics.

We also used the synthetic data to formally assess the effect of adding new monitoring stations on model fit and parameter identifiability. We initiated the model fitting process using only the loads calculated at the 13 exiting Neuse monitoring stations. We then incrementally added new stations and refit the model with every new addition. Ten scenarios were assessed in total, ranging from 13 to 48 stations. The stations were added either on first order or higher order stream segments (Strahler 1952). Stratifying the new stations based on stream order allowed us to see if the location of the monitoring stations had any appreciable effects on the model fitting process. This is a particularly important point to assess, given that many states and countries are interested in adopting SPARROW and are looking towards augmenting their existing monitoring networks. The results from the defined scenarios should help guide future monitoring work.
Weak priors were defined on the model coefficients, which allowed the information in the data to solely guide parameter estimation (Table 3). The specification and use of weak/non-informative priors in Bayesian analysis is common and is discussed in more detail by Gelman et al. (2004) and Van Dongen (2006). Note that we constrained the coefficients on source terms and decay rates to be positive, as is typically done in SPARROW applications (Alexander et al. 2002a, Schwarz et al. 2006)

Additionally, we also assumed that the coefficients did not vary over time; as such the 12 years of data were treated exchangeably and fit in the same model.

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Semi-informative prior distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$\text{Norm}(\mu = 0.5, \sigma^2 = 4) \mid [0, +\infty)$</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$\text{Norm}(\mu = 1, \sigma^2 = 25) \mid [0, +\infty)$</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>$\text{Norm}(\mu = 3, \sigma^2 = 25) \mid [0, +\infty)$</td>
</tr>
<tr>
<td>$\beta_{\text{point}}$</td>
<td>$\text{Norm}(\mu = 3, \sigma^2 = 25) \mid [0, +\infty)$</td>
</tr>
<tr>
<td>$k_b$</td>
<td>$\text{Norm}(\mu = 1, \sigma^2 = 25) \mid [0, +\infty)$</td>
</tr>
<tr>
<td>$k_m$</td>
<td>$\text{Norm}(\mu = 1, \sigma^2 = 25) \mid [0, +\infty)$</td>
</tr>
<tr>
<td>$k_u$</td>
<td>$\text{Norm}(\mu = 1, \sigma^2 = 25) \mid [0, +\infty)$</td>
</tr>
<tr>
<td>$k_r$</td>
<td>$\text{Norm}(\mu = 1, \sigma^2 = 100) \mid [0, +\infty)$</td>
</tr>
<tr>
<td>Precision ($\tau = 1/\sigma^2$)</td>
<td>$\text{Gamma}(\text{shape} = 2, \text{scale} = 4)$</td>
</tr>
</tbody>
</table>
2.2.3.2 Model Fitting with Annual Data from the Neuse River Basin

After we made sure that the developed model structure was accurate, we attempted to calibrate the Neuse Bayesian SPARROW using the annual TN loads at the 13 existing monitoring stations within the Neuse River basin. Since the informational content of the data from the 13 Neuse stations may be inadequate to fully calibrate a model as complicated as SPARROW, the use of vague priors runs the risk of poor model convergence. Nevertheless, our developed Bayesian SPARROW framework allows us to make use of the informational content from the larger scale Southeast SPARROW models through the specification of semi-informative priors on the Neuse model parameters. Defining semi-informative priors places loose constraints on model parameters. This allows the data to influence the posterior distributions, while making sure that the model parameters remain constrained in physically plausible ranges (Stow and Scavia 2009). This is particularly important in a model like the Bayesian SPARROW that shows strong correlations between model parameters. This behavior has been observed in simpler models, like the Streeter-Phelps model (Qian et al. 2003, Stow and Scavia 2009). Note that the use of semi-informative priors is a common approach in Bayesian water quality modeling (Borsuk et al. 2001a, Stow and Scavia 2009, Liu et al. 2010, Liu and Scavia 2010).

We chose to use positively constrained Gaussian distributions as semi-informative priors on the model coefficients. These distributions were centered on the model parameter values reported for the Southeast TN SPARROW and with variances...
set to be four times larger than those reported for the Southeast model (Table 4). Priors with variances 16 times larger than those reported for the Southeast model were also used in order to assess model sensitivity towards the choice of the priors.

Given that we combined the TN loads from atmospheric deposition, fertilizer, and manure application into one term that we are calling the “agriculture” source, the prior distribution on that combined source had to take into account the Southeast reported values on each of the three terms. This was achieved by adding the three prior distributions together to generate a Gaussian prior distribution that appropriately reflects the sum of the three sources. Equation 7 presents the Gaussian distribution on the sum of the three independent variables. It should be stated that we are assuming that the three source coefficients are independent in order to be consistent with the Southeast model. This assumption may not be entirely true given that the source coefficients may show correlations. Figure 6 shows the prior distribution on each of the three source coefficients in relation to the prior distribution for the combined “agriculture” term.

\[ \beta_i \sim \text{Norm}(\mu_{\beta_i}, \sigma_{\beta_i}^2) \]

\[ \sum_{i=1}^{3} a_i \beta_i \sim \text{Norm}\left(\sum_{i=1}^{3} a_i \mu_{\beta_i}, \sqrt{\sum_{i=1}^{3} (a_i \sigma_{\beta_i})^2}\right) \]

Equation 7

\[ a_i = \frac{\text{TN from Source}_i}{\sum_{i=1}^{3} \text{TN from Source}_i} \]
Where $a_i$ is the relative weight for each of the three source coefficients; it was assigned based on the relative TN load contribution from each of the three “agricultural” sources. $\mu_{\beta_i}$ is the Southeast reported coefficient for source $i$ and $\sigma_{\beta_i}^2$ is the assigned variance for that source.

Figure 6: The prior on the “agricultural” source term, which is the weighted sum of the prior distributions on $\beta_1$ (Fertilizer source coefficient), $\beta_2$ (Manure source coefficient), and $\beta_3$ (Atmospheric deposition source coefficient). Also shown are the prior distributions on each of the three terms.
The prior distribution on the model variance $\sigma^2$ was chosen to be an $Inv - Gamma(\text{shape} = 2, \text{scale} = 2)$. Table 4 summarizes the semi-informative priors that were specified. We believe that the use of the Southeast TN SPARROW model to generate semi-informative priors for our Neuse Bayesian model is consistent with the recommendations of Schwarz et al. (2006), who recommend nesting regional SPARROW models within larger scale models.

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Semi-informative prior distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$Norm(\mu= 0.12, \sigma^2= 0.02) \ I(0,+\infty)$</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$Norm(\mu= 0.11, \sigma^2= 0.02) \ I(0,+\infty)$</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>$Norm(\mu= 1.99, \sigma^2= 1.78) \ I(0,+\infty)$</td>
</tr>
<tr>
<td>$\beta_{point}$</td>
<td>$Norm(\mu= 0.8, \sigma^2= 0.04) \ I(0,+\infty)$</td>
</tr>
<tr>
<td>$k_b$</td>
<td>$Norm(\mu= 0.23, \sigma^2= 0.05) \ I(0,+\infty)$</td>
</tr>
<tr>
<td>$k_m$</td>
<td>$Norm(\mu= 0.13, \sigma^2= 0.01) \ I(0,+\infty)$</td>
</tr>
<tr>
<td>$k_u$</td>
<td>$Norm(\mu= 0, \sigma^2= 0.002) \ I(0,+\infty)$</td>
</tr>
<tr>
<td>$k_r$</td>
<td>$Norm(\mu= 13, \sigma^2= 31) \ I(0,+\infty)$</td>
</tr>
<tr>
<td>Precision ($\tau = 1/\sigma^2$)</td>
<td>$Gamma(\text{shape} = 2, \text{scale} = 2)$</td>
</tr>
</tbody>
</table>

Table 4: Semi-informative priors defined for the Neuse Bayesian SPARROW model
Another feature of the Bayesian approach that we capitalized on was the ability to transparently and seamlessly update the model distributions over time by combining the informational content in the priors with “new” information collected from the monitoring data. This process generated updated posterior distributions using Bayes Theorem (Equation 8). We used this Bayesian feature to update the model over time. Six time periods were defined between 1990 and 2001; each period comprised two years worth of data. The model was successively run for each time period starting with the 1990-1991 time period. The posterior distributions from the previous time step acted as the priors for the next time step. This guaranteed that as we marched forward in time, the effects of the semi-informative priors got overwhelmed by the local data collected from the Neuse River basin. The sequential updating procedure allowed us to substitute space (which is often a function of the number of stations) with time (years of monitoring) and to fit the model using annual data.

\[
p(\theta_t|y_t) = \frac{p(\theta_t|\theta_{t-1}) \times p(y_t|\theta_{t-1})}{\int p(\theta_{t-1}) \times p(y_t|\theta_{t-1}) d\theta}
\]

Equation 8

Where \(p(\theta_t|y_t)\) is the posterior probability on \(\theta\) (the set of model parameter) conditional on the observed data, \(y\); \(p(\theta_{t-1})\) is the prior probability distribution on \(\theta\) (the probability of \(\theta\) before observing the current data), and \(p(y_t|\theta_{t-1})\) is the likelihood function, which represents the likelihood of the data given the model parameters.
Due to the nonlinear structure of the SPARROW model, numerical methods were needed to explore the full multidimensional parameter space formed by the posterior distributions. We used Gibbs sampling in WinBUGS (Lunn et al. 2000, Spiegelhalter et al. 2003) to sample from the joint probability distribution of the random variables (Gelman et al. 2004, Gill 2008). The posterior distributions on the Neuse Bayesian SPARROW model parameters were defined by calculating their respective sufficient statistics using the posterior samples (Raiffa and Schlaifer 2000, Casella and Berger 2002). This was accomplished through the process of central moment matching. For the Gaussian distribution the sufficient statistics are the mean and variance, while those for the Inverse-Gamma are the shape and the scale parameters (Casella and Berger 2002).

2.3 Results

2.3.1 Synthetic Data: the Value of Monitoring and the Curse of Dimensionality

The model results in Table 5 as well as the plots in Figure 7 and Figure 8 indicated that the model performed relatively well in terms of the commonly adopted goodness of fit measures, even with the 13 existing stations. These measures included the coefficient of determination ($R^2$), the Root Mean Square Error (RMSE), and the model deviance. The RMSE and model deviance were calculated based on Equation 9 and Equation 10, respectively.

The adopted goodness of fit measures improved as we increased the number of stations; but the gains were modest. This improvement is to be expected, given the
increased information associated with the addition of new stations. The results also showed that the overall improvements in model fit were more evident for stations placed on higher order stream segments; yet these improvements did not appear to be dramatic. We hypothesize that this improvement by stream order is probably associated with the fact that stations placed on higher order stream segments cover a wider range of sources and transport pathways, which improves model convergence.

\[
RMSE = \sqrt{\frac{\sum (f(x_i) - y_i)^2}{n}} \quad \text{Equation 9}
\]

\[
Deviance = -2\log \left( p(y|\hat{\theta}) \right) \quad \text{Equation 10}
\]

where \( n \) is the number of observations \( f(x_i) \) correspond to the logarithm of the modeled SPARROW loads, \( y_i \) are the logarithms of the observed SPARROW loads; \( p(y|\hat{\theta}) \) is the likelihood of the model given the model parameters. Clearly, a lower RMSE and deviance values indicate better model fit.
Table 5: Model fit as a function of the number of monitoring stations and their locations

<table>
<thead>
<tr>
<th># of stations</th>
<th>Stream order</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Existing station (mixed)</td>
<td>0.92</td>
<td>0.51</td>
<td>12.2</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0.94</td>
<td>0.39</td>
<td>20.9</td>
</tr>
<tr>
<td>20</td>
<td>&gt;1</td>
<td>0.94</td>
<td>0.40</td>
<td>4.6</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>0.94</td>
<td>0.37</td>
<td>-3.6</td>
</tr>
<tr>
<td>27</td>
<td>&gt;1</td>
<td>0.95</td>
<td>0.36</td>
<td>31.6</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>0.94</td>
<td>0.37</td>
<td>53.6</td>
</tr>
<tr>
<td>34</td>
<td>&gt;1</td>
<td>0.95</td>
<td>0.36</td>
<td>64.1</td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td>0.94</td>
<td>0.37</td>
<td>40.6</td>
</tr>
<tr>
<td>41</td>
<td>&gt;1</td>
<td>0.96</td>
<td>0.35</td>
<td>24.0</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
<td>0.94</td>
<td>0.37</td>
<td>117.7</td>
</tr>
<tr>
<td>48</td>
<td>&gt;1</td>
<td>0.96</td>
<td>0.35</td>
<td>48.8</td>
</tr>
</tbody>
</table>
Figure 7: Model fit for the SPARROW model using synthetic data as a function of the number of monitoring stations placed on first order streams within the Neuse River basin. Reported in each of the panels is the corresponding $R^2$. 
Figure 8: Model fit for the SPARROW model using synthetic data as a function of the number of monitoring stations placed on >1 order streams within the Neuse River basin. Reported in each of the panels is the corresponding $R^2$.

Perhaps more importantly than assessing model fit, was the ability to monitor the capacity to recover the “true” set of model parameters that were used to generate
the synthetic data, as a function of the number of monitoring stations. Upon closer inspection of the posterior distributions of the model coefficients, we found that the data from the 13 existing monitoring stations were insufficient to accurately recapture the initial set of model parameters. Moreover, the posterior distributions were diffuse and showed large variability (wide credible intervals), which in turn makes coefficient interpretation hard. Yet, even under these diffuse posteriors the model converged and had decent goodness of fit metrics (high $R^2$ and reasonable RMSE) (Figure 9 and Table 5). This indicates that SPARROW is prone to equifinality, as multiple model formulations are equally supported by the data. The “curse of dimensionality” and equifinality are common limitations associated with most mechanistic and semi-empirical models (Franks et al. 1997, Beven and Freer 2001, Beven 2006, Stow et al. 2007). Under such cases, it is key to provide a distribution for the model parameters rather than focusing on the “best” point estimates for each of the model parameters (Stow et al. 2007). Acknowledging and accommodating equifinality allows us to transparently convey our uncertainties. Fortunately, the results from our developed Bayesian model were in the form of posterior distribution.

As we increased the number of monitoring stations, the posterior distributions on the model parameters got tighter and converged to the predefined set of model parameters (Figure 9). Yet, for some model parameters, the model only managed to converge to values close to the predefined parameter set even with 48 stations (13 exiting + 35 new). This was due to interactions and correlations between the model
coefficients as we discuss below. Note that there does not seem to be an appreciable difference between having stations on 1st order stations versus on higher order streams.

Figure 9: Variations of the model coefficients with respect to the number and order of the monitoring stations used to fit the model. The blue lines are the originally set values that were used to generate the synthetic data. The unfilled circles represent the mean posterior values of the parameters. The thick horizontal segments correspond to the intervals associated with ±1 standard deviation. The thin segments show the intervals associated with ±2 standard deviation away from the mean.
Another key goal of this work was to assess the posterior distributions of the model coefficients in order to properly characterize and quantify the correlations between the SPARROW model parameters. Of particular interest was the relationship between $\alpha$, the land-to-water delivery coefficient, and the coefficients on the non-point TN sources ($\beta_1$ and $\beta_2$). Based on the SPARROW model formulation (Equation 4), the relationship between $\alpha$ and the source coefficients ($\theta_1$, $\theta_2$) is intrinsically prone to show strong correlations.

The Bayesian implementation of the SPARROW model allowed us to account for these complex and often strong correlations between the variables by sampling from the joint posterior parameter space. As such, we drew samples from the joint posterior parameter space and calculated the non-parametric Spearman's rank correlation coefficient (Hollander and Wolfe 1999). The results indicated high correlations between some of the model coefficients, with the correlations between $\alpha$, $\theta_1$, and $\theta_2$ being especially strong (Figure 10 through Figure 13). It was also evident that the ability to appropriately capture the parameter correlations improved as we increased the number of stations. With the exiting 13 stations, the joint distribution space was diffuse and over-dispersed (Figure 10). Moreover, the “occupied” joint parameter space contracts significantly with the addition of new monitoring stations as illustrated in Figure 13. Differences between 1st order and higher-order stations were minor when it came to model parameter correlations.
Not surprisingly the correlation between $\alpha$ and the source coefficients ($\beta_1$ and $\beta_2$) was found to be positive. These strong positive correlations indicate that multiple model parameterizations can generate similar TN loadings. Increasing the values of the source input coefficients ($\beta_1$ and $\beta_2$), which leads to an increase in the availability of TN on the landscape, can get attenuated by increasing the value of the land-to-water attenuation coefficient, $\alpha$. It is worth noting that the correlation between the coefficients is purely an artifact of the SPARROW model formulation, as we did not a priori specify any correlation structure between the model parameters.

The observed high correlations between $\alpha$, $\beta_1$, and $\beta_2$ emphasizes the need to use the joint distribution on the model variables for model predictions and parameter interpretation. The valid parameter space that the three variables occupy is tightly constrained (Figure 13); therefore treating these variables independently can result in erroneous load prediction estimates as well as high uncertainties. Gaining marginal inference on one variable should be done conditional on the others. These findings are consistent with Schwartz et al. (2006) recommendation that for interpretation purposes $\alpha$, the land-to-water delivery factor, should be combined with the source coefficients ($\beta$) to form a single aggregate delivery factor. Nevertheless, most SPARROW models often interpret the model parameters independently and associate them with a mechanistic interpretation.
Figure 10: Bivariate plots showing the correlations between $\alpha$, $\beta_1$, $\beta_2$, and $\beta_{\text{point}}$ from the SPARROW model. Plotted values are based on the posterior distributions associated with the 13 existing monitoring stations. Spearman's rank correlation coefficients are reported in each of the panels. The grey shading in each the panel is proportional to the bivariate density. The red dot shows the location in 2-dimensional space of the “true” parameters in each panel.
Figure 11: Bivariate plots showing the correlations between $\alpha$, $\beta_1$, $\beta_2$, and $\beta_{\text{point}}$ from the SPARROW model. Plotted values are based on the posterior distributions associated with 48 monitoring stations (13 existing +35 1st order new stations). Spearman’s rank correlation coefficients are reported in each of the panels. The grey shading in each the panel is proportional to the bivariate density. The red dot shows the location in 2-dimensional space of the “true” parameters in each panel.
Figure 12: Bivariate plots showing the correlations between $\alpha$, $\beta_1$, $\beta_2$, and $\beta_{\text{point}}$ from the SPARROW model. Plotted values are based on the posterior distributions associated with 48 monitoring stations (13 existing +35 higher order new stations). Spearman’s rank correlation coefficients are reported in each of the panels. The grey shading in each the panel is proportional to the bivariate density. The red dot shows the location in 2-dimentional space of the “true” parameters in each panel.
Figure 13: Three-dimensional scatterplot showing the parameter space shared by $\alpha$, $\beta_1$, and $\beta_2$. In the right panel the values are based on the posterior distributions associated with data from the 13 existing monitoring stations. The panel on the right shows the results based on 48 stations (13 existing + 35 order 1 stations). The color gradient is based on the values of $\alpha$. Note that the scales are set to be the same in both plots. The blue asterisk shows the location in 3-dimntional space of the “true” parameters.

2.3.2 Neuse Data: Dynamic Updating over Time

The developed dynamic Bayesian SPARROW model for the Neuse had three source variables (point sources, “agricultural” sources, and urban non-point sources), a land-to-water delivery term based on the natural logarithm of soil permeability, three in-stream loss-rate coefficients that vary as a function of mean annual river flow, and one reservoir decay-rate coefficient. All of the model coefficients were statistically different from zero, except for the in-stream TN attenuation coefficient that was
assigned for high flow rate stream segments \( (k_{su}) \). This is to be expected as very little nitrogen removal occurs within the main stems of rivers (Alexander et al. 2000).

A comparison between the model coefficients for the Neuse SPARROW model and those of the Southeast nitrogen regional model indicated that they were different. This is apparent in Figure 14 and Figure 15, where the mean parameter values for the Southeast model are shown by the dotted horizontal lines and the estimates on the Neuse model coefficients are summarized by their posterior means and standard deviations.

The median values for the Neuse land-to-water coefficient, \( \alpha \), was around 0.2 over the 12 year time period; the values appear to increase slightly over time, although the changes were not significant. The land-to-water delivery coefficient was larger than that reported by Hoos and McMahon (2009) for the Southeast SPARROW model. Note that \( \alpha \) can be interpreted as the percent drop in TN loading to the streams due to a 1 percent increase in soil permeability.

The model results indicated that catchments with higher soil permeability have lower effective TN loading rates. This is to be expected as higher permeability allows for more contact time between the TN and the microbial community, which in turn increases the denitrification rate as well as nitrogen assimilation by both heterotrophes and autotrophies. Moreover, soils with low permeability are associated with tile drainage particularly in the coastal plain area of the Neuse basin. Tile drains are very effective in transporting any accumulated onsite water along with available nitrogen to
nearby stream segments. This curtails the effectiveness of TN removal. The changes in $\alpha$ across time were minor, indicating that the denitification rates have been constant over time.

The source coefficient on the “agricultural” term (which is a combination of TN load from fertilizer, manure, and atmospheric deposition) was around 0.1 across the 12 years of data. This value is close to the value derived from the Southeast SPARROW model (0.12). Meanwhile, the coefficient on the non-point source TN loading from urban areas in the Neuse was between 2.5 and 4 Kg/1000 m$^2$ over the 12 year study period. This value was larger than the 1.99 Kg/1000 m$^2$ value reported for the Southeast model. Additionally, the Neuse values were larger than those reported for the New England (0.8 Kg/1000 m$^2$) and Chesapeake Bay (0.9 Kg/1000 m$^3$) SPARROW models. Nevertheless, field studies have reported TN export rates from urban areas ranging between 0.16 and 38.5 Kg/1000 m$^2$ (Beaulac and Reckhow 1982, Dodd et al. 1992, Bales et al. 1999). A recent study by Line et al. (2002) reported that the average nitrogen export coefficient from residential areas in the Upper Neuse basin was 2.39 Kg/1000 m$^2$.

Since both the source coefficients as well as the land-to-water attenuation coefficient for the Neuse SPARROW were larger than those for the Southeast, we decided to compare the $\alpha \times \beta_1$ and $\alpha \times \beta_2$ from the Neuse dynamic Bayesian SPARROW to their Southeast model counterparts. Looking at the products of the variables aims to account for potentially strong collinearities between these coefficients. The results indicated that the Neuse estimates for effective land-based loading were indeed larger.
than those for the Southeast model. By comparison, the median $\alpha \times \beta_1$ for the Southeast model was around 0.014, while the median values for the Neuse model ranged from 0.024 for the 1994-1995 period to 0.028 for the 1992-1993 period. Similarly, the median value for $\alpha \times \beta_2$ was around 0.24 for the Southeast model, while that for the Neuse ranged between 0.49 in 1990-1991 period to 0.76 in 1998-1999.

In SPARROW models, the coefficient on point sources is often expected to be close to unity, as point sources are assumed to discharge directly to the stream segments without any land-delivery attenuation. The median value for the Neuse point-source TN coefficient ($\beta_{point}$) ranged between 0.96 and 1.18 Kg/Kg. The value of the coefficient tended to increase slightly over time before stabilizing. The coefficient is around 1.5 times larger than that reported by Hoos et al. (2009) in their Southeast model (0.8). The fact that $\beta_{point}$ was slightly larger than one may suggest that the current TN point source loading estimates for the Neuse (McMahon et al. 2007) were underestimated.
Figure 14: The development of the posterior distributions of the source coefficients ($\beta_1$, $\beta_2$, and $\beta_{\text{point}}$) and the land-to water delivery factor ($\alpha$) over time. The black dots represent the posterior median values; the gray vertical segments are ± 1 standard deviations from the median; the black dashed horizontal line represents the mean value reported for the Southeast SPARROW model.
The Neuse in-stream aquatic attenuation rates (Figure 15) were comparable to the ones reported in the Southeast model and in other SPARROW applications including the Chesapeake Bay (Preston et al. 1998), the National SPARROW model (Smith et al. 1997, Alexander et al. 2008a), and the Waikato River basin (Alexander et al. 2002a). The variations of the in-stream decay variables were largely temporally invariant, indicating that the factors affecting the denitrification mechanism did not change appreciably across the study period. As expected the attenuation rate decreases with increasing mean annual river flow; a finding that has been corroborated in several other studies (Smith et al. 1997, Alexander et al. 2000, Schwarz et al. 2006, Alexander et al. 2008a, Alexander et al. 2009).

The variation in the lake attenuation coefficient for the Neuse across the 12 years was very close to those reported for the Waikato River basin (Alexander et al. 2002a) and to field measurements reported in other lakes (Schwarz et al. 2006); however, the values were significantly lower than those reported for both the Southeast (Hoos and McMahon 2009) and the North Carolina models (McMahon and Roessler 2002, Qian et al. 2005). This may indicate that the role that the Neuse reservoirs’ play in nitrogen attenuation is less important than what is expected from the Southeast model results. The lake attenuation rate was largely constant across time (Figure 15). Note that the uncertainty around the stream and lake attenuation coefficients decreased consistently with every Bayesian updating.
Figure 15: The development of the aquatic attenuation coefficients (Ks.u, Ks.b, Ks.m, and Kr) over time. The black dots represent the median values; the gray vertical segments are ± 1 standard deviations from the median; the black dashed horizontal line represents the mean value reported for the Southeast model.

The development over time of the model posterior distributions, as seen in Figure 14 and Figure 15, was largely due to Bayesian model updating. This process
involved combining the new information (observations) with the prior information from the previous step to generate new posterior distributions. As we march forward in time more data becomes available, which in turn decreases the standard errors on the posterior distributions. This framework also accounted for changes in the mechanisms governing landscape loading and/or delivery. Changes in delivery and attenuation rates over time are accounted for through the Bayesian updating process. It is hard to separate the contribution of each of these two processes given the limited data we had.

Overall, the model’s RMSE decreased while the $R^2$ increased with sequential updating (Figure 16). This is to be expected as new information became available over time. The plots of observed versus predicted TN loads indicate relatively good agreement between observed and predicted values. Figure 17 shows the temporal development of the posterior distribution on the precision. Given that we started with a semi-informative prior that was almost flat, $Inv - Gamma(shape = 2, scale = 2)$, the posterior distribution on precision consistently got more concentrated and shifted towards larger values. This corresponds to a better fit and higher precision. The model’s precision following the last update was around 3 (i.e. $\sigma^2 \approx 0.33$). This value was comparable to the model errors reported for the North Carolina model ($\sigma^2 \approx 0.19$) and slightly larger than those for the Southeast model ($\sigma^2 \approx 0.12$).
Figure 16: Predicted versus observed log TN loading in the Neuse as a function of time. The red solid line is the 1:1 line, which indicates a perfect fit. The solid points are the mean log TN loads; the dark grey vertical segments indicate the log load $\pm$ 1 standard deviation (68 % credible interval); the lighter grey vertical segments represent the log load $\pm$ 2 standard deviation (95 % credible interval).
Figure 17: The development of the posterior Gamma distribution on the model precision ($\tau = 1/\sigma^2$) over time. The dashed line represents the density of the semi-informative prior we initially set on the precision: $\text{Gamma}(\text{shape}=2, \text{scale}=2)$

The $R^2$ for the Neuse Bayesian SPARROW was slightly lower than those reported in other SPARROW applications. Similarly, the RMSE was marginally larger. Yet, this is to be expected given that the SE and the North Carolina models used a larger number of stations to calibrate the model against. Additionally, those models used long-term detrended average annual loading data, which are less variant than the annual loads we used in our dynamic Neuse Bayesian SPARROW model. We also assessed the model fit of the Southeast SPARROW parameterization with respect to the Neuse data. The results clearly showed that our regionalized approach provided a significantly better fit
both in terms of the $R^2$ and RMSE (Figure 18). This highlights the importance of regionalization SPARROW to the basin scale.

![Figure 18: Predicted loads versus observed TN loads in the Neuse based on the Southeast TN SPARROW model (Hoos and McMahon 2009). The red solid line is the 1:1 line, which indicates a perfect fit. The solid points are the mean log TN loads.]

Non-Bayesian SPARROW models have often suffered from scale invariant model residuals (Smith et al. 1997, McMahon et al. 2003, Schwarz et al. 2006). The Neuse Bayesian SPARROW model residuals did not show a clear spatial pattern as seen in Figure 19, which plots the residuals as a function of river distances upstream from the Neuse Estuary. Yet, as is expected the magnitude of the residuals decreased over time as more data became available, which in turn improved model fit. These findings
indicate that the Neuse Bayesian SPARROW model formulation did not violate the assumption of having independent and identically distributed errors with no spatial patterns.

![Graph showing spatio-temporal variability of residuals from the Neuse SAPRROW model.](image)

**Figure 19:** Spatio-temporal variability of the residuals from the Neuse SAPRROW model

### 2.4 Discussion

#### 2.4.1 Bayesian Framework: Model Regionalizing

We developed a Bayesian SPARROW model for the Neuse River Basin in North Carolina that adopts a novel regionalization approach, which nests the Neuse model
within the larger-scale Southeast SPARROW model through the use of semi-informative priors. These priors get updated and the model structure gets regionalized as more data from the Neuse become available over time. This dynamic updating approach culminates with regionalized basin-specific posterior distributions that can be effectively used to assess nitrogen loading to the Neuse River and ultimately to the estuary. As seen earlier, the optimized SPARROW parameterization for the Neuse river basin showed marked differences from the larger scale Southeast SPARROW model. These differences can lead to different management decisions.

Our analysis indicated that the developed basin-scale model provides far more accurate load predictions when compared to those based on the regional-scale model. These differences are primarily due to the fact that the regional/national SPARROW model coefficients have been optimized based on data that cover a wide range of river basins that have different intrinsic and extrinsic properties. Therefore any attempt to downscale the model results and/or to make basin-specific predictions become particularly challenging. This limitation has hampered the effective use of SPARROW in basin-scale management. As discussed earlier, there has been a push by USGS towards more regionalized SPARROW models. Recent models, like the Southeast SPARROW model, have incorporated dummy variables within the model structure in order to account for spatial differences between sub-regions; yet such an approach entails two main assumptions. The first being that the differences within a sub-region (between river basins in a given sub-region) are insignificant and can be ignored, while the second
considers that the differences between the sub-regions can be adequately accounted for by allowing the land-to-water delivery coefficients to vary by sub-regions. These two assumptions are hard to justify.

A more appropriate approach to adopt for developing regional SPARROW models will incorporate our developed approach within a hierarchical Bayesian framework. This will provide a hierarchical structure for the explanatory variables, and allow for the partial pooling of data from different regions/river basins. Partial pooling allows for the shrinkage of the predicted model parameters toward a common mean. This feature also allows for making load predictions in basins/regions with few monitoring stations by borrowing strength from the rest of the regions. The level of shrinkage towards the overall mean is dependent on the available basin data as well as the within and between basin/region variances (Gelman et al. 2004, Gelman and Hill 2007, Kashuba et al. 2009, Qian et al. 2010).

2.4.2 Bayesian Framework: Temporal Updating

Our developed Neuse Bayesian SPARROW model was temporally dynamic, which allowed the model parameters to be updated with time. As such, the model structure and the load estimates were updated incrementally between 1990 and 2001. To our knowledge this is the first application of a temporally dynamic SPARROW model. Moreover, this approach moved SPARROW past its temporally static framework towards
a dynamic platform, which increases the availability of calibration data. This is particularly important for regions with few stations that have long monitoring records.

Temporal updating also permitted the identification of any shifts in pollutant loading/delivery processes over time. The results from the Neuse application have indicated that the model coefficients are not constant over time. While in this application the temporal changes in model parameters were largely a function of the Bayesian updating process rather than basin-level changes in TN delivery/attenuation, our developed framework is capable of identifying and accounting for system changes especially with the availability of long-term monitoring records and the implementation of large scale management decisions.

2.4.3 Parameter Correlations and Model Inference

Smith et al. (1997) have stated that the mass balance constraints on SPARROW improve the interpretability of the model coefficients and allow for their comparison with literature values. Yet, our adopted simulation work made it possible to fully characterize and quantify the correlations between the model coefficients. While strong parameter correlations have been previously recognized as a potential shortcoming of SPARROW (Qian et al. 2005, Schwarz et al. 2006), this is the first time that it has been properly characterized and quantified. As seen from our results, the joint posterior parameter space of the model coefficients is asymmetrical and highly constrained. This makes conventional numerical optimization algorithms, such as minimizing the residual
sum of squares, inefficient. Moreover, these algorithms can get trapped in local minima especially when parameter correlations are strong (Qian et al. 2005). In this study, we used MCMC, which allows for sampling from the entire joint posterior distribution. This approach adequately captures the joint posterior parameter space and accounts for parameter correlations. Another approach that has proven to be effective under such cases is the Generalized Likelihood Uncertainty Estimation (GLUE) process proposed by Beven and Binley (1992).

Strong correlations between the SPARROW model parameters can diminish the mechanistic/physical interpretability of the model coefficients. Yet, these correlations are not expected to diminish spatial accountability of the pollutant fluxes in a given study area. This indicates that several behavioral models can adequately reproduce the observational data. This concept is referred to as equifinality (Beven and Freer 2001, Beven 2006).

2.4.4 Increased Monitoring: What Do We Gain?

In our simulation analysis, we were able to evaluate the value of adding monitoring stations both in terms of model fit and parameter identifiability; the effects on the latter were found to be more pronounced. In terms of model fit, we observed that even with few monitoring stations (13 stations) the generated SPARROW model had respectable goodness of fit measures. Yet, the model parameter space was poorly constrained, which limits inference and the ability to make predictions. We also saw
that placing monitoring stations on higher order streams resulted in minor improvements (in terms of model fit) as compared to placing them on 1st order streams.

Data limitation will remain the main constraint curtailing the regionalization of SPARROW in the foreseeable future. In conventional SPARROW models, the impetus has been towards increasing the number of monitoring stations by expanding the geographic boundaries of the study area. It seems to us that any gains from such an approach may get diminished by the increased spatial variations and the inability of the conventional SPARROW model structure to properly account for these variations. However, our Bayesian framework resolves many of these limitations, particularly through the adoption of a Bayesian hierarchical model structure.

Our findings have implications on current attempts aiming to use SPARROW in developing countries (e.g. China and India). Under these cases, we recommend that the SPARROW model should be initially fit with few source coefficients, while adopting weakly informative priors based on the current US applications. The model can then be updated sequentially as new data becomes available. Over time the model will get regionalized and its predictions will become more accurate. Moreover, the results generated from each updating step can help identify hotspots and allow for targeted sampling as McMahon et al. (2003) showed in their North Carolina SPARROW application.
3. Developing a Bayesian Changepoint-Threshold Flow Concentration Model

*Note: the results and text from this chapter appeared in a recently published peer-reviewed article in Water Research (Alameddine et al. 2011b). The paper was co-authored by Dr. Song S. Qian and Dr. Kenneth H. Reckhow.*

3.1 Introduction

Anthropogenic nitrogen reaching rivers, lakes, estuaries, and coastal areas has been linked to eutrophication, acidification, adverse human health effects, the disruption of ecosystem functions, as well as the lowering of biodiversity in affected water bodies (Kelly 2008 and references therein). High profile events such as the dramatic fishkills in the Neuse River and Estuary in mid 1980 and early 1990s as well as the development of extensive hypoxic zones in the Gulf of Mexico have been linked to excessive nitrogen release and delivery (Turner and Rabalais 1994, Paerl et al. 1995, Paerl 1997, Alexander et al. 2000, Scavia et al. 2003, Stow and Borsuk 2003a, Borsuk et al. 2004b, Alexander et al. 2008b). Such events have stimulated an impetus towards the implementation of aggressive management and mitigation measures to limit the amount of nitrogen reaching the aquatic environment. While successes have been made in some water bodies, water quality impairment from excessive nitrogen loading continues to be a pressing issue on the national as well as international levels. Currently nutrient impairment ranks fourth on the national impairment list with 6,826 water
bodies listed as impaired due to excessive nutrient loading (out of a total of 75,677 impairment causes). Of these around 17% have nitrogen explicitly listed as the cause of impairment (Environmental Protection Agency 2009).

The Neuse River and its associated estuary in North Carolina have experienced all the symptoms of eutrophication with extensive algal blooms, fish kills, hypoxia and anoxia that have captured the public attention in the 1980s and 1990s (Paerl et al. 1998). This led to designating the Neuse as a nutrient sensitive water and prompted its listing on the 303(d) list with nitrogen identified as the main culprit behind eutrophication (Paerl et al. 1995, Stow et al. 2003). In 1998, the USEPA settled a lawsuit brought by the Neuse River Foundation which required North Carolina to establish a TMDL for nitrogen reaching the estuary. The TMDL was approved by USEPA on August 26, 1999. Meanwhile, the State of North Carolina, through the North Carolina Division of Water Quality (NCDWQ), adopted in 1997 a set of rules that aimed at reducing the amount of nitrogen delivered to the Neuse River Estuary by 30% based on mean 1991-1995 loads. Despite almost a decade of post-TMDL monitoring, there has been no consensus on whether the TMDL has achieved its stated goal and if the implemented management measures have been successful (Deamer 2009).

To better understand the nitrogen dynamics and the effectiveness of the TMDL program in the Neuse over time, we made use of daily flow measurements to estimate nitrogen concentrations and nitrogen loading rates in the Neuse. The use of flow measurements to predict nutrient concentrations (and thus load) is common in water
quality modeling given that in-stream nutrient concentrations have been observed to exhibit a relationship with river/stream flow (Johnson 1979, Reckhow and Stow 1990, Stow and Borsuk 2003a). The development of flow-concentration (as well as flow-load) models are often used to draw upon the large databases of daily flow measurements in order to augment infrequent water quality sampling measurements.

The use of regression-based empirical methods to predict daily nutrient loading through the use of daily averaged river flow measurements is one of the more commonly used approaches to determine nutrient concentrations/loads (Cohn et al. 1992, Green and Haggard 2001, Hooper et al. 2001, Haggard et al. 2003, Runkel et al. 2004, Ide et al. 2007). This approach is based primarily on the work of Cohn et al. (1992) who developed the “rating curve” method that involves a log-linear multivariate regression model linking flow to concentration and load. While refinements have been added to the original “rating curve” method, most of the adopted models assume that the relationship between flow and concentration is fixed over time as well as across the range of river flows. Yet, the implementation of environmental management measures at a river basin scale can often result in changes to the underlying relationship linking flow to concentration measurements, and ultimately affect load estimates. With the implementation of basin-scale water quality management plans in different river systems, it is becoming increasingly imperative to evaluate the effects that such basin-scale management plans (like the TMDL program) have on these systems. So far, however, there has been little chance to conduct such an assessment due to the
difficulty of finding a river system with both a long monitoring record and that has had TMDL mitigation measures put in place. The Neuse River presents a unique opportunity to study these changes and demonstrate their impacts due to the presence of a monitoring program that stretched for over 30 years, during which a TMDL program has been enforced.

The objective of this chapter is to assess the dynamics of the relationship of nitrogen concentration and flow between 1979 and 2008 in order to determine whether the relationship has been time invariant or if it has experienced a major changepoint across time. The occurrence of a system changepoint at a specific point in time often signifies an abrupt change in the way a system operates or behaves. These system changes may involve changes to the model parameters, the underlying model structure, or changes to both. We use Total Oxidized Nitrogen (TON) (which is the sum of nitrate ($\text{NO}_3^-$) and nitrite ($\text{NO}_2^-$)) concentrations in the Neuse, given its long historical data. We then check the identified temporal changepoint against the backdrop of environmental management measures undertaken in the Neuse as well as a set of natural major shocks to the system in order to gain a better understanding of nitrogen delivery to the system. Additionally, we explore whether a flow threshold is discernable, while exploring the implications of its presence/absence. Finally, we use the empirical model to evaluate whether the implemented TMDL mitigation measures have been successful in reducing nitrogen loading by 30 %, while taking into account associated model and parameter uncertainties. To our knowledge this is the first attempt to study nutrient delivery in the...
form of TON to the Neuse River system that spans three decades of data (1979-2008) during which major natural as well as watershed-scale management measures have occurred.

3.2 Materials and methods

3.2.1 Study Area and Data

The Neuse River at the Fort Barnwell station drains an area of 10,100 km$^2$ (United States Geological Survey 2010). The basin has a diverse landuse/landcover. Just east of its head waters, an urbanized area -that includes the cities of Raleigh, Durham, and Cary- dominates the basin. Intensive agricultural areas and CAFOs become more prominent towards the lower portions of the Neuse basin, where the river traverses the North Carolina coastal plain (Figure 20). The major point source nitrogen emitters in the basin are the 20 major wastewater treatment plants that service the cities and townships in the 19 counties that fall within the Neuse basin. Moreover, continued landuse changes in the basin have resulted in an increase in the relative importance of non-point nutrient sources to the overall Neuse nutrient budget.
Figure 20: Neuse River Basin showing the major urban areas, the river network, as well as the town of New Bern, where the Neuse River opens to form the Neuse Estuary before discharging in Pamlico Sound. The image in the small frame shows the Neuse River Basin within the context of the State of North Carolina.

TON concentrations and water temperature data between 1979 and 2008 were primarily collected from the USEPA’s STORET service that publishes ambient monitoring data collected by NCDWQ at Fort Barnwell (Lat = 35.3125 N; Lon = 77.3022 W). TON sampling frequency at Fort Barnwell has changed significantly over the years. Sampling was conducted on a monthly basis between 1979 and 1995, while the years 1996 through 2002 saw a major increase in the sampling frequency (>200 samples per year) before the sampling effort was reduced to weekly post 2002. Water temperature values were included in order to capture the seasonality in TON concentrations that are known
to vary by season as a result of changes in the biological and physio-chemical changes in the river and its contributing watershed (Malone et al. 1996).

Flow data for the same period were collected for the Fort Barnwell Station through the USGS’s National Water Information System (NWIS). The NWIS database also reported data on TON concentrations and ambient river water temperature. These were used to augment data from USEPA’s STORET. Missing values for flow measurements at Fort Barnwell were estimated through the regression model that was developed by Stow and Borsuk (2003a). Their model predicts flow at Fort Barnwell from two upstream USGS operated gauging stations.

In the same context, missing water temperature values were imputed using a linear regression model linking water temperature to ambient air temperature, with a constraint placed to ensure that imputed water temperatures do not drop below freezing (Equation 11). Both empirical models used to impute the data resulted in high coefficients of determination ($R^2$ was 0.97 and 0.85, respectively). As such, the uncertainties associated with both models were not included within the modeling framework.

$$\text{Water Temp } (^\circ C) = 3.26 + 0.89 \times \text{Air Temp } (^\circ C) + \varepsilon, \varepsilon \sim N(0, \sigma^2); \sigma = 3.02 ^\circ C$$

if Water Temp $\leq 0 ^\circ C$

then:

- Water Temp $= \text{Water Temp } (^\circ C)_{i-1}$, if Water Temp $> 0$;
- Water Temp $= \text{Water Temp } (^\circ C)_{i+1}$, if Water Temp $> 0$;
- else Water Temp $= \text{Min(Water Temp } (^\circ C))$

Equation 11
Fort Barnwell was selected as the point of interest for this study given that 1) it is the most downstream long-term USGS maintained gauge on the Neuse (~20 km upstream from the Neuse Estuary), 2) it drains around 85% of the Neuse watershed, and 3) it has been used in recent Neuse River water quality studies (Stow et al. 2003, Borsuk et al. 2004b).

3.2.2 Model Development

Given the changes that have occurred at the Neuse River Basin, we developed a changepoint-threshold Bayesian model that is capable of predicting TON concentrations from river flows and water temperature. To avoid confusion, we point out that we will be using the term *changepoint* to signify a change occurring over time, while the use of the term *threshold* will be reserved to indicate a change over the range of river flows. Even though the inclusion of a changepoint along with a threshold adds to the complexity of the model, it allows for a better understanding of the system by removing the constraining assumption of invariance both in time and across flows.

Accounting for a threshold in the relationship governing nutrient concentration and flow in the Neuse is not new. It was previously suggested by Borsuk et al. (2004b) in their regression-based estuarine model for the Neuse Estuary. However, their model-like most other water quality empirical models such as the LOAD ESTimator (LOADEST) (Runkel et al. 2004) and the SPARROW (Smith et al. 1997) models- assumes that the flow-concentration relationship is constant over time. In many cases, assuming that a
system is statistic over time can be a reasonable assumption; yet in many cases failing to properly account for the temporal dynamics in the system can lead to incorrect parameter estimation, inflated errors, as well as to poor predictive power. This is particularly true when the system under study experiences an external intervention such as a policy change (Congdon 2006).

The incorporation of temporal changepoints in statistical models has been well established in the social sciences (e.g. Western and Kleykamp 2004, Park 2006); nevertheless, far too little attention to our knowledge has been paid to their use in water quality models. The incorporation of a changepoint in a water quality model is relatively easy when the changepoint is known beforehand. This is usually done through the inclusion of a dummy variable or through fitting separate models for each period. Yet in most cases, we are seldom sure of the occurrence or the exact timing of changepoints. Thus, what we are often more interested in is: 1) to be able to recognize whether a changepoint actually occurs or not, 2) to have a rigorous method to estimate its timing, and 3) to have an associated probability distribution that accounts for the uncertainty in its time of occurrence.

The changepoint-threshold Bayesian model that we propose is capable of predicting TON concentrations ($TON$) from flow measurements ($Q$) and water temperatures ($Temp$) without constraining the relationship linking flow to TON concentrations to be fixed over time or over the range of observed river flows (Equation 12). We assume that the random variable $TON_{i=1,\ldots,n}$, like most water quality
concentration variables, follows a lognormal distribution (Ott 1995). As such we can describe the changepoint-threshold model as:

\[
\log(TON_i) = N\left(\alpha_{j[i]} + \beta_{j[i],k[i]} \times (\log(Q_i) - \text{threshold}) + \beta_{\text{Temp}} \times \text{Temp}, \sigma_{j[i],k[i]}^2\right)
\]

\[
j[i] = \begin{cases} 
1; & \text{if } \text{year}_i - \text{changepoint} < 0 \\
2; & \text{if } \text{year}_i - \text{changepoint} \geq 0 
\end{cases}
\]

\[
k[i] = \begin{cases} 
1; & \text{if } \log(\text{Flow}_i) - \text{threshold} < 0 \\
2; & \text{if } \log(\text{Flow}_i) - \text{threshold} \geq 0 
\end{cases}
\]

where \(\alpha_j\) is the model intercept. It corresponds to the log(TON) concentration when \(\log(Q)\) is zero in the event that no flow threshold is identified, and to the log(TON) concentration when flow is equal to the median flow threshold in the event a flow threshold is recognized. \(\beta_{j,k}\) correspond to the slope on \(\log(Q)\) with different values assigned depending on whether the system has passed the changepoint and/or threshold. Note that in order to improve parameter identifiablity when running the MCMC procedures, \(\beta_{2,2}\) was redefined as \(\beta_{2,1} + \delta\). \(\beta_{\text{Temp}}\) represents the slope on water temperature. We opted to maintain a common slope on water temperature, as there was no reason to suspect that the seasonal patterns captured by temperature values changed over time or flow. Moreover, the model allows for different normally distributed error terms (\(\sigma_{j,k}\)) for pre and post changepoint as well as for pre and post flow threshold. This allows for the possibility that the model may perform better under certain ranges of flow and/or over certain periods of time.
The decision to include a changepoint within the model framework permits the model to incorporate any change in the flow-concentration relationship over time, while also allowing for the pooling of data into two main groups, namely pre and post changepoint. This partial pooling of the data offers a clear advantage over completely combining all measurements into one system or treating each year separately (Gelman et al. 2004, Gelman and Hill 2007, Qian et al. 2010). For this study the temporal partial-pooling also helps to overcome problems associated with the inter-annual variations in the Neuse hydrograph, whereby some years are wet while others are relatively dry. This ensures that the TON sampling events covered the entire flow range in the Neuse for both pre and post changepoint groups. This reduces model bias towards low flow conditions as described by Ide et al. (2007), especially during low sampling frequency periods in the Neuse (1979-1995 and 2003-2008) (Figure 21).
Figure 21: Boxplot of TON concentrations in natural logarithm scale from 1979 till 2008 for the Fort Barnwell station. Grey line indicates a locally-weighted polynomial regression LOWESS curve that traces TON concentrations across time. Sampling frequency for each year is also illustrated in the inverted histogram.

The incorporation of a flow threshold within the model structure allows for the detection of an environmental threshold along the range of observed daily mean flows. The presence of the threshold signifies a change in the manner that TON concentrations respond to flow measurements. For this study, the inclusion of the threshold within the
model structure allows for the accommodation of a piecewise-linear relationship linking river flows to TON concentrations. The adopted piecewise-linear model in this framework constrained the two linear pieces to intersect at the identified flow threshold. For a more detailed discussion of piecewise linear regression models and their uses, refer to Qian and Richardson (1997). Notice that in the event that the data do not support the presence of a threshold and/or of a changepoint, the model will simply collapse back into a simplified flow-concentration linear regression model.

Under a Bayesian framework, prior distributions have to be specified on the model parameters \( \theta = (\alpha_j, \beta_{j,k}, \beta_{\text{Temp}}, \sigma_{i,j}, \delta, \text{Threshold}, \text{Changepoint}) \). We defined diffuse prior distributions on all model parameters. A constrained weak prior was specified on the \( \text{Changepoint} \) distribution. This was done by defining a discrete uniform distribution over the range of sampling years i.e. \( \text{Changepoint} \sim \text{Unif} (\text{Min(Year)}, \text{Max(Year)}) \). All years in that range were given equal probabilities of being identified as potential changepoints. The assignment of a weak prior on \( \text{Changepoint} \) lets “the data speak for themselves” (Gelman et al. 2004), and thus assures that any posterior inference on the timing of the changepoint is unaffected by information external to the current data. Similarly, the prior on the flow threshold was chosen to be a continuous uniform distribution that is bounded to the range of observed river flows i.e. \( \text{Threshold} \sim \text{Unif} \left(\text{Min}(\log(Q)), \text{Max}(\log(Q))\right) \). Meanwhile, the priors on \( \sigma_{i,j} \) were constrained to be nonnegative and to be upper bound to 55 mg/L.
i.e $\sigma_{i,j} \sim Unif(0,4)$, which is more than one order of magnitude larger than the maximum observed TON concentration observed in the past 30 years (max TON = 3.235 mg/L reported on 10-24-2000). Finally, the priors on the model intercepts ($\alpha_j$) and slopes ($\beta_{j,k}, \beta_{\text{Temp},l}, \delta$) were given vague normal distribution of the form $\mathcal{N}(0, 10^3)$. The motivation behind adopting vague and noninformative priors for model parameters is primarily due to the desire to reduce subjectivity and let the data drive inference. The specification and use of weak and noninformative priors in Bayesian analysis is common and is discussed in more detail by Gelman et al. (2004), Gelman (2006), and Van Dongen (2006).

We used a MCMC procedure to determine the posterior distributions of these parameters using the Bayesian software package WinBUGS (Lunn et al. 2000). Six chains were initiated at different arbitrary initial values for the parameters and were monitored. The generated posterior distributions were all based on 10,000 MCMC samples. Convergence was assured by monitoring that the potential scale reduction factor, $\hat{R}$, for each parameter was equal to 1.0 (Gelman et al. 2004, Gelman and Hill 2007). The computer code for the proposed Bayesian model is included in the Appendix B.

### 3.2.3 Calculating TON Load Reductions

The posterior predictive distribution for TON concentrations was used to predict the probability that the stipulated 30 % TON load reduction has been achieved in the
Neuse River basin. This was carried out by first drawing 10,000 simulated daily flow measurements from the 30 year daily hydrograph for Fort Barnwell (Figure 22 a). For each flow value, 2,500 corresponding TON concentrations were sampled for both pre and post changepoint periods using the posterior predictive distribution of log(TON) (Figure 22 b). The generated pre and post changepoint TON concentrations were then compared in order to determine the distribution of the percent reduction in TON loading given a specific flow value (Figure 22 c). The process was repeated over the 10,000 simulated daily flow values in order to construct the posterior predictive distribution of TON percent load reduction over all flow values. Note that since we are using the same simulated flow measurements for both time periods, this ensures that a 30 % reduction in concentration translates directly into a 30 % reduction in TON load. This methodology ensures that both periods cover the whole range of average daily flow value. Simulation procedures for both concentration as well as load reductions were conducted using the rv package (Kerman and Gelman 2007) in the statistical software program R (R Development Core Team 2010).
Figure 22: The adopted simulation methodology for estimating the probability distribution for % of TON load reduction goal for a given flow. In panel (a), a flow is drawn from the Fort Barnwell hydrograph. In this example the drawn flow corresponds to 231.48 m$^3$/sec. Panel (b) shows the posterior predictive probability distribution of TON for both pre and post 1999 that correspond to a flow event of 231.48 m$^3$/sec. Panel (c) presents the histogram of the % TON reduction given the pre-1999 and post-1999 distributions in panel (b).
3.3 Results

Model results indicate that when it comes to TON concentrations in the Neuse, the system has experienced a regime shift in 1999 with respect to the flow-TON concentration relationship (Table 1). The posterior distribution around the changepoint sharply peaks at the year 1999 and has almost no associated uncertainty, indicating strong evidence that compels us to conclude that a regime change did indeed occur in 1999. The identified changepoint coincides with the implementation of one of the key mitigation measures that were stipulated as part of the TMDL action plan for the Neuse, which saw the major point-source discharger, namely the Neuse River Wastewater Treatment Plant for the city of Raleigh (design capacity of 227,125 m$^3$/day; currently treating 138,811 m$^3$/day) reduce nitrogen levels in its discharge by 55% from the 1995 levels. Moreover, the 1999 changepoint also coincides with an active hurricane season in North Carolina that saw the landfall of two major hurricanes, namely Floyd in September 1999 as well as Bonnie in August of 1998. A temporal regime change within the Neuse had been suspected earlier by Paerl (2006), who indicated that the increased tropical storms activity between 1996 and 1999 along with basin-scale management measures could have resulted in a change in the dynamics of the phytoplankton biomass (as chlorophyll) in the Neuse estuary. While both point-source reduction and increased flushing from hurricane activity could have played a role in forcing the regime change, we believe that the latter had less of an impact on the system. This is based on the fact
that the flow-concentration relationship in the Neuse following the landfall of hurricane Fran (1996) did not deviate appreciably from the pre-1996 relationship. Furthermore, it is also evident from Figure 23 that the flow-concentration relationships at high river flows are comparable for both pre and post 1999 models, consequently challenging the idea that the system’s response in terms of TON delivery during extreme flow events has changed over time.
Table 6: Posterior model summary statistics for the TON flow-concentration changepoint-threshold Bayesian model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Percentiles 2.5%-97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changepoint year</td>
<td>Year</td>
<td>1999</td>
<td>0.000</td>
<td>1999 - 1999</td>
</tr>
<tr>
<td>$\alpha_1$ (Pre-1999 intercept)</td>
<td>log mg/L</td>
<td>1.837</td>
<td>0.265</td>
<td>1.318 – 2.357</td>
</tr>
<tr>
<td>$\alpha_2$ (log TON concentration at identified flow threshold for the post-1999 model)</td>
<td>log mg/L</td>
<td>0.289</td>
<td>0.032</td>
<td>0.220 - 0.350</td>
</tr>
<tr>
<td>$\beta_{1,1}$ (Pre-1999 slope on log flow)</td>
<td>-</td>
<td>-0.378</td>
<td>0.013</td>
<td>-0.410 - -0.350</td>
</tr>
<tr>
<td>$\beta_{2,1}$ (Post-1999 slope on log flow before flow threshold)</td>
<td>-</td>
<td>1.239</td>
<td>0.149</td>
<td>1.020 - 1.630</td>
</tr>
<tr>
<td>$\beta_{Temp}$ (Slope on water temperature)</td>
<td>-</td>
<td>-0.028</td>
<td>0.001</td>
<td>-0.030 - -0.025</td>
</tr>
<tr>
<td>$\delta$ (Change in the slope on flow subsequent to the flow threshold in the post-1999 model)</td>
<td>-</td>
<td>-1.594</td>
<td>0.145</td>
<td>-2.000 - -1.400</td>
</tr>
<tr>
<td>$\sigma_1$ (Standard error for the pre-1999 model)</td>
<td>log mg/L</td>
<td>0.379</td>
<td>0.008</td>
<td>0.360 - 0.400</td>
</tr>
<tr>
<td>$\sigma_2$ (Standard error for the pre-flow threshold &amp; post-1999 model)</td>
<td>log mg/L</td>
<td>0.537</td>
<td>0.023</td>
<td>0.490 - 0.590</td>
</tr>
<tr>
<td>$\sigma_3$ (Standard error for the post-flow threshold &amp; post-1999 model)</td>
<td>log mg/L</td>
<td>0.399</td>
<td>0.009</td>
<td>0.380 - 0.420</td>
</tr>
</tbody>
</table>
Figure 23: Flow-TON relationship both for pre and post the 1999 system changepoint. Reported TON concentrations have been adjusted for the effect of water temperature. The relationship pre 1999 was linear and dilution dominated. Post 1999, the relationship is piecewise linear with a threshold in flow at 30.429 m³/sec.

The relationship between nutrient concentrations and river flow can be divided into two main behaviors, namely: dilution dominated regimes and flow-driven regimes (Johnson 1979). The results indicate that prior to the 1999 changepoint, the TON concentration-flow relationship is typical of a dilution dominated system, whereby
increased river flows dilutes the overall TON load ($\beta_{1,1} = -0.378$). These findings are consistent with the inverse flow concentration relationship that Stow and Borsuk (2003a) described earlier for the Neuse, using data from 1979 till 2000. Such a relationship is typical of water bodies where point-sources play the main role in nutrient delivery. Note that no flow threshold was identified for the pre-1999 period (and as such there is no $\theta_{1,2}$). Following the identified 1999 changepoint, our findings point to a transition in the relationship between flow and concentration. This transition involves moving the system away from a dilution dominated system towards a more complicated system where the flow-concentration relationship is no longer continuous or monotonic in nature. We believe that this added complexity is a result of the implementation of aggressive point-source load reduction measures that tightened the discharge of nitrogen from point sources. Given that TON levels at low flows are indicative of mainly point-source discharges- as low flows are associated with very limited overland flow- the results point to a significant drop in the contribution of point sources to the TON loads reaching the Neuse at Fort Barnwell post 1999 (Figure 4). In addition, the results show that at low flow levels (below the flow threshold of 30.429 m$^3$/sec) TON concentrations post 1999 tend to increase with flow, as evident by the positive slope on flow ($\beta_{2,1} = 1.239$) (Table 1). This positive relationship can be attributed to the delivery of accumulated nutrients from agricultural fields, croplands, as well as from residential areas through overland flow and drainage canals. This behavior suggests that post 1999
non-point sources have come to play a more prominent role in TON delivery to Neuse. The relationship between flow and TON concentration shifts towards a dilution dominated system past the identified flow threshold for the post 1999 period \((\hat{\beta}_{2,2} = \hat{\beta}_{2,1} + \delta < 0)\) (Table 1 and Figure 4). This shift is most probably due to the fact that past the flow threshold most of the nutrient load on the landscape has already been washed off. These findings have also been corroborated in a recent study by Libo et al. (In Press).

With respect to the seasonal component of the model (namely \(\beta_{\text{Temp}}\)), the posterior mean for the coefficient on water temperature is found to be negative and significantly different from zero (Table 1). This illustrates that riverine water temperature has an inverse relationship with respect to the measured TON concentrations.

It is worth mentioning that given the sampling frequency bias that is inherent to the data (Figure 21), we tested for the possible impacts that this bias might have on the model results. For this purpose, a balanced dataset was generated and used to run the developed model. The results from this test indicated that the model results were robust and did not significantly change as the data was balanced over the years. Details on the adopted method to balance sampling frequency over time, as well as the corresponding generated model results are presented in Appendix C.

A comparison of the predictive distribution for both pre and post changepoint periods indicates that the mean predicted TON concentrations post-1999 are lower than
the levels pre-1999 over all flows (Figure 23). These results are a good indicator that the mitigation measures adopted by the State of North Carolina as part of the Neuse TMDL program have had an impact on reducing TON concentrations in the river and ultimately in the estuary too. Moreover, what is more important is to be able to assess whether the drop in concentration is consistent with the mandated 30 % decrease in nitrogen loading. Simulation results for the change in TON concentration over the Neuse River hydrograph (and as such also loads) indicate that the predicted mean drop in TON concentration post-1999 is around 32 %. Yet, what is more relevant than the mean load reduction in TON is the ability to capture the uncertainty around that reduction. Adopting the previously detailed simulation methodology (section 3.2.3), allows us to easily determine the probability distribution for the TON load reduction. The model findings indicate that the probability that TON loading has achieved the stipulated reduction goal is around 49 %. Note that even though there is strong evidence of a drop in TON loading, the model results still indicate that there is a 28 % chance that the concentrations (and similarly loads) have either not changed over time or have even increased post 1999.

When the change in TON concentration following the 1999 changepoint is more closely inspected, we observe that the percent reduction along with the probability of attaining the 30 % reduction goal is flow dependent. Figure 24 shows that for low flows we are very confident that the reduction goal has been achieved. This can be seen by the fact that the 95 % credible intervals around the load reduction are all above the 30
% load reduction target. However, as flow levels increase the confidence in achieving the 30 % load reduction diminishes, while the uncertainty in the percent change between pre and post 1999 TON levels grows. This is consistent with the observed data that indicate that post the identified flow threshold, the flow-concentration relationship for pre and post 1999 periods are almost identical (Figure 23). This translates to the fact that at high flows the mean percent load reductions are close to zero and the credible intervals around them are considerably wide.

![Diagram](image)

**Figure 24:** Percent reduction in TON levels as a function of flow. Black solid points indicate the predicted mean change in TON levels between pre and post 1999 levels for the defined flow measurements. The dark grey bands correspond to ± 1 standard error credible interval, while the light grey bands indicate ± 2 standard error credible intervals. Dashed black line shows the TMDL stipulated reduction in nitrogen loading reaching the Neuse. Gaps towards the high flow values are a direct result of sampling from a right-tailed distribution.
3.4 Discussion and Conclusions

3.4.1 Dynamics of the Flow-Concentration Relationship

The results from this study strongly support the occurrence of a changepoint during the year 1999 for the Neuse River. The change in system behavior in 1999 is most probably attributed to the implementation of the TMDL program that involved the execution of a set of point as well as non-point source nutrient control measures. While the TMDL recommended mitigation measures were executed over a period of time starting as early as 1997, the introduction in 1999 of biological nitrogen removal processes to the Neuse River Wastewater Treatment Plant, which is the major point-source contributor of TON in the Neuse watershed, seems to have culminated in a regime change when it comes to the flow-concentration relationship in the system.

Since temporal invariance remains at the heart of many empirical water quality models, such as SPARROW and LOADEST, what our findings highlight is the need to adapt these models so that they can better reflect large-scale management decisions that occur on the watershed level over time.

The results from this study also indicate that the classification of the nutrient delivery system in the Neuse River basin by Stow and Borsuk (2003a) as point source dominated is only valid for the timeframe prior to the changepoint year of 1999. Post-1999, the system transitioned into a more complex system that is non-monotonic in nature. This added complexity is a direct result of the implementation of successful point source discharger mitigation measures, while non-point source reductions have
lagged behind at best and thus have come to play an increasingly important role in nutrient dynamics. In the pre-1999 period, one would have expected to find the highest TON concentrations at low flows; post-1999 low flows are now associated with low TON concentrations, while the highest concentrations now occur at medium flow levels (around $30.429 \text{m}^3/\text{sec}$). Moreover, our results indicate that dilution remains the dominating process at high flows in the Neuse as noted earlier by Borsuk et al. (2004b).

We should note that the developed model tends to over predict TON concentrations during major hurricane (Fran in 1996 and Floyd in 1999), when extreme flow events are recorded (Figure 23). This shortcoming of the model may warrant considering the inclusion of a second flow threshold to further govern the flow-concentration relationship at these extreme flows. Yet, we think that the inclusion of another threshold at this stage is not advisable due to the lack of enough data that adequately captures several severe hurricane events in order to ensure that the model is not biased towards a single event.

### 3.4.2 The Effect of Water Temperature on TON Concentrations

Model results suggest that the dynamics of TON concentrations in the Neuse are seasonal in nature. The negative coefficient on water temperatures in the developed model indicates that given the same flow values, winter to early spring TON concentrations are expected to be higher than their counterparts during the summer to early fall period. This is consistent with our understanding of the Neuse system and can
be explained by three main processes. The first involves the seasonal patterns of fertilizer application, with most fertilizer applied during the late winter and early spring period. The second is associated with the seasonal algal dynamics within the river that affect nutrient uptake, with late spring and summer blooms as described by Borsuk et al. (2004b) and Paerl (2006). The last process involves an acceleration of the benthic denitrification rate as water temperatures increase (Alperin et al. 2000, Borsuk et al. 2004b).

3.4.3 Assessing Load Reductions and Their Management Implications

The findings from this work have implications when it comes to assessing the effectiveness of the nitrogen mitigation measures executed by the State of North Carolina as part of the Neuse TMDL. The assessment of compliance for a nutrient like TON is complicated by the fact that the applicable standard of interest is set for chlorophyll concentrations and not for TON concentrations. With the absence of a TON criterion for the Neuse, compliance assessment has to be conducted through the assessment of load reduction over time. Given that load cannot be directly measured and since load itself is mostly dominated by flow rather than concentration, any adopted load reduction verification methodology has to take into account the underlying flow regime over the assessment period. Several studies have tried to quantify the temporal changes in nitrogen loading in the Neuse River Basin. The results from most of these studies have been inconclusive as to whether the reduction in
nitrogen loading has met the 30 % reduction goal or not (Stow et al. 2001, Burkholder et al. 2006, Paerl et al. 2006b, Rajbhandari 2007), with some studies even indicating that nitrogen load has actually increased post-TMDL implementation (McNutt 2007). We think that the major source of this inconclusiveness pertains to three main reasons: 1) the inability of the adopted methodologies to account for a changepoint in time, 2) the failure of some studies to properly account for flow, and 3) finally the use of dissimilar and relatively short time periods by different studies.

Our developed methodology has tried to address these three limitations, while also focusing on accounting for the uncertainty in both model parameters as well as model structure. The results from our modeling approach, concerning load reduction in the Neuse, indicate that the TMDL for the Neuse appears to have had some success in reducing TON levels reaching the Neuse Estuary, with a 32 % estimated mean drop in TON loading. Nevertheless, the results highlight that load reduction is a function of the observed flow regime in the Neuse. At low flow values, model results very clearly show that the load has been reduced by more than 30 % (Figure 24). Since low flows conditions are associated more closely with point-source loading, the results indicates that point source mitigation measures targeting load reduction have been successful. Yet, the model results show that at high flow values the estimated load reductions are significantly lower than the 30 % mark and are associated with a large degree of variability. Our findings are consistent with those reported by Libo et al. (In Press), who reported a marked drop in TON concentrations along several stations in the Neuse River
network. Since the contribution of non-point TON sources increases with overland flow, this implies that either the state run non-point source nutrient control programs (the NC Agricultural Cost Share Program, the Conservation Reserve Enhancement Program, as well as the United States Department of Agriculture’s (USDA’s) Environmental Quality Improvement Program) have had little success in reducing nitrogen washoff or that whatever success has been achieved in reducing non-point source nitrogen loading has been offset by an increase in the total load within the Neuse basin. We tend to think that the latter is more plausible, given the proactive engagement that the state has adopted as well as given the findings reported by the state (Neuse Basin Oversight Committee 2008).

Our findings corroborate the conclusions that have been reported in the 2009 Neuse River Basinwide Water Quality Plan, where nutrient delivery through non-point sources was named as the primary cause of impairment to surface water in the Neuse River basin (Deamer 2009). As such, any future reductions in nitrogen delivery to the Neuse Estuary, particularly TON concentrations, should concentrate on targeting non-point sources in the basin. We strongly believe that in order to increase the efficiency of the currently implemented non-point source nutrient reduction programs more emphasis has to be placed on monitoring and site inspection in order to better address the challenges associated with regulating non-point pollution sources that are diffuse and intermittent in nature, diverse in origin, and hard to identify. Continued efforts towards further load reduction is of particular importance in the Neuse watershed,
where denitrification has been shown by Whalen et al. (2008) to be a minor sink for nitrogen removal, with an average removal rate of around 5% of the nitrogen reaching the river.

Finally, it should be noted that the analysis we conducted has focused solely on TON and not on TN due to data limitations with TN measurements. TN monitoring limitations are not unique to this study, but are common to many river basins (Lewis et al. 1999). This is partly due to the fact that TON monitoring is often preferred by many monitoring agencies, given that it is instantaneous as compared to TN monitoring, which requires an extra digestion step that adds costs, requires further off-site analysis, and can even add bias (Lewis 2002, Dodds 2003). Fortunately, the relationship between TON and TN in most river systems is linear in nature, which allows for the utilization of TON to better describe TN dynamics (Dodds 2003, Turner et al. 2003). This linearity between TON and TN is also evident for the Neuse (Appendix D). Furthermore, it has been shown that TON forms the major constituent of the TN pool in major rivers, particularly in developed watersheds (Dodds 2003, Turner et al. 2003, Sauer et al. 2008). This also holds true for the Neuse River, where TON concentrations contribute around 61% (± 15%) of the measured TN concentrations. The strength as well as the linearity of the relationship between TON and TN permits using the TON results from our model to make conclusions on the TN dynamics in the Neuse, particularly when it comes to TN load reductions. Nevertheless, there has been evidence to indicate that the strength of the linear relationship between TN and TON weakens at low concentrations of TONs,
where the balance between biota uptake and remineralization plays a significant role in observed TON concentrations (Dodds 2003). It is imperative to look more closely at the relationship between TON and TN loads in the Neuse given the recent findings reported by Libo et al. (In Press), who adopted a simplified modeling approach to compute TN loading in the Neuse. They found that organic nitrogen loads have increased at several stations within the Neuse River basin, which counteracts some of the observed reductions in TON loading.

### 3.4.4 Bayesian Framework, Adaptive Management, and Monitoring

The adopted Bayesian methodology provides the ability to quantify the uncertainty in the load reduction and to transparently report the outcomes associated with the success/failure of the adopted nutrient management policies to the decision makers with a full disclosure of the uncertainties involved. The adoption of such a framework has been promoted by Reckhow (2003) and helps policy makers better assess their decisions by facilitating the evaluation of trade-offs between accepting the risk of exceeding the TMDL’s stipulated reductions with a given probability versus incurring additional mitigation and monitoring costs in order to increase the frequency of attainment. Furthermore, the developed model framework removes many of the constraints that are often placed on the relationship linking flow to nutrient concentration. This added flexibility to the model structure allows for a much better understanding of the nutrient transport process and its dynamics in the Neuse River. It
also helped resolve a large portion of the uncertainty in the observed data and to reconcile many of the conflicting studies both with respect to the success/failure of the TMDL program for the Neuse as well as the type of the relationship that governs the way flow affects TON concentration in the Neuse River.

The fact that we had almost 3 decades of uninterrupted data presented us with an incomparable opportunity to better understand nutrient loading and its sensitivity to mitigation measures across time. This work underscores how our understanding of the system evolves as new data are collected. Our understanding of nutrient delivery in the Neuse River Basin started with a simple linear regression model (Stow and Borsuk 2003a), then it evolved to a piecwise threshold model (Borsuk et al. 2004b), and now to a threshold-changepoint model. We observe this succession as a dynamic learning process, whereby we build on our previous knowledge and incorporate new data to continuously update the models to better our understanding of our riverine systems. We strongly believe that our approach fits well with the process of adaptive management or as Walters and Holling (1990) called it “learning by doing”, which has become widely accepted for environmental systems in general and has been endorsed by the National Research Council (NRC) (2001) when it comes to dealing with TMDLs (Reckhow 2003). Even though the results indicates that there is strong evidence to imply that the year 1999 resulted in a changepoint, there still is a crucial need to follow through this analysis as more data become available to make sure that alternative explanations to the model are formulated and evaluated. Moreover, the findings
highlight the need to adopt monitoring programs that operate on long temporal scales that often exceed individual research projects.

While the methodology that we presented has been applied to the Neuse River estuary, it can easily generalized to other systems that have had a large-scale management or perturbation occur that may warrant the presumption of the occurrence of a changepoint in time. Even though the developed model in this study has been constrained to identify a single changepoint in time, it can also be easily expanded to identify multiple changepoints if such an assumption is warranted. Furthermore, the model is not constrained to identify either a changepoint or a threshold if the data do not support such a claim.
4. Evaluating Automated Structure Learning with Bayesian Networks: Chlorophyll Dynamics in the Upper Neuse Estuary

Note: the results and text from this chapter appeared in a recently published peer-reviewed article in Environmental Modeling and Software (Alameddine et al. 2011a). The paper was co-authored by YoonKyung Cha and Dr. Kenneth H. Reckhow.

4.1 Introduction

Environmental systems are complex and are often described by deterministic or stochastic models that tend to simplify the system under study. The stride towards simplification is ubiquitous amongst scientists, who are looking for order and a set of basic laws to account for the complexities observed in nature (Peters 1991). Given that our knowledge is never complete and that our models are merely simplifications of reality, variable selection, model structure development, and properly addressing uncertainty become vital tasks when dealing with environmental systems. BN modeling is well suited to handle these issues.

A BN is a graphical model for reasoning under uncertainty, with a set of variables (or nodes) and directed arcs that describe the sets of conditional dependencies between these variables (Pearl 2000, Korb and Nicholson 2004). The use of BNs as tools for characterizing and predicting the dynamics of complex systems has increased over the past decade. This increase can be attributed to better understanding and addressing of uncertainty, improvements in the computational power of computers, and the
development of easy to use software such as Hugin (Hugin Expert A/S 2008), Netica (Norsys Software Corp 1997), and GeNIe (University of Pittsburgh 2005). Refer to Murphy (2008) and Uusitalo (2007) for a comprehensive overview of commercially available BN software.

A key property of BNs is their ability to reduce the joint probability distribution of the model, which is often intractable and hard to assess, into a set of conditional probabilities through the application of the chain rule (Equation 13) (Pearl 2000, Korb and Nicholson 2004, Jensen and Nielsen 2007). Additionally, the capabilities of BNs to express model uncertainties, propagate information quickly, represent complex topologies, combine subjective expert knowledge with hard data, and update model parameters as new information becomes available, has promoted their use within the environmental research community.

\[ P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | pa(X_i)) \]  
Equation 13

Where \( X_i \) are the variables, \( pa(X_i) \) are the parents of node \( X_i \).

Starting with Varis’s (1995) and Reckhow’s (1999) seminal papers that utilized BN models to address modern environmental challenges, BNs have been used extensively in the environmental field. They have been adopted to conduct water quality assessments (Stow and Borsuk 2003b, Borsuk et al. 2004a, Dorner et al. 2007,

Even with the increased use of BNs, building model structure remains daunting; using previously accumulated subject-matter information to configure the model is common. Such information is often acquired through an expert elicitation process along with supplemental literature synthesis; both are then used to develop a conceptual diagram consisting of boxes and arrows (Bromley et al. 2005, Marcot et al. 2006, McNay et al. 2006, Pollino et al. 2007b, Uusitalo 2007, Borsuk 2008, Kjaerulff and Madsen 2008).

While using expert elicitation to build model structure is adequate in many instances, particularly when one is lacking enough observational data, it may be hindered by the absence of a widely accepted interpretation of the process being studied. For example, many new environmental problems we face today, such as the environmental implications of nanoparticle use, the fate and transport of estrogens, and the initiation of harmful algal blooms in coastal areas, are all relatively new and highly
contentious subjects. Eliciting a prior structure in these cases may often prove very difficult. The use of structure learning algorithms to generate plausible BN model structures under such conditions becomes a major advantage toward beginning to understand these systems.

Structure learning can also help minimize the need for expert elicitation, a process that is time consuming, prone to errors, and might prove to be too expensive to conduct (Korb and Nicholson 2004, Uusitalo et al. 2005, Taleb 2007). Our experience with expert elicitation has shown that when experts are elicited for a model structure that spans different disciplines and/or scales, they often find it hard to agree on a unifying model and are prone to generate elaborate models with a large number of variables, a high degree of interconnections between variables, and the inclusion of feedback loops that violate the acyclical nature of BN models. Such model structures, if not reduced, are often of limited use, particularly when adequate data is not available.

In this chapter, we developed a BN model that describes estuarine algal biomass in the Upper Neuse River Estuary (UNRE) in North Carolina, using data collected following the implementation of the nitrogen TMDL program. Our purpose is to examine the applicability and usefulness of the automated constraint based structure learning algorithms for building model topology in ecological and environmental studies. We have chosen to study the dynamics of estuarine chlorophyll concentrations in the UNRE, as it is a well studied area with considerable subject matter knowledge (Rudek et al. 1991, Stow et al. 2003, Borsuk et al. 2004a, Paerl et al. 2006a, Valdes-Weaver et al. 2007).
2006, Arhonditsis et al. 2007b, Paerl et al. 2007). Previous work on the area has shown that increased river flow generally exerts a negative effect on chlorophyll concentrations in the UNRE, while chlorophyll concentrations have a positive correlation with temperature and nitrogen concentrations. Since work in the study area was conducted pre-TMDL, we were also interested in using the BN model to assess any changes that may have occurred to chlorophyll dynamics post-TMDL.

We begin by providing an overview of the different structure learning algorithms that are currently used to build BN model structures. These algorithms are then applied to the chlorophyll dataset for the UNRE to generate plausible model structures. The resulting BN models are subsequently evaluated and compared based on a number of criteria. We then examine the results from the selected model to draw inference about the post-TMDL chlorophyll dynamics in the UNRE. The chapter also discusses the effects that various parameters have on the structure learning process and provides a summary of best practices that can be adopted when executing these algorithms.

4.2 Materials and Methods

4.2.1 Study Area

The Neuse River drains an area of approximately 16,000 km². The river has its headwaters to the north of the city of Durham and travels approximately 443 km before discharging into the Pamlico Sound just below the city of New Bern (Figure 25). The Neuse River and its associated estuary have experienced eutrophication, which was
evidenced by extensive algal blooms, fish kills, hypoxia and anoxia that captured public attention in the 1980s and 1990s (Paerl et al. 1998). These events prompted the listing of the estuary on the 303(d) list, with nitrogen identified as the main culprit behind eutrophication (Stow et al. 2003). In 1998, the USEPA settled a lawsuit brought by the Neuse River Foundation which required North Carolina to establish a TMDL for nitrogen reaching the estuary. The TMDL was approved by USEPA on August 26, 1999 (NC Department of Environment and Natural Resources 1999). Meanwhile, the State of North Carolina, through the NCDWQ, adopted a set of rules in 1997 that aimed to reduce the amount of nitrogen delivered to the Neuse River Estuary by 30% based on 1991-1995 loads.

4.2.2 Data Description

Chemical and physical data were collected for the study area. Mean daily flow measurements were collected through the USGS NWIS service for the Fort Barnwell Station (refer to Figure 25). Fort Barnwell was selected as the point of interest, as it drains around 85% of the Neuse watershed and is the flow gauge used by most studies on the Neuse River and Estuary. Chlorophyll concentrations, nutrient concentrations, as well as water temperatures for the period between 2000 and 2005 were compiled from the ModMon program (http://www.unc.edu/ims/neuse/modmon), using data collected every other week at monitoring station # 30 in the upper section of the Neuse Estuary (Figure 25). ModMon is a collaborative program that was established between the
University of North Carolina and the State of North Carolina. ModMon’s water quality sampling is conducted by the Institute of Marine Sciences at the University of North Carolina at Chapel Hill. The distributions of the selected variables in this study as well as the bivariate co-plots between the variables are presented in Figure 26. The data show that water temperature has a negative correlation with both dissolved inorganic nitrogen (DIN) and nitrites and nitrates (NO₃), while a positive relationship with Ortho-Phosphate (PO₄) is observed. Other relationships, especially those involving flow and chlorophyll are obscured by the log-normal nature of both distributions.

Figure 25: The locations along the Neuse Estuary of Station 30 from the Modmon monitoring program as well as the USGS gauge station in relation to the city of New Bern, NC.
Figure 26: Matrix scatterplot of ModMon selected variables for Station # 30 with the corresponding histograms. The scatterplot plots each variable against the other and fits a locally weighted scatterplot smoothing (LOESS) line to help discern the underlying relationship between each pair. The histograms illustrate the distribution of each variable. Temperature (Temp) is plotted in °C; Nitrates and nitrites (NO$_x$), dissolved inorganic nitrogen (DIN), Ortho-Phosphate (PO$_4$) as well as Chlorophyll (Chl-a) are all plotted in (μg/l); Flow is plotted in (m$^3$/day).

4.2.3 Structure Learning Algorithms

Structure learning in BNs is conducted through the use of collected data. Each set of observations is termed a case, with each case representing a realization of all the variables within the network. Given that BN models generally require the use of discrete
variables, continuous observations need to be properly discretized. While complete cases are preferred, the expectation-maximization (EM)-algorithm is often used to salvage information from incomplete cases (Little and Rubin 2002, Jensen and Nielsen 2007).

Structure learning from a set of case files is conducted under a set of assumptions that include: 1) the existence of a “true” network in the real world that describes the process at hand; 2) the cases are actually sampled from that “true” network; 3) the cases have been fairly sampled with no intentional bias; 4) the number of cases is large enough to cover the network adequately; and 5) there are no hidden variables in the system (Pearl 2000, Korb and Nicholson 2004, Jensen and Nielsen 2007). Even under the stated assumptions, structure learning may only be able to generate a set of Markov equivalent structures. Markov equivalent structures are models that share the same set of variables and are equally supported by the joint probability distribution of the available data (Korb and Nicholson 2004, Kjaerulff and Madsen 2008).

One may be tempted to find the most adequate model structure by considering all possible permutations of structures. However, the space of all possible BNs that describe a given set of variables grows faster than exponential with the number of variables (Robinson 1977), as seen in Figure 27. This makes such an approach inefficient. Clearly, a more efficient approach is required to learn the structure of a BN. Two main approaches have been developed based on machine learning techniques, namely the constraint based and the score-based approaches (Jensen and Nielsen 2007).
approach makes use of the concept of conditional independence and directional-separation (d-separation) to build the network. D-separation is a concept used to test for conditional independence. Two variables (e.g., “C” and “D”) are said to be d-separated given a set of variables Z if and only if Z blocks every path from variable “C” to variable “D” (Pearl 2000, Jensen and Nielsen 2007). The score-based approach scores several candidate models based on a defined criterion and selects the one with the highest score. Both approaches automate the process of structure learning in BNs and have shown promise in retrieving the structure from simulated data. It should be noted that some algorithms combine aspects from both of these approaches when searching for the most plausible structure. In this chapter, we focus primarily on the constraint based approach with emphasis on the Peter-Clark (PC) and Necessary Path Condition (NPC) algorithms.
4.2.3.1 PC-Algorithm

The PC algorithm is a constraint based structure learning algorithm that was developed by Spirtes and Glymour (1991) and Spirtes et al. (2000), as an upgrade to the Inductive Causation (IC) algorithm developed earlier by Verma and Pearl (1992). The PC algorithm is a standard structure learning algorithm in the Hugin software (Hugin Expert A/S 2008). The PC algorithm starts by first connecting all of the variables being assessed...
by undirected arcs. This leads to the generation of a fully connected skeleton, whereby all nodes are neighbors to one another. Subsequently, the algorithm operates iteratively to trim down the fully connected model through a set of tests that remove arcs between variables that are d-separated and retain the arc connecting two variables if and only if there exists no possible subset in the BN space that can make the two neighboring variables conditionally independent given that subset (Korb and Nicholson 2004, Jensen and Nielsen 2007). The PC algorithm proceeds through several iterations whereby it searches for variables with an increasing number of neighbors and assesses their conditional independence with respect to those neighbors. Any conditional independence that is found between a variable (e.g., variable “X”) and any of its neighbors (e.g., variable “Y”) given the remaining neighbors (e.g., the set Z, which includes variables “C”, “D”, and “E”), will directly result in arc removal between the two variables (e.g., the arc between X and Y is removed if \( X \perp Y | Z = [C, D, E] \), where \( \perp \) indicates independence and | symbolizes conditionality. The iterative steps of the PC algorithm are summarized in Table 7.
Table 7: Summary of the main steps of the PC algorithm

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Start with a complete undirected graph</td>
</tr>
<tr>
<td>Step 2</td>
<td>Start with $i = 0$</td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td>Look for all variables with at least $i+1$ neighbors and place these nodes in a set which we will call $\Delta$</td>
</tr>
<tr>
<td><strong>Step 4</strong></td>
<td>Look for the neighbors of each variable node that is an element of $\Delta$. Each element in $\Delta$ will have a set of neighbors as identified from the undirected graph</td>
</tr>
<tr>
<td><strong>Step 5</strong></td>
<td>Start with the first variable in $\Delta$ and see if it is conditionally independent of any of its identified neighbors, given the rest of its neighbors that are to be combined into sets of size $i$.</td>
</tr>
<tr>
<td><strong>Step 6</strong></td>
<td>If conditional independence is proven between that variable and one of its neighbors given the rest of the neighbors, then proceed to remove the arc linking the variable and that neighbor. Proceed until you have analyzed all identified neighbors.</td>
</tr>
<tr>
<td><strong>Step 7</strong></td>
<td>Move on to the next variable in $\Delta$ and repeat the process identified in Steps 5 and 6 until all the variables in $\Delta$ have been cycled through. This will generate an undirected graph that has less than or an equal number of arcs as compared to the one we started out with.</td>
</tr>
<tr>
<td><strong>Step 8</strong></td>
<td>Increment $i$ by one ($i = i+1$) and repeat Steps 3 through 7. The PC algorithm is terminated once there are no variables with $i+1$ neighbors</td>
</tr>
</tbody>
</table>

One may wonder how knowledge concerning conditional independence can be acquired given that the structure is unknown. The answer lies within the observational data itself. Statistical significance testing checks for the presence or absence of conditional independence between two variables based on the cases in the dataset (Spirtes et al. 2000, Korb and Nicholson 2004). The test assumes that the null hypothesis ($H_0$) states that the two variables under consideration are conditionally independent (or marginally independent), while the alternative hypothesis indicates that the variables
are dependent (Kjaerulff and Madsen, 2008). The hypothesis test then uses a likelihood test statistic, such as the \( G^2 \), to reject or accept \( H_0 \) based on a user defined significance level \((\alpha)\). \( G^2 \) is asymptotically Chi-squared distributed \( (\chi^2_{df}) \) with \( df \) degrees of freedom. Details concerning the calculation of the \( G^2 \) score and its degrees of freedom are described in Kjaerulff and Madsen (2008). Whenever the tail probability of the \( \chi^2_{df} \) distribution of the test statistic \( G^2 \) is smaller than the user defined significance level \((\alpha)\), the null hypothesis is rejected in favor of the alternative hypothesis (Kjaerulff and Madsen 2008). Failing to reject the \( H_0 \) leads to the removal of the arc connecting the two variables.

The finalized skeleton produced by the PC algorithm is then converted into a Directed Acyclical Graph (DAG) by applying four prioritized rules (Jensen and Nielsen 2007). The first rule requires creating a number of local V-structures within the finalized skeleton. The introduction of V-structures, which are also known as colliders, should be explored whenever there are three nodes (which we will call nodes “C”, “D”, and “E”) that are connected in a manner whereby an arc connects “C” to “E” and another arc connects “D” to “E”, but there is no connection between “C” and “D”. If that situation holds while at the same time we find that the nodes “C” and “D” are conditionally (or marginally) independent given any set of nodes that excludes “E”, then a V-structure is to be introduced. The resulting structure between the three nodes would thus be \( C \rightarrow E \leftarrow D \). V-structures allow for the incorporation of the “explaining away” concept.
that allows for intercausal inference (Kjaerulff and Madsen 2008). After introducing as many V-structures as possible within the skeleton, the second rule prohibits the addition of any directed arrow that would lead to the creation of a new V-structure. The third rule requires that any arc that could lead to the creation of a cycle should have its direction reversed. This principle preserves the acyclical nature of BNs. After exhausting Rules 1 to 3, any remaining undirected arcs are given a random direction. This last step is termed Rule 4. The stochastic nature of the fourth rule can lead the structure learning procedure to identify multiple equivalent solutions for the same set of variables.

4.2.3.2 NPC-Algorithm

Besides the PC algorithm, the NPC algorithm is the standard structure learning algorithm in the Hugin software (Hugin Expert A/S 2008). Although the NPC algorithm shares the same basic procedures detailed in the PC algorithm, it enforces the execution of an additional criterion called the necessary path condition. This condition limits the search for conditional independence between two variables “X” and “Y” to the variable sets that have their elements along paths linking “X” to “Y” (Jensen and Nielsen 2007, Kjaerulff and Madsen 2008).

The NPC algorithm often results in the introduction of ambiguous regions. The ambiguity is a direct result of the fact that retaining some arcs is dependent on the removal of other arcs linking other variables, and vice versa. Ambiguous regions indicate inconsistency in the conditional independence and dependence relations that are
derived by the statistical tests (Jensen and Nielsen 2007, Kjaerulff and Madsen 2008). These inconsistencies can be attributed to a variety of factors that include measurement error, failure to recognize and incorporate additional variables within the model, absence of a clear cause-effect relationship between the variables, and natural variability of some processes.

The Hugin software allows the user to interact and resolve these ambiguous regions based on user judgment and expertise. In the event that the user is incapable of resolving some of these ambiguous regions, the default setting in Hugin will be applied, which incorporates the smallest number of ambiguous links possible (Hugin Expert A/S 2008). It should be noted that the creation of ambiguous regions enables the NPC algorithm to better address the inconsistencies of conditionally independent and dependent relations.

4.2.4 Defined Structure Learning Scenarios

The primary purpose of our developed BN model is to predict chlorophyll levels given the nutrient concentrations, water temperature, and the river flow regime. As such, chlorophyll concentration, a surrogate for the amount of phytoplankton, is selected as the end point of interest in the model. Five scenarios were defined by choosing different user defined settings that are known to affect model structure recovery by the constraint based structure learning algorithms. Table 8 provides a summary of the five scenarios that were adopted in this study.
Table 8: Defined scenarios to test the sensitivity of the constraint based structure learning algorithms to several user defined options

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
<th>Modification to base scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Base Case</td>
<td>NPC algorithm executed with 3 bins defined for each variable, while adopting an equal frequency distribution of cases amongst the bins. Significance level ($\alpha$) was set at 0.05</td>
<td>Not applicable</td>
</tr>
<tr>
<td>(b) PC versus NPC</td>
<td>Compares the Base Case structure with the structure generated by the PC algorithm while preserving the binning and significance level options selected for the Base Case scenario.</td>
<td>Change in type of algorithm</td>
</tr>
<tr>
<td>(c) Changing binning method</td>
<td>Compares the Base Case structure with the structure generated by running the NPC algorithm while changing the distribution of cases in the bins from equal frequency to equal interval. Number of bins for each variable and the significance level ($\alpha$) were preserved.</td>
<td>Change in binning method</td>
</tr>
<tr>
<td>(d) Increasing the number of bins for each variable</td>
<td>Compares the Base Case structure with the structure generated by running the NPC algorithm while increasing the number of bins defined for each variable from 3 to 7. Equal frequency distribution of cases and the significance level ($\alpha$) were preserved.</td>
<td>Change in number of bins</td>
</tr>
<tr>
<td>(e) Decreasing the significance level</td>
<td>Compares the Base Case structure with the structure generated by running the NPC algorithm while changing the significance level from 0.05 to 0.001. Number of bins and case distribution between bins were preserved.</td>
<td>Change in significance level</td>
</tr>
</tbody>
</table>

4.3 Results

4.3.1 Generated Model Structures

Different model structures were generated based on the defined scenarios, indicating that the constraint based structure learning algorithms are highly sensitive to changes in the selected settings governing the algorithms. Figure 28 presents the
different model structures that were generated for algal biomass (measured in terms of chlorophyll concentrations) in the UNRE. It is evident that while the generated structures change from one scenario to the other, some structures are more closely related to the base structure (Figure 28 (a)) than others, indicating a varying degree of sensitivity between the defined scenario settings.

The structures generated by the PC and NPC algorithms are often different. The NPC generated structure (Figure 28 (a)) had an ambiguous region between Temp, NO$_x$, and DIN; the ambiguity is primarily due to the high correlation between DIN and NO$_x$ (Pearson’s correlation = 0.94). As a result of this high correlation, Temp and DIN were found to be independent conditional on NO$_x$ (DIN⊥Temp| NO$_x$); similarly, Temp and NO$_x$ were independent conditional on DIN (NO$_x$⊥Temp| DIN). Moreover, both DIN and NO$_x$ are marginally dependent on Temp. This situation requires the removal of the Temp - DIN link and the retention of the Temp - NO$_x$ link, or the other way around. In our case, we selected to remove the Temp – DIN link and retain the Temp - NO$_x$ link, which resulted in the final structure presented in Figure 28 (a). In the case of the PC algorithm, the algorithm removed the links between Temp and DIN and those between Temp and NO$_x$ based solely on the findings of (DIN⊥Temp| NO$_x$) and (NO$_x$⊥Temp| DIN), thus generating the structure in Figure 28 (b). Comparing Figure 28 (a) with (b), it becomes clear that the NPC algorithm generates a structure that better describes our understanding of chlorophyll dynamics in the estuary, as compared to the structure generated by the PC algorithm. The model generated by the PC algorithm, while similar
in many ways to the base-case model, has several arcs that are misdirected. For example, PO₄ points towards the ambient temperature node, while DIN is the parent of the NOₓ node. As previously mentioned, these discrepancies are largely the result of the stochastic nature in which arcs are introduced by the fourth rule in the PC algorithm.

![BN structure comparison](image)

*Figure 28: BN structure comparison as a function of the selected constraint based algorithm, the number of bins, the binning method, and the significance level. Note that all NPC generated models included an additional step that involved resolving ambiguous regions and selecting a directed path for the undirected arcs.*
All BN software, including Hugin, require the discretization of continuous variables into binomial/multinomial variables, with each variable including a given number of intervals that cover the original continuous variable space (Korb and Nicholson 2004, Jensen and Nielsen 2007). Discritizing results in partitioning the data into a set of bins (also known as states or ranges). Discritization is required to speed up both model parameterization and information propagation, while allowing for the use of the Dirichlet-Multinomial conjugacy to perform Bayesian updating of the model as new data are made available. Nevertheless, the process of binning a continuous variable always leads to a loss of statistical power (Myllymaki et al. 2002, Uusitalo 2007, Gelman and Park 2009).

Our results have found that the binning process has a profound impact on the generated model structure. By comparing Figure 28 (a) and (c), one can clearly see the impact of setting different cutoff points on the final generated structures. The equal frequency binning based model indicates that chlorophyll concentrations can be predicted directly by river flow, water temperature, and PO₄ (Figure 28 (a)). All of these links are lost when the equal interval binning process is used (Figure 28 (c)), implying independence between chlorophyll levels and all other variables in the system. Clearly such a model does not capture our understanding of chlorophyll dynamics in the UNRE, which is known to be flow and temperature dependent (Borsuk et al. 2004b, a, Valdes-Weaver et al. 2006, Arhonditsis et al. 2007a).
Finding the most optimal binning procedure, when dealing with continuous variables in BNs, is still an active area of research in machine learning. Recent developments in the field of discretization have primarily focused on supervised discretization methods (Liu et al. 2002 and references therein). Nevertheless, these methods are often hard to implement in BNs due to the lack of relevant class information. It should be noted that Hugin and other BN software have started to allow for the incorporation of mixed BN models that are capable of handling some continuous nodes in conjunction with discrete nodes, particularly when the relationship is linear Gaussian. However, the restrictions and assumptions required in such models are often difficult to meet, thus limiting their wider use (Kozlov and Koller 1997, Jensen and Nielsen 2007).

Changing the number of bins that are assigned for each variable also results in changes to the generated model structure. Increasing the number of bins will change the values of the $G^2$ test statistic thus changing the generated structure. As expected, increasing the number of bins from 3 to 7 per variable resulted in: 1) the generation of a model structure that had fewer arcs connecting the variables to each other (Figure 28 (d)), 2) the introduction of additional ambiguous regions, and 3) an increase in the number of uncertain links in the intermediate model structure. These results are attributable to having too few cases in each of the defined bins, making it hard to ascertain conditional (in)dependence and to determine the direction of the relationships linking variables. These results highlight the problems associated with increasing the
number of bins when trying to discretize continuous variables, a process that has also been shown to result in a drop in the performance of the developed BN model due to over-fitting (Sun and Shenoy 2007). Deciding on the number of bins and the adopted binning method remains one of the more challenging tasks associated with developing BNs with continuous variables.

Finally, the significance level ($\alpha$) is another user defined measure that impacts the generated BN structure. Decreasing the value of $\alpha$ will result in the generation of fewer links between two variables; smaller $\alpha$ increases the probability of accepting the $H_0$. As expected, reducing the value of $\alpha$ from 0.05 to 0.001 led to fewer directed arcs in Figure 28 (e). Note that while the number of arcs was reduced, the remaining structure (which includes the more strongly defined relationships) was very similar to the original structure in Figure 28 (a). On the other hand, adopting a larger value for $\alpha$ will increase the number of arcs and type-I errors that manifest in an increased ambiguity in arc directions within the generated BN structure.

### 4.3.2 BN Model Evaluation

BN models should be evaluated both in terms of their abilities to describe the data at hand satisfactorily and in terms of their usefulness in answering the main question of interest (Myllymaki et al. 2002). While the first criterion is relatively easy to quantitatively determine using such measures as maximum likelihood, Akaike's information criterion (AIC), and the Bayesian information criterion (BIC), assessing
model usefulness is more often done qualitatively. Table 9 summarizes the scores that each model structure received based on the AIC, BIC, and log-likelihood measures. As expected, AIC and BIC measures for any model are always smaller in magnitude than the log-likelihood measure given that both measures incorporate a penalty term that accounts for model complexity and rewards parsimonious models.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>AIC</th>
<th>BIC</th>
<th>Log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Base Case</td>
<td>-702.51</td>
<td>-808.75</td>
<td>-626.51</td>
</tr>
<tr>
<td>(b) PC versus NPC</td>
<td>-711.70</td>
<td>-812.35</td>
<td>-639.70</td>
</tr>
<tr>
<td>(c) Changing binning method</td>
<td>-452.42</td>
<td>-497.15</td>
<td>-420.42</td>
</tr>
<tr>
<td>(d) Increasing the number of bins for each variable</td>
<td>-1542.27</td>
<td>-2045.52</td>
<td>-1182.27</td>
</tr>
<tr>
<td>(e) Decreasing the significance level</td>
<td>-703.75</td>
<td>-703.70</td>
<td>-679.75</td>
</tr>
</tbody>
</table>

While it might seem tempting to compare the five generated models based on their BIC, AIC, or log likelihood scores (Table 9), such comparisons are not always possible or prudent. Comparisons across models using these information criteria are only valid when the likelihoods of all models are conformable to each other (StataCorp LP. 2009). In this study, some of the generated model structures have different endpoints and/or are assessing different events based on the adopted binning process. Attempting to quantitatively compare models based on score criteria does not make
much statistical sense in these cases. From the five structures we generated, we can use
the scoring criteria to compare the models from scenarios (a), (b), and (e) since these
three structures fit the same data (use the same binning process) and include the same
endpoint, namely chlorophyll a. The comparison indicates that both scenarios (a) and (e)
outperform the structure implied by scenario (b).

Other ways of comparing across structures could be done through assessing
model usefulness or through model corroboration and verification. Usefulness can be
evaluated based on maximizing the ability of a model to answer the question(s) of
interest, incorporating the relevant endpoint (chlorophyll concentration in this case),
and showing adequate predictive ability and precision. For example, the model
generated by equal interval binning (Figure 28 c) is useless as a decision-support tool,
given that: 1) it places more than 95 % of all observed phosphate and chlorophyll
concentrations in one of the three assigned bins for each, and 2) it fails to establish any
links between any of the predictor variables and chlorophyll. In such a case, the model’s
discretization is very coarse and adds little to our knowledge of the dynamics of
chlorophyll concentrations in the estuary. This coarseness allows the model to achieve a
high score (Table 9), yet this is achieved at the expense of building a useful model that
can help predict chlorophyll concentrations from other predictors at a reasonable
resolution.

Model corroboration and verification on the other hand focus on testing the
model’s compatibility with current events while also assessing the model’s performance
on an independent data set (Reckhow and Chapra 1983). Some of the validation tools include Relative Operating Characteristics Curves (ROC) and quantitative reliability indices (Garrote et al. 2005, Hugin Expert A/S 2008). Given the lack of an independent dataset to use at this stage, we did not attempt to conduct model verification.

4.3.3 Model Interpretation: Post-TMDL Chlorophyll Dynamics

Based on the five developed models, the base-case scenario model is the most informative and instructive. It mirrors our understanding of how the system works by indicating that flow and temperature are the most important factors affecting chlorophyll concentration in the UNRE followed by phosphate concentrations. The strength of links between the variables has been assessed by computing their mutual information values. A high mutual information value indicates that a variable is informative with respect to the response variable, while a value of zero signifies independence between two variables (Korb and Nicholson 2004).

The results from the BN model suggest that the relationships between flow and temperature on one hand and chlorophyll dynamics on the other have remained unchanged from those reported for the pre-TMDL period (Mallin et al. 1991, Rudek et al. 1991, Paerl et al. 1998, Borsuk et al. 2004b). The model indicates that flow plays a strong negative role on chlorophyll concentrations in the UNRE. This negative relationship is attributed to high flow levels resulting in reduced residence time, lowered salinity, and increased turbidity (Borsuk et al. 2004b). Flows in excess of
10.3×10^6 m^3/day result in a 60% probability of experiencing low chlorophyll concentrations (< 2.3 μg/L). In contrast, increases in water temperature promote algal growth. Water temperatures in excess of 23.5 °C are associated with a 58% probability of observing high chlorophyll concentrations (>8 μg/L), as these temperatures represent the spring and summer growing seasons. Overall, low flows (< 10.3×10^6 m^3/day) and high water temperatures (>23.5 °C) seem to be the most conducive combination for algal growth. Under these conditions the probability of observing high chlorophyll concentrations (> 8 μg/L) is greater than 96%. Our results support the findings of Arhonditsis et al. (2007a), whose structural equation model results stressed the strong influence of the physical environmental variables (flow and temperature) on chlorophyll dynamics in the UNRE.

Phosphorus concentrations (as PO_4) were found to play a minor role in the algal dynamics of the UNRE post-TMDL. Surprisingly, model results indicated that nitrogen (measures as DIN and NO_x) did not play a role in modulating chlorophyll levels in the UNRE, as suggested by the absence of a link between the two nitrogen species and chlorophyll (Figure 28a). This is at odds with previous findings that found that nitrogen concentrations had a positive effect on chlorophyll levels in the Neuse Estuary (Paerl et al. 1998, Pinckney et al. 1999, Borsuk et al. 2004b).
4.4 Discussion and Conclusions

4.4.1 Structure Learning

The model results indicate that structure learning in BN is highly sensitive to several settings that are often difficult to set by the user, especially when a system is poorly understood and hard to characterize. While this limitation may reduce the appeal of using constraint based structure learning algorithms to construct BNs for environmental applications, we think that the structure learning process provides a new perspective on the problem, a better appreciation of the complexity of nature, and a better understanding of the system and the limitations of the monitoring programs.

Although it may seem trivial, defining the main goal(s) of the BN model before conducting structure learning is paramount. Such a step will dictate the end point of interest, the selection of an appropriate binning procedure, the definition of the spatio-temporal scale, as well as the identification of the variables that need to be considered. Some suggestions that may help simplify the process of structure learning are summarized here based on the model results:

- First, when it comes to the binning process, make sure that the selected binning is not affected by outliers that can reduce accuracy and skew the model results. Robust multivariate outlier detection techniques are useful in identifying and minimizing the impact of outliers (Alameddine et al. 2010). When equal interval binning is used, outliers can affect the range of the data to be discretized, resulting in few sparsely populated bins. Indeed, one needs to ensure that enough data
points are present in each bin. Selecting such a number is very subjective and depends on the complexity of the model; Liu et al. (2002) recommend a minimum of 6 cases in each bin. Moreover, one should keep in mind that while a BN model does not require any transformation for the data, care should be taken to make sure that the selected binning technique is appropriate given the underlying (marginal as well as conditional) distributions of the variables. An optimal discretization would provide finer partitioning for the regions where the variable distribution changes rapidly while allocating wider bins for areas that are relatively flat (Kozlov and Koller 1997). Assigning bin cutoffs points, where an environmental threshold is expected or where an environmental standard is set, is critical. One should always keep in mind that once the binning is decided upon, all data points falling within a given bin are treated equally. Failing to incorporate important cutoffs will result in the loss of important connections or the addition of faulty links to the network.

- Second, adding new variables as well as changing the spatio-temporal scale might result in the identification of new structures with links being deleted, or new links being added, or exiting links being reversed. For example, the fact that we used instantaneous measurements in this study is a major source leading towards less well defined links between chlorophyll levels and nutrient concentrations. Indeed, the lack of a link between dissolved nitrogen and chlorophyll concentrations, as described earlier, is a clear example of the temporal mismatch that conceals the
relationship between the two variables. One should always keep in mind that establishing a cause-effect relationship is always conditional on the “universe” one defines (Pearl 2000).

- Third, BN structure learning assumes the existence of a large number of cases that adequately cover the “true” network. In an ideal world there is an infinite set of observations, yet in the real world sampling is often constrained by limited resources and time. As such, the user should always constrain the number of variables to include and the associated number of bins assigned for each. Nonetheless, one should keep in mind that failing to account for an important variable can lead to the generation of ambiguous regions and could distort the generated BN structure. Unaccounted variables play the role of hidden variables. While it may seem tempting to increase the number of bins in order to achieve better accuracy and precision, it is often the case that such an act will lead to increased computational complexity, ambiguity, and uncertainty in the model predictions. Under such cases, conditional probabilities are often populated with few cases (that could be outliers) or in the absence of cases fulfilling the underlying conditions forced to have uninformative probabilities. These limitations can be partially surmounted by deciding on the binning process prior to data collection and adopting a targeted sampling design.

- Fourth, bear in mind that the structure learning algorithms are not perfect. These algorithms have been shown to miss dependencies even when analyzing simulated
data from a constructed model, particularly as the model gets more complicated.

Given that environmental data are inherently noisy, generated model structures should be examined by domain experts (when possible) for consistency and usefulness or through designing a monitoring program based on the implied structure, whereby additional data are collected to validate the proposed model. Nevertheless, it should be taken into account that any finite set of observations in nature can be explained equally well with a potentially infinite set of theories and models (Peters 1991).

4.4.2 Chlorophyll Dynamics in the UNRE

The implementation of the TMDL program in the Neuse Estuary has changed our understanding of algal biomass dynamics in the UNRE. While the links between chlorophyll levels and the physical environment (river flows and water temperature) have remained unchanged post-TMDL, the relationship between chlorophyll and dissolved nutrient concentrations has been modified.

The lack of a link between dissolved nitrogen and chlorophyll in the BN model raises several questions, particularly since the Neuse Estuary, like most temperate estuaries and coastal areas, is known to be nitrogen limited (Nixon 1995, Paerl et al. 1998, Paerl et al. 2004, Smith and Schindler 2009). We believe that the absence of this link can be attributed to three main reasons. The first is related to the temporal mismatch between algal uptake and growth on one hand, and the instantaneous
dissolved nitrogen concentrations on the other. This mismatch weakens any potential link between nitrogen and chlorophyll. Paerl et al. (1998) had previously reported a lag between nutrient concentrations and chlorophyll levels in the Neuse Estuary. The second reason can be attributed to the success of the TMDL program in reducing nitrogen concentrations reaching the UNRE and in damping their temporal variations. Prior to TMDL implementation, the UNRE showed a more distinct seasonal nitrogen loading cycle, with high values observed in the late summer (Qian et al. 2000). Current concentrations show occasional dips during the summer (Figure 29); these dips are probably associated with increased uptake rates by phytoplankton. Nevertheless, other than those dips, nitrogen concentrations are temporally stable to a large extent. This stability partially explains the lack of a link between nitrogen and chlorophyll in our BN model. The last reason is the interaction that inorganic nitrogen has with inorganic phosphorus and their combined impact on primary production in the UNRE.
Figure 29: Temporal variations of (a) Dissolved Inorganic Nitrogen (DIN), (b) Ortho-phosphate (PO$_4$), (c) DIN:PO$_4$ ratio (molar), and (d) chlorophyll (Chl-a) in the Upper Neuse Estuary. The concentrations of DIN, PO$_4$ and Chl-a are in μg/L and are log transformed. The grey shaded areas represent the summer season (from June 21 until Sep 21) of each year. The solid horizontal grey line in each graph represents the mean of the corresponding variable.
The model results indicate that phosphorus (measured as PO$_4$) plays a minor role in regulating algal biomass in the UNRE. PO$_4$ concentrations show a clear temporal signal, with strong peaks in the summer season (Figure 29). This relationship is depicted by the BN model in the form of a link between water temperature and PO$_4$ levels. This link, along with the lack of a significant association between flow and PO$_4$ levels, points to the importance of sediment release in the phosphorus dynamics. Previous studies conducted on the Neuse have described this process (Paerl et al. 1995, Paerl et al. 1998). The connection between PO$_4$ levels and chlorophyll concentrations is relatively weak (mutual information = 0.05). The low information value between PO$_4$ and chlorophyll is mostly due to the strong correlation between PO$_4$ concentrations and water temperature, interaction with dissolved nitrogen species, and the temporal lag.

The importance of the interaction between the dissolved species of nitrogen and phosphorus on the dynamics of chlorophyll concentrations in the UNRE pre-TMDL implementation has been highlighted by Qian et al. (2000). They found that in the late 1990s the DIN:PO$_4$ molar ratio in the UNRE varied from over 140:1 in the winter to near 30:1 in the summer. In the present study, the DIN:PO$_4$ molar ratio varied between 150:1 and 1:5, with a median of 30:1. These changes signify a drop in DIN concentrations in relation to PO$_4$ levels post-TMDL. Note that high DIN:PO$_4$ ratios are observed in the winter and early spring seasons (Figure 29). The ratio’s temporal variations are dominated by the seasonality of PO$_4$ concentrations. High ratios in the winter and early spring indicate that phosphorus is potentially limiting, while low ratios in the summer
point to increased nitrogen limitations, as phosphorus becomes readily available due to sediment release.

Our findings support the conclusion that nitrogen load reduction, through the TMDL program, has changed the way algae responds to nutrient loading. The observed drop in the DIN:PO$_4$ ratio is a clear indication that nitrogen loading to the Neuse Estuary has dropped. These findings also corroborate the need to adopt a dual phosphorus and nitrogen management strategy to achieve effective eutrophication management in estuaries similar to the Neuse. This dual nutrient management approach to combat eutrophication has been promoted by Paerl et al. (2004) and Conley et al. (2009). Yet, recently there is evidence to indicate an increase in the loading of organic nitrogen (Lebo et al. In Press), which has been shown to be readily assimilated by estuarine phytoplankton (Bradley et al. 2010).

In this chapter, we have tried to explore the functionality of the structure learning capabilities that can be implemented in many BN software programs. To our knowledge, this is the first attempt to assess the BN structural learning algorithms with real-time data from an environmental system. A combination of structure learning along with expert involvement seems to be most promising. A domain expert can first express his/her beliefs and background knowledge before allowing the structure learning algorithms to make use of the data to fill gaps in our knowledge of the system being studied. All constraint based structure learning algorithms allow the expert to specify known (in)dependences in the system prior to executing the structure learning
algorithm. Providing information on (in)dependencies whenever it is available can greatly improve the model, as shown by our intervention to resolve the ambiguous regions generated by the NPC algorithm. This hybrid approach towards structure learning allows for the incorporation of an expert’s career worth of experience into the process of model formulation. This combined approach showed its potential when adopted by Wooldridge and Done (2004), who built a BN network for predicting coral bleaching in Australia.

We hope that our results highlight the complexity of establishing (or even interpreting) causal links between environmental variables, even when dealing with a relatively simplified model of aquatic chlorophyll dynamics. The complexity of most environmental systems, the interplay of variables and processes operating at different spatio-temporal scales, the lack of adequate monitoring data, the difficulty of inducing interventions, as well as our limited understanding of the systems we study, leads us to question our ability to identify causal relationships in ecological systems solely based on observational data. Ascertaining that a directed link is in fact a cause-effect relationship requires the adoption of set of causal calculus rules (Pearl 1995, 2000). These rules provide a suitable mathematical platform to discover causal relationships. Yet, embarking on such a task is challenging and more often than not unfeasible, particularly due to our limited ability to intervene when we are dealing with observational data.
5. Optimizing an Estuarine Water Quality Monitoring Program Through an Entropy-Based Hierarchical Spatio-Temporal Bayesian Framework

5.1 Introduction

The collection of environmental monitoring data has the fundamental purpose of gaining a better understanding of the system we are monitoring. In an ideal world with no financial, temporal, or technical constraints, monitoring can be conducted at an infinite number of sites with instantaneous data collection and analysis. While recent advances in remote sensing, deployable sensors, and autonomous vertical profilers have reduced the above-mentioned constraints, monitoring at predefined locations remains by far the most common method for environmental data collection. Yet, the process of locating monitoring stations is more often than not driven by convenience and makes little use of more efficient methods that have been developed (Sanders et al. 1987, Berthouex and Brown 2002, Simeonov et al. 2003, Strobl and Robillard 2008). An exhaustive review of common approaches used in developing water quality networks is beyond the scope of this study. Interested readers are encouraged to consult the review paper by Strobl and Robillard (2008).

Much of the advances in the selection of monitoring locations have been brought about by developments in the geostatistical field (Christakos 2000, Banerjee et al. 2003). Recognizing and accounting for both the spatial and temporal correlations of environmental variables provides the proper framework towards optimizing an existing
monitoring network or designing a new one. However, acknowledging the spatio-temporal correlations through a model is insufficient by itself to guarantee a proper optimization. There is a need to couple that effort with an efficient approach that allows for the quantification of information (gain or loss) resulting from any proposed network redesign. The adoption of the concept of entropy from information theory allows for the quantification of system uncertainties (and thus the system’s information content) over an appropriate probability distribution (Caselton et al. 1992, Christakos and Li 1998, Christakos 2000). The concept also lends itself towards assessing uncertainties associated with water quality standard violations. Assessing these violations has great relevance from a management perspective particularly with respect to the TMDL program that entails the expenditure of significant resources (~ $ 17.3 million/year) in order to monitor more than 40,000 standard violations in around 20,000 impaired water bodies across the US (Environmental Protection Agency 2001, National Research Council 2001).

Entropy-based optimization methods have been mostly implemented towards redesigning air quality networks (Le and Zidek 2006, Puangthongthub et al. 2007, Ainslie et al. 2009). Moreover, their adoption for redesigning water quality networks has been less common. Some studies have successfully incorporated entropy in the design of water quality/quantity networks; however, most have fallen short of making full use of the spatio-temporal correlations in the data (Harmancioglu and Alpaslan 1992, Ozkul et al. 2000, Markus et al. 2003, Karamouz et al. 2009). Additionally, these studies were
constrained by the need to discretize the underlying distribution of the variable of interest in order to simplify the calculation of entropy. Although discretizing the distribution on the water quality parameter simplifies the calculations of entropy, it runs the risk of not achieving a unique design solution, as the estimated entropy values are dependent on the underlying discretization method. LoBuglio et al. (2007) were able to resolve this limitation by developing a Bayesian maximum entropy model that was capable of handling continuous distributions. Yet, their model assumed a spatially homogenous and temporally stationary process as they attempted to optimize chlorophyll (Chl-a) monitoring in the Catawba River reservoir system in North Carolina.

Though assuming isotropy and stationarity may be valid in some cases, these assumptions are often hard to justify in dynamic systems, such as estuaries. Riverine inputs, wind driven mixing, and saltwater influx lead to complex circulation patterns and make the estuaries anisotropic and non-stationary. Failing to properly account for the spatio-temporal correlations in the data can lead to the loss of information, inaccurate model predictions, and can ultimately impede monitoring optimization in these complex water systems.

In this paper, we explore the use of entropy as an information measure to properly quantify system uncertainties, while using a spatio-temporal hierarchical Bayesian model that is not restricted by either spatial isotropy or stationarity. Moreover, the adopted approach is not constrained towards optimizing a monitoring
network based on a single water quality parameter nor on a particular criterion or attribute.

We apply our developed methodology on the Neuse River Estuary in North Carolina in order to identify locations where our knowledge about the system is most limited given the existing monitoring program. The methodology does not place any constraints on the random spatial fields assigned for the environmental variables of interest. This makes the approach particularly valuable to adopt in flow-complex water bodies.

The adopted framework also incorporates a MADM process, which allows us to integrate several competing design criteria in order to generate a single optimal design. To our knowledge this is a novel approach towards assessing the adequacy of existing water quality monitoring networks and towards identifying vulnerable areas that would benefit most from any additional monitoring.

This chapter is organized as follows. In section 5.2, we describe the existing Neuse Estuary monitoring network and present our Bayesian Hierarchical spatio-temporal model. We then describe the entropy-based optimization criteria that we adopted along with the Analytical Hierarchical Process (AHP) that we implemented to weight and prioritize the different optimization criteria/attributes. In section 5.3, we present the optimization results from each of the design criteria and compare them with the final compromise design. Finally, in section 5.4 we summarize our adopted methodology and
results and discuss the associated limitations. It is hoped that our work will be of use for both water quality monitoring agencies and policy makers.

5.2 Methods

5.2.1 Study Area and Data Description

The Neuse River drains a 16,000 km$^2$ watershed. The river has its headwaters to the north of the city of Durham and travels approximately a distance of 443 km before discharging into an estuary located just below the city of New Bern. The basin has a diverse landuse/landcover. Just east of its head waters, an urbanized area -that includes the cities of Raleigh, Durham, and Cary- dominates the basin. Intensive agricultural areas and CAFOs become more prominent towards the lower portions of the Neuse basin, where the river traverses the North Carolina coastal plain. Continued landuse changes in the basin have resulted in an increase in the relative importance of non-point nutrient sources to the overall Neuse nutrient budget. The estuary has been increasingly under pressure from anthropogenic activities in its contributing watershed as well as from climatic perturbations (Paerl et al. 2006b, Paerl et al. 2009a, 2010). The estuary eventually connects to the Pamlico Sound after traveling around 78 km (Figure 30). The Neuse Estuary is the largest of the four major riverine tributaries feeding into the Pamlico Sound (PS) (Paerl et al. 2007).
Figure 30: (a) Locations of the existing monitoring stations that form the ModMon monitoring program. Inter-station separation distances range between 1.5 km (between stations 60 and 70) and 9.4 km (between stations 160 and 180), (b) the 25 potential locations generated from the 5x5 km representation of the Neuse Estuary.

The river and its associated estuary have experienced eutrophication with extensive algal blooms, fish kills, hypoxia and anoxia in the 1980s and 1990s (Paerl 1987, Paerl et al. 1995, Borsuk et al. 2003, Stow et al. 2003). The Neuse Estuary is known to be a predominantly nitrogen-limited system when it comes to primary production and algal biomass formation (Paerl 1987, Rudek et al. 1991, Paerl et al. 2010). The estuary was listed on the 303(d) list of impaired water bodies (Stow et al. 2003). In 1998, the USEPA settled a lawsuit brought by the Neuse River Foundation which required North Carolina to establish a TMDL for the nitrogen reaching the estuary. The TMDL was approved by the USEPA on August 26, 1999 (NC Department of Environment and Natural Resources
Following the implementation of the TMDL program, chlorophyll concentrations in the estuary have decreased and the frequency of standard violations (chlorophyll ≥ 40 μg/L) have been significantly reduced. Similarly, bottom DO levels have since increased, yet they still drop below the 4 mg/L standard during the summer season (Figure 31b). Currently many sections of the estuary have been delisted from the federal 303(d) list of impaired water bodies; nevertheless, around 16 percent of the total estuary area remains impaired (Deamer 2009). Given these changes in both mitigation and legislation, reassessing the existing monitoring plan at this point to take into account these changes is constructive.

Monthly surface chlorophyll and bottom DO concentrations between 2000 and 2005 were compiled from the ModMon program (http://www.unc.edu/ims/neuse/modmon). ModMon was initiated in 1997 and remains operational (Luettich et al. 2000b, Valdes-Weaver et al. 2006). One of the main goals of the program has been (and still is) to gain a better understanding of primary production dynamics in the Neuse Estuary in light of changing anthropogenic and climatic forcings. Figure 30(a) shows the locations of the ModMon monitoring stations along the Neuse Estuary. Note that the existing monitoring stations are mostly placed equidistantly from each other. Such a design is usually the most appropriate in the absence of prior information on the system.

To better characterize the study area, a 5x5 km grid was constructed to account for the estuary’s shape. The centroids of the pixels were then used as ungauged
locations on which we were interested in quantifying prediction uncertainty. The generated grid yielded 25 potential locations (Figure 30b). We found that the selected 5 km resolution grid was the most optimal as it achieved a reasonable approximation of the irregular shape of the estuary with a limited number of pixels. While we could have selected a finer resolution to better cover the estuary, increasing the number of potential locations beyond the limits of the data (11 stations monitored over 5 years) can pose problems from a Bayesian computation perspective, as it becomes harder to guarantee that the predictive confidence intervals at the ungauged sites will remain simultaneously valid (Le and Zidek 2006).

Monthly chlorophyll concentrations were log-transformed to normalize the data. The use of log-transformation is a common practice when dealing with environmental data, which approximately follow a lognormal distribution (Environmental Protection Agency 1994, Ott 1995, Murphy and Morrison 2002). Bottom DO concentrations did not undergo any transformation.

The chlorophyll concentrations across the ModMon stations showed spatio-temporal patterns. Measured concentrations start out low in the upper section of the estuary before increasing downstream. They reach a maximum around the bend section (between station 120 and station 140 in Figure 30a) before decreasing as nutrient delivery drops and salinity levels increase. These dynamics are consistent with chlorophyll patterns previously documented for the Neuse Estuary (Pinckney et al. 1998, Qian et al. 2000, Borsuk et al. 2004b, Paerl 2006). The system also shows some seasonal
dynamics with early spring and fall blooms and elevated chlorophyll concentrations during the summer season (Pinckney et al. 1998). As expected, the winter season is the least productive (Figure 31a). As observed previously by Qian and Reckhow (2007) monthly chlorophyll levels in the estuary seldom exceed the set water quality standard of 40 μg/L (Figure 31 a,b). Bottom DO levels, on the other hand, showed a pronounced temporal pattern with limited spatial structure (Figure 31 c,d). DO levels in the estuary experience a summer minima with the median values for June, July, and August all falling below the set 4 mg/L standard (Figure 31c). This drop is mainly due to increased water temperatures, less frequent vertical mixing, and increased bacterial activity in the sediment. These patterns are consistent with those described previously by Borsuk et al. (2001b).
Figure 31: Variations in chlorophyll concentrations in the Neuse Estuary by months (a) and across the monitoring stations (b). Variations in Dissolved Oxygen (DO) levels in the Neuse Estuary by months (c) and across monitoring stations (d). The notch in each box represents the median. The solid grey area represents the interquartile range (IQR). The whiskers capture the outliers within 1.5 times the IQR. Values that are beyond the 1.5xIQR are shown as solid points in the graph. The dash-dotted lines represent the 40 μg/L and the 4mg/L standards for chlorophyll and bottom DO concentrations in North Carolina, respectively.
5.2.2 Bayesian Hierarchical Model

We adopt a Bayesian hierarchical model that jointly models the distribution on surface chlorophyll concentrations (log transformed) and bottom DO levels in the Neuse Estuary. This permits us to account for both the inter-variable correlations and the observed spatio-temporal patterns (Figure 31). The model incorporates time varying covariates, which included year and month as categorical covariates, along with the average monthly water temperature in the Neuse Estuary as a continuous predictor. The decision to have monthly average water temperature as an additional predictor in the model stems from the fact that both surface chlorophyll and bottom DO concentrations are known to vary as a function of the water temperatures (Borsuk et al. 2001b, Borsuk et al. 2004b). These temporally varying covariates allow the model to account for differences in the annual and monthly means. The incorporation of a high-frequency component (months) along with a long-term trend is a common approach when dealing with environmental time series (Qian et al. 2000). The use of temporal covariates within the hierarchal model eliminates the need to detrend and deseason the data in a separate step and accounts for strong AR(1) autocorrelations. This approach has also been shown to reduce spatial correlation leakage as described by Le and Zidek (2006).

Our two response variables are modeled through a bivariate random field that can be described by a joint matrix normal distribution (Equation 14). The mean of this matrix distribution is itself a matrix that can be expressed as a linear regression model
that predicts the mean of the two environmental variables concurrently using a set of appropriate covariates.

**MODEL:**
\[ Y \mid X, B, \Sigma \sim \text{MVN}(XB, I_t \otimes \Sigma) \]  
Equation 14

**PRIORS:**
\[ B \mid B_0, \Sigma, F^{-1} \sim \text{MVN}(B_0, F^{-1} \otimes \Sigma) \]  
Equation 15
\[ \Sigma \mid \Psi, \delta \sim \text{GIW}(\Psi, \delta) \]  
Equation 16

Where \( Y \) is an \((t \times 2n)\) dimensional matrix of the monthly mean concentrations of the two environmental variables under consideration at \( n \) sites over \( t \) time steps. \( X \) is a \((t \times k)\) dimensional matrix, where \( k \) is the number of temporally changing but spatially constant covariates. \( B \) is a \((k \times 2n)\) matrix of regression coefficients that are allowed to vary over sites (Le and Zidek 2006, Pollice and Jona Lasinio 2010). \( \otimes \) is the Kronecker product operation on two matrices and \( I_t \) is a \( t \times t \) identity matrix.

The regression coefficients matrix, \( B \), was assigned a prior multivariate matrix normal (MVN) distribution (Equation 15). The covariance between sites as well as the correlation between the two response variables is captured in the covariance matrix \( \Sigma \) (Equation 14). \( \Sigma \) is separable into two components namely, the between-variables covariance matrix that is fixed across space, and the between-site covariance matrix, which captures the spatial correlations. \( \Sigma \) was assigned a conjugate Generalized Inverted Wishart (GIW) prior distribution (Equation 15). The defined priors on \( B \) and \( \Sigma \) have a set
of hyper-parameters that include $B_0$, $\Psi$, $F^{-1}$, and $\delta$. While in a formal Bayesian methodology one would add an additional layer to the above hierarchy by assigning priors (often diffuse) to the hyper-parameters themselves, in this study we use empirical Bayes to define the hyper-distributions. This improves computation and has been previously shown not to bias the generated model posteriors (Le and Zidek 2006). Details concerning the assignment of the hyper-distributions is beyond the scope of this paper and can be found in Le and Zidek (1994). The use of Empirical Bayes in water quality assessment is common particularly when dealing with a hierarchical or nested data (Solow and Gaines 1995, Reckhow 1996, Butcher et al. 2003, Canham et al. 2003). The defined hierarchical Bayesian model (Equation 14-16), leads to a joint posterior predictive distribution that is a matrix-Student $t$-distribution. A more complete overview of the model specification is provided in Le and Zidek (2006) and is therefore not repeated here.

5.2.3 Entropy

Entropy (Shannon’s entropy) is a measure of the amount of information that is missing before learning the occurrence of an event (Caselton et al. 1992). It describes the uncertainty associated with a given probability distribution. Entropy for the continuous bivariate random field is defined by $H(Y) = E \left[ -\frac{\log f(Y)}{h(Y)} \right]$, where $f(Y)$ is the joint probability distribution for chlorophyll concentrations and bottom DO levels and $h(Y)$ is a measure that represents complete ignorance. $h(Y)$ is included in order to
ensure that entropy is invariant to transformations (Jaynes 1963, Le and Zidek 1994).
We use entropy to quantify model uncertainty, which is used to guide monitoring
optimization in the Neuse Estuary. Three different entropy-based criteria are adopted,
namely total system entropy, chlorophyll standard violation entropy, and DO standard
violation entropy.

5.2.3.1 Total System Entropy

The optimization of the monitoring design based on total system entropy focuses
on minimizing residual entropy following data collection (Caselton et al. 1992, Le and
Zidek 1994, Le and Zidek 2006). This entails choosing to monitor locations that have
maximum values of entropy given the existing data. These stations hold the maximum
amount of latent information. Monitoring at these locations guarantees that the
uncertainty in model parameters and in the predictive distributions at the rest of the
ungauged locations will be minimized (Le and Zidek 2006). Since the posterior predictive
distributions depend on the model parameters, reducing the uncertainty of these model
parameters is also a key objective to be incorporated in optimizing the network
(Caselton et al. 1992, Le and Zidek 2006). The optimization proceeds by selecting the
locations whose entropy is greatest. Monitoring those locations will maximize system
information (Ainslie et al. 2009). This is guaranteed by the fact that the total a priori
entropy of the system is fixed (Le and Zidek 1994).
5.2.3.2 Standard Violation Entropy

While the system’s entropy is an adequate criterion when it comes to choosing the most appropriate sites for locating additional monitoring stations, it does not explicitly account for violations in the relevant water quality standards. To account for these violations, we need to account for violation probabilities. The chlorophyll standard in North Carolina is set to $40\mu g/L$, while that for bottom DO states that levels should not drop below $4\text{ mg/L}$. We calculated the probability of chlorophyll standard violation over time and across all locations in order to identify locations that have the highest violation probabilities. Similarly, we quantified the violation probabilities for bottom DO.

While violation probabilities can be used to design a network, they will not always guarantee the most efficient selection of monitoring sites. This is due to the fact that they are bias towards the selection of locations that have the highest probability of violations. From an information theory perspective, the locations that are in need of monitoring are those whose compliance status we are most uncertain about. The ability to identify these locations can be accomplished through the calculation of their violation entropies. Note that for low probabilities of violation both probability and entropy based designs lead to similar solutions. However, the two methods diverge in their conclusions when the violation probabilities increase. Locations with high probabilities of violation will have low corresponding entropies. The latter reaches its maximum when the probability of violation is $50\%$. We argue that using violation entropy is more relevant both from an information and management perspective.
5.2.4 The AHP Framework

The optimization of the monitoring network based on each of the three adopted criteria resulted in a different solution. Therefore, an efficient approach was needed to weight and incorporate the different design elements from the three criteria in order to generate a single optimal solution. This was achieved by adopting a MADM process that weighs the three design attributes or optimization criteria. The same approach was also used to rank the 25 potential monitoring sites given each of the three criteria.

In this study, we adopted the AHP approach for MADM. AHP is an efficient decision analysis tool that can be used to rank a finite number of options or alternatives based on a set of multiple attributes (Saaty 1977, 1986, Saaty 1987, Saaty 1990, 1994, Yoon and Hwang 1995, Saaty 2005, Stein and Mizzi 2007). This is achieved through subjective pairwise comparisons. The procedure is one of the most widely used MADM tools (Marshall and Oliver 1995, Vaidya and Kumar 2006, French et al. 2009). Several AHP procedures have been developed over the years, with conventional AHP being mostly used when a single decision-maker is involved. Advances in AHP have focused in expanding the capabilities of the procedure to handle group decisions and its associated uncertainties. Multiplicative AHP is one of these advances (Barzilai et al. 1987, Lootsma 1988, Lootsma 1993, Van Den Honert and Lootsma 1997, Huang et al. 2009).

In the present study, we have applied both conventional and multiplicative AHP to prioritize the selection of monitoring stations and to verify the stability of the
obtained results by the two techniques. The results did not show any sensitivity towards the adopted AHP procedure. This implies a stable solution.

MADM methods require information on the relative importance of each attribute either in ordinal or cardinal scale. Given that ranking multiple attributes at the same time places a heavy cognitive burden on decision-makers, pairwise judgment is often the preferred approach (Morris 1964). In this study, a pairwise comparison matrix (PCM) was developed and the Saaty scale was adopted for ranking. Saaty's scale ranges from 1 to 9, with 1 indicating equal or no preference and 9 for extreme preference between 2 actions/criteria (Harker and Vargas 1987, Pohekar and Ramachandran 2004, Saaty 2005). In order to rank the three monitoring optimization criteria, we elicited the opinion of four water quality experts. The adopted questionnaire is presented in Appendix E. The results from each expert was equally weighted and combined. This implies that our pairwise inter-expert matrix was populated with a Saaty score of 1, indicating that there was no preferences for valuing the opinion of one expert over the others.

A similar PCM approach was adopted to rank the 25 ungauged locations based on each of the three criteria. However, in this case the PCMs were numerically derived from modeled entropy values calculated at each of the 25 locations. Inter-station entropy ratios were determined and normalized. Ratios greater than or equal to 1 were then linearly interpolated to the Saaty scale. Their transpose in the 25x25 comparison matrix evidently had ratios less than unity. Those were assigned the reciprocal value of
the Saaty score of their respective transpose (Saaty 1986, Harker and Vargas 1987, Saaty 1990, 1994).

Consistency tests on the PCMs were conducted through the calculation of the Consistency Ratio (CR) (Saaty 1990, Stein and Mizzi 2007). This step is important to ensure that the priorities provided by the four experts were internally consistent. Appendix F presents a detailed account of the adopted AHP process.

5.3 Results

Model results were able to capture both the long-term as well as the seasonal patterns observed for surface chlorophyll and bottom DO concentrations. The results indicated that on average the chlorophyll concentrations have decreased post 2000, with the median concentration dropping from 13.1 µg/L in 2000 to 8.8 µg/L in 2005. Bottom DO concentrations on the other hand, were found to have increased during the same period, with the largest increase occurring in 2005. The observed improvements in the water quality of the Neuse Estuary are probably associated with the implementation of the nitrogen TMDL program in the Neuse River basin that limited flow-driven nitrogen delivery to the estuary (Alameddine et al. 2011b).

The observed seasonal patterns for surface chlorophyll levels (Figure 31a) were accounted for by the model coefficients on the months. The highest mean monthly concentrations occurred between July and September. Yet, most of these coefficients were not statistically significant. The seasonality in bottom DO levels (Figure 31c) was
mostly accounted for by the seasonal variations in water temperatures. This strong negative linear relationship between bottom DO concentrations and the water column temperature measurements has been previously documented by Borsuk et al. (2001b) for the Neuse Estuary.

The correlation between bottom DO levels and surface chlorophyll concentrations was found to be weakly negative (-0.13). On the other hand, the spatial correlations between the existing stations in ModMon were captured in the inter-station correlation matrix. As expected, stations that were spatially closer together had higher correlations (Table 10). Yet, the magnitudes of these correlations were not constant across space. We observed that the spatial correlations between the stations and their immediate neighbors in the upper part of the estuary became weaker as we moved downstream. This is probably due to changes in the riverine flow dynamics along this section of the estuary. Note that the scale of the spatial correlations increased once again past station 70 (Table 10).
Table 10: The marginal correlation matrix between the currently monitored stations in the ModMon program. Station 0 is located towards the beginning (river section) of the estuary, while Station 180 is the last station currently monitored before the Neuse Estuary opens to Pamlico Sound (Figure 30). Spatial correlations are evidently stronger in the lower part of the estuary. Grey shaded cells indicate correlations > then 0.5.

<table>
<thead>
<tr>
<th>Station</th>
<th>0</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>100</th>
<th>120</th>
<th>140</th>
<th>160</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00</td>
<td>0.70</td>
<td>0.18</td>
<td>0.11</td>
<td>-0.08</td>
<td>0.04</td>
<td>-0.07</td>
<td>-0.03</td>
<td>0.00</td>
<td>0.09</td>
<td>0.20</td>
</tr>
<tr>
<td>20</td>
<td>1.00</td>
<td>0.56</td>
<td>0.26</td>
<td>-0.06</td>
<td>-0.01</td>
<td>-0.18</td>
<td>-0.15</td>
<td>-0.13</td>
<td>0.01</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.00</td>
<td>0.55</td>
<td>0.15</td>
<td>-0.02</td>
<td>-0.09</td>
<td>-0.06</td>
<td>-0.10</td>
<td>-0.07</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.00</td>
<td>0.44</td>
<td>0.36</td>
<td>0.16</td>
<td>0.19</td>
<td>0.02</td>
<td>-0.10</td>
<td>-0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>1.00</td>
<td>0.70</td>
<td>0.41</td>
<td>0.19</td>
<td>0.12</td>
<td>-0.04</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>1.00</td>
<td>0.54</td>
<td>0.40</td>
<td>0.24</td>
<td>0.19</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.00</td>
<td>0.60</td>
<td>0.42</td>
<td>0.18</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>1.00</td>
<td>0.65</td>
<td>0.48</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>1.00</td>
<td>0.61</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>1.00</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The lack of uniformity in the spatial correlations in the Neuse Estuary system hampered the use of krigging, which assumes second-order stationary and isotropy (Qian 1997, Banerjee et al. 2004). As such, the methodology developed by Sampson and Guttorp (SG) was adopted to deal with nonstationarity and possible anisotropy (Guttorp et al. 1992, Sampson and Guttorp 1992, Guttorp and Sampson 1994). Under the SG methodology, the spatial correlations were modeled by an isotropic semivariogram that was fit using the inter-station separation distances measured in a deformed plane termed the dispersion space (D-space). The D-space is created by iteratively warping the geographic space to generate a plane where the spatial
dependency of the bivariate random field could be assumed stationary and isotropic. An exponential variogram was then fit using the data in D-space (Guttorp et al. 1992, Sampson and Guttorp 1992, Guttorp and Sampson 1994). Mapping between D-space and geographic space was made possible by means of a thin-plate spline (Wahba and Wendelberger 1980). For a more in-depth explanation of the SG method we refer the reader to the references cited above.

The deformation results for the Neuse Estuary did not show drastic deviations from the original geographic space. However, these deformations showed an overall trend of spatial contractions along the direction of the flow lines (northwest to southeast direction), while stretching occurred along the axis perpendicular to flow. This is expected given the unique shape of the Neuse Estuary, and the fact that the monitoring stations are located along the flow centerline. The results also showed that Stations “60” and “70” appeared to be more far removed from Stations “50” than what Euclidian separation distances alone would indicate. This highlights the existence of appreciable differences between the upper and middle sections of the estuary.

5.3.1 Optimizing Based on Total System Entropy

The optimization based on total system entropy proceeded by identifying the locations that would add the most to our knowledge about the system. In this study, we limited our analysis to the identification of a single station. Yet, the same methodology can be expanded to explore the addition of multiple stations. Clearly from an entropy
perspective, the most suitable station to add will be the one whose future monitoring is expected to carry the most information for the system as a whole. This entails reducing both the predictive uncertainty for surface chlorophyll and bottom DO levels in the Neuse Estuary, as well as decreasing the model parameter uncertainties. This is achieved by looking for a location that has the largest log determinant computed from the residual covariance matrix conditional on the observed data (Le and Zidek 1994, Le et al. 1997, Le and Zidek 2006).

Figure 32 shows the relative magnitude of the entropy measure for each of the 25 proposed sites. Of these sites, monitoring at E1 would add the most information to the system. The choice of E1, which is located towards the lower Neuse Estuary, is not surprising given that the station is located the furthest away from the Neuse flow centerline. This also reinforces our previous finding that the spatial correlations are weaker along the axis perpendicular to the flow direction. Moreover, a closer examination of the estimated covariance matrix for the 25 ungauged locations showed higher model uncertainties associated with that location. Adding a monitoring station at E1 would give us a better understanding of surface chlorophyll and bottom DO dynamics in that section of the estuary and add the most information to the system as a whole. Note that location E2, which had the second highest entropy, is also located away from central flow line. The location with the largest entropy, but still located on the main flow path of the estuary, was found to be located in E3 (Figure 32). The selection of E3 was expected given that our initial assessment of the correlations between existing stations
revealed that the spatial correlation between Station “50” on one hand and Stations “60” and “70” on the other were lower than expected (Table 10). That section of the estuary is probably the section where changes in riverine flow, nutrient loadings, and salinity levels are largest.

![Map of proposed locations](image)

**Figure 32:** The locations of the proposed 25 locations are shown in the map. The relative size of each symbol is proportional to the computed entropy for that station. E1 is the station with the largest system entropy and as such is the most optimal site to augment the existing ModMon monitoring program. E2 and E3 are second and third in terms of entropy, respectively.

### 5.3.2 Optimizing Based on Standard Violation Entropy

Standard violation entropy was calculated through model simulation. A thousand joint samples were drawn from the posterior predictive distribution of the bivariate
random field across the set of proposed locations over a period of six years. For each
time-step (t), we calculated the water quality standard violation probabilities ($P_{DO}$ and
$P_{chl}$) at the 25 ungauged locations. The violation probabilities were estimated as the
fraction of observations violating a given standard (Equation 17 and Equation 18). Our
adoption of a simulation-based assessment avoided integration over the high
dimensional space of the joint bivariate random field.

$$P_{Chla_{i,t}} = \frac{\sum_{j=1}^{1000} I(Chl_{a_{i,t,j}} \geq 40 \mu g/L)}{1000}$$

**Equation 17**

$$P_{DO_{i,t}} = \frac{\sum_{j=1}^{1000} I(DO_{i,t,j} \leq 4 \text{ mg/L})}{1000}$$

**Equation 18**

Where $Chla_{i,t,j}$ = the chlorophyll concentration at site $i$, and time $t$, $P_{Chla_{i,t}}$ is the
associated probability of standard violation. Similarly, $DO_{i,t,j}$ = the bottom DO level at
site $i$, and time $t$, $P_{DO_{i,t}}$ is the associated probability of standard violation at that site for
that time period. $I$ is an indicator function that returns 1 if the condition is satisfied and
zero otherwise.

We then calculated the associated Bernoulli violation entropy for each time-step
using Equation 19. Given that the predictions at time $t$ are independent of $t+1$ and $t-1$
given the model parameters, the joint entropy for each potential monitoring location
across time can be estimated through Equation 20. Choosing the locations with the
largest standard violation entropy ensures that priority is given to sites where our
knowledge on standard violations is most uncertain. Given that we have two water
quality standards to account for, we generated an optimized design based on each of
the two standards.

\[ H(X_t) = - p_t \times \log_2 p_t - (1 - p_t) \times \log_2 (1 - p_t) \]  
Equation 19

\[ H(X_1, \ldots, X_t) = H(X_1) + \ldots + H(X_t) \text{ iff } X_1, \ldots, X_t \text{ are } \perp \]  
Equation 20

Where \( p_t \) represents the probability of standard violation at time \( t \) at a given
location and \( \log_2 \) is the binary logarithm.

With respect to chlorophyll standard violation, site CHL1E was associated with
the largest violation entropy (Figure 33 b). Note that the second (CHL2E) and third
(CHL3E) locations with highest chlorophyll violation entropies were located in the
immediate vicinity of CHL1E. This spatial pattern indicates that the section of the Neuse
Estuary that stretched between Stations “30” and “70” is associated with the largest
amount of uncertainty with respect to chlorophyll standard compliance. This is primarily
due to the elevated chlorophyll concentrations observed in that section of the estuary
as a result of nutrient delivery and favorable physical conditions (Figure 31).

Nevertheless, the model results indicated that there was a very small probability of
exceeding the chlorophyll standard across the 25 locations in the Neuse Estuary (Figure
33 a). The predicted median probability of standard violation across the 25 ungauged
locations over the 6 years of analysis did not exceed 2 percent. This is not surprising
given that between 2000 and 2005, the standard was only exceeded twice (Qian and Reckhow 2007). Under low violation probabilities, the corresponding exceedance entropies are low as well. Therefore the optimization results based on either probability or violation entropy produced comparable results as seen in Figure 33.
Figure 33: (a) Chlorophyll standard exceedance probability based on the 40 µg/L bottom standard at the 25 proposed locations. The size of each point is proportional to its exceedance probability; (b) standard violation entropy for the 25 proposed locations. The size of each point is proportional to its standard violation entropy value.
For bottom DO concentrations, the same approach was adopted. Figure 34 compares the optimization results that are based on DO standard violation probability with those based on the violation entropies at the 25 proposed locations. Given that bottom DO levels often drop below the 4 mg/L, the entropy optimization (Figure 34b) was significantly different from that generated based on violation probabilities (Figure 34a). The differences were compounded by the strong seasonality that DO levels exhibit in the Neuse estuary (Figure 31c).

Violation probabilities for bottom DO in many stations showed a strong bimodal behavior, with one mode concentrated at 0 % for the months when the water temperatures in the estuary were low, and another mode around 100 % for summer and early fall months when temperatures were high. A probability of violation-based design will thus be biased towards selecting stations that had very strong and distinct seasonal cycles with consistent violations all across the summer and fall seasons. On the other hand, the entropy based design will be more predisposed to select stations that exhibited less predictable seasonal patterns. This is clearly seen in Figure 35, which compares the violation probabilities across time for station DO1P, which had the highest violation probability and station DO1E that had the largest standard violation entropy. The temporal variations of DO1P are more regular as compared to those at DO1E, which is also co-located with location E1, with the largest total system entropy (Figure 32). Similarly, location DO2E was co-located with E2. These two stations showed significant
violation probabilities and a less structured seasonal signal. Similarly, DO3E showed a disordered seasonal pattern for bottom DO, which results in its high violation entropy.
Figure 34: (a) DO standard violation probability based on the 4 mg/L bottom DO standard at the 25 proposed locations. The size of each point is proportional to its violation probability; (b) standard violation entropy for each of the 25 proposed locations. The size of each point is proportional to its standard violation entropy value.
Figure 35: The probabilities of violating the bottom DO standard of 4 mg/L at the two locations DO1P and DO1E across time. The former was associated with the largest violation probability, while the latter was associated with the largest DO violation entropy. The solid horizontal line represents the 50% probability of standard violation. The dotted horizontal lines represent the 25 and 75% chance standard violation. Note that x-axis shows the months and the years.

5.3.3 AHP Based Results: A Compromise Solution

The multiplicative AHP approach was used to find a compromise solution between the three competing optimization designs. The expert panel inter-criteria comparison resulted in prioritizing the three design criteria. The total system entropy based optimization design received a 75% overall weight, the chlorophyll standard-violation entropy design was assigned a 16% overall weight, and the bottom DO
The CR of the aggregated PCM for the panel of four experts was found to be 0.053, which is less than the 0.10 threshold level defined by Saaty (Saaty 1990, Stein and Mizzi 2007). This indicated a consistent judgment. Similarly, the 25x25 PCM matrices that were constructed for inter-site comparisons had CR values below the Saaty score threshold (CR=0.011 for total system entropy, CR=0.014 for chlorophyll standard violation entropy, and CR=0.008 for DO standard violation entropy). This indicated consistency in derivation.

The final ranking of the 25 locations based on the three criteria were computed by multiplying the expert priority matrix by the inter-site priority matrices for the three attributes. The final optimization results based on the three criteria are shown in Figure 36. It is apparent from the final optimization results that the south-eastern shoreline of the Neuse Estuary, which stretches between stations 160 and 180, is the section where our existing information is most limited. Monitoring that part of the estuary should reduce our model and standard violation uncertainties. This section of the estuary was identified both based on total system entropy and the DO violation criterion. Moreover, given the priorities that the experts placed on each of the three criteria, the final design was heavily weighted by the total system entropy-based design.

On the other hand, the locations just before the Neuse Estuary bend (between stations 100 and 140) were associated with the lowest monitoring priority, which indicates that with the existing monitoring program our knowledge of chlorophyll and
DO dynamics is highly certain in that section of the estuary. This is supported by the work of Luettich et al. (Luettich et al. 2000a), which shows that this section of the estuary has very predictable vertical stratification dynamics. Moreover, this area seems to have both low predictive uncertainties along with low standard violations entropies.

Figure 36: The relative importance of monitoring at the 25 proposed locations is reflected in the height of the associated vertical bars. The bar heights range from 1 to 25. The southern shoreline section that stretches between stations 160 and 180 has the highest values. The ranking is based on the three criteria we considered namely, total system entropy, Chl-a standard violation entropy, and DO standard violation entropy.
5.4 Discussion

In this study we assessed the existing water quality monitoring network that is currently operated in the Neuse Estuary based on multiple design criteria. We also identified areas with maximum uncertainties and that would benefit from future monitoring. The developed methodology incorporated a MADM component that permitted different experts to weigh-in and give their perspective on the relative importance of different optimization criteria. Moreover, the adopted framework was probabilistic in nature, which allowed it to transparently account and propagate the uncertainties both in the data and in the model parameters. Additionally, the developed framework did not place any limitations on the spatio-temporal field (stationarity or isotropy); an advantage that is particularly important in modeling dynamic water bodies (estuaries, coastal waters, dammed rivers, current driven waters, etc).

Our results explored the possible difference between a probability and an entropy-based design. The results highlighted the advantages of adopting entropy as a monitoring design criterion. We think that the developed framework is a valuable tool for assisting scientists and managers in optimizing their monitoring efforts in impaired waters that are listed under section 303(d) of the Clean Water Act (Environmental Protection Agency 2001, National Research Council 2001).

The optimization of a monitoring network should also be coupled with a cost analysis. While our methodology did not explicitly account for cost, we assumed that the cost of monitoring at any locations was identical and represented an equal fraction.
of the total operational cost of the program. This assumption may not hold true from an operational stand-point, since the costs associated with the addition of a station may be non-linear; however, it is still adequate for our purposes. More elaborate cost functions can also be incorporated within our developed framework if the need arises.

Incorporating costs while accounting for uncertainty is an application of statistical decision theory (Berger 1985, Borsuk 2001). Another implicit assumption in our methodology was that information across the estuary is equally valuable. This might not be true in some cases, where certain areas may have a higher informational value. Such locations include shell-fishing areas, recreational beaches, and ecologically sensitive environments. Under such cases, an approach similar to the one adopted by Puangthongthub et al. (2007) can be incorporated within the modeling framework.
6. Remote Sensing of Chlorophyll Concentrations in the Pamlico Sound

6.1 Introduction

Coastal oceans comprise less than 10% of the total ocean area, but contribute around 20% of the total global primary production in the oceans (Knauer 1993, Schlesinger 2007). These highly productive zones are also home to a growing proportion of the world’s population, whereby nearly 75% of the human population is expected to live within 200 km from the coast by 2025 (Hinrichsen 1998). The human impact on the physical, chemical, and biological environment of these coastal areas is evident. One of the principal sources of coastal pollution is nutrient enrichment, a process that leads to eutrophication (Nixon 1995, Paerl et al. 2007, Boesch et al. 2009). Eutrophic waters become susceptible to fishkills, harmful algal blooms, changes in the water’s biogeochemistry, and the spread of dead zones (hypoxic) (Nixon 1995, Jørgensen and Richardson 1996, Bricker et al. 1999a, Borsuk et al. 2001b, Livingston 2001, Environmental Protection Agency 2007, Paerl et al. 2007, Scavia and Donnelly 2007, Boesch et al. 2009, Liu et al. 2010).

The natural variability in coastal primary production is large, necessitating extensive monitoring to gain a better understanding of these systems. Collected data can improve our understanding on how anthropogenic and physical forcings interact to alter these natural dynamics. Information gained from conventional monitoring programs often lack the required spatio-temporal resolution needed to adequately
assess, quantify, and eventually mitigate negative anthropogenic effects. The advancement of remote sensing has largely alleviated this limitation, with satellites providing data on a global scale at temporal and spatial resolutions that surpass existing in-situ monitoring efforts. This is accomplished through the use of bio-optical algorithms that are capable of mapping water leaving radiances \( L_w(\lambda) \) to geophysical properties (Werdell et al. 2007, Werdell et al. 2009).

The availability of data at a high spatio-temporal resolution is important for detecting high intensity short bursts of productivity, which are often missed by conventional monitoring efforts (Hyde et al. 2007). Additionally, remotely sensed data can help track the response of coastal systems to anthropogenic perturbations and extreme climatic events. This is made possible by a constant stream of data that is limited only by cloud cover. Nevertheless, our ability to use these satellites effectively to track and better understand the condition of coastal regions has not been fully realized.

The bio-optical complexities and the lack of adequate atmospheric correction algorithms in coastal regions have made the use of remotely sensed ocean color data challenging. Global chlorophyll algorithms have been optimized to predict chlorophyll-concentrations in open ocean conditions, which are often referred to as Case-I waters. Open oceans are considered to be a single component system, where phytoplankton is the main constituent (other than water) affecting the way the water body responds to radiation (Arst 2003, Komick et al. 2009). On the other hand, coastal waters and
estuaries are more optically complex and are considered multi-component systems. They are often referred to as Case-II waters.

Predicting chlorophyll concentrations in Case-II waters through NASA’s global Ocean Color (OC) algorithms has proven to be difficult, uncertain, and non-robust. A recent comparison of the relationship between in-situ chlorophyll data from Case-II waters and satellite based predictions has shown poor correspondence (Schalles 2006). The lack of fit in Case-II waters has been attributed to several factors including shallow depth, a dynamic and productive phytoplankton community, contamination from nearby land (e.g. aerosols, stray light, air pollution, runoff, etc.), and to the presence of other optically active components such as suspended sediments and colored dissolved organic materials (CDOM) (Bukata et al. 1995, Arst 2003, Kowalczuk et al. 2006, Gitelson et al. 2007, Werdell et al. 2009). This added complexity limits the validity of having a linear dependency between the optical properties of the water and the concentrations of chlorophyll (Slade et al. 2003).

In North Carolina, several attempts have been made to use remotely sensed data to predict chlorophyll dynamics in coastal waters. One of the earliest efforts focused on using Landsat data (Khorram and Cheshire 1985). This was followed by the work of Lunetta et al. (2009), who used the Advanced Visible-Infrared Imaging Spectrometer (AVIRIS) to study chlorophyll dynamics in the Albemarle-Pamlico Estuarine System (APES). Nevertheless, both studies were limited in their spatial and temporal coverage, a limitation primarily associated with the sensor choice. Recently, a more encompassing
approach was adopted by Sokolovsky et al. (2011), who conducted a validation assessment on remotely sensed chlorophyll concentrations in the Neuse Estuary and APES using the Medium Resolution Imaging Spectrometer (MERIS) sensor. Nevertheless, we know of no previous systematic attempt to validate and make use of the extensive chlorophyll time series data from the SeaWiFS satellite to predict chlorophyll levels in the PS.

In this chapter, we assess the performance of the existing SeaWiFS based global algorithms in predicting chlorophyll levels in the PS. We also test whether regionalized versions of these algorithms are capable of achieving a higher fidelity to the in-situ data. We then investigate the effectiveness of empirical algorithms that make use of the red and near-infrared optical ratios instead of the traditional Blue:Green ratio in predicting chlorophyll levels. Specifically, we assess the effectiveness of the Red:Green \( \frac{R_{\text{R}(670)}}{R_{\text{R}(554)}} \) and the Infrared:Red \( \frac{R_{\text{R}(765)}}{R_{\text{R}(670)}} \) optical ratios. A novel modeling approach, the Multivariate Adaptive Regression Splines (MARS), is also introduced and used to generate a hybrid algorithm that uses both the Blue:Green and Red:Green ratios simultaneously. Towards these goals, we make use of two independently collected datasets for the PS. Using the data, we calibrate, assess, and validate the ability of the different empirical ocean color models to predict chlorophyll concentrations in the study area.
6.2 Methods

6.2.1 Study Area and Data Acquisition

The PS in North Carolina (Figure 37) is the largest lagoon along the US East Coast. It is connected to the adjacent Albemarle Sound and together they form APES, which is the second largest estuarine complex in the US after the Chesapeake Bay (Paerl et al. 2007, Paerl et al. 2009b). The PS is separated from the Atlantic Ocean by a row of low lying sandy barrier islands that restrict water exchange with the Atlantic. This increases the system’s residence time, which is known to exceed one year (Pietrafesa et al. 1996). A relatively long residence time makes the system sensitive to nutrient enrichment (Paerl et al. 2010). On the other hand, the system is an important habitat for many ecologically and commercially important species. It supports more than 90% of North Carolina’s commercial and 60% of recreational finfish and shellfish catches (Copeland and Gray 1991, Paerl et al. 2001). Overall the system is shallow (mean depth = 4.5 m), periodically stratified, affected by wind-driven tides, and is prone to major perturbation from large scale Atlantic hurricane events (Paerl et al. 2001, Paerl et al. 2009a).

The PS is the point of discharge for several river systems that drain an area around 70,000 km$^2$. This makes the system susceptible to anthropogenic activities occurring in its contributing watersheds. Impairment in the PS is primarily attributed to non-point pollution sources (Environmental Protection Agency 2007). The system was designated by the USEPA in 1987 as “an estuary of national significance” within the National Estuary Program. In 2007, the USEPA National Coastal Assessment (NCA)
project rated the overall condition of the PS as “good to fair” using four indices of estuarine condition, with chlorophyll conditions receiving most concern (Environmental Protection Agency 2007).

Figure 37: The Pamlico Sound in North Carolina showing the chlorophyll sampling locations by the FerryMon program and the samples that are part of the SeaBASS database. Also shown on the map are the total unique locations of the FerryMon samples, of which only a few matched with SeaWiFS scenes.
6.2.1.1 In-Situ Data

In-situ chlorophyll measurements were collected from two water quality monitoring programs that operate in the PS. The first dataset included surface chlorophyll samples measured by the FerryMon program that operates a series of water quality sensors on three ferries that run within the study area. The program monitors water quality along established ferry routes (Table 11). An excess of 950,000 samples were available from Ferrymon between 2001 and 2007 (Figure 37). Chlorophyll levels in FerryMon are measured directly through onboard sensors before being transmitted wirelessly to a receiving station on the mainland (Buzzelli et al. 2003, Paerl et al. 2009a). Chlorophyll concentrations are estimated by converting fluorescence to chlorophyll concentrations. Fluorescence is measured through a YSI-MA 6600 fluorescence sensor that is capable of taking samples once every 3 minutes (~ 0.8 km between data points). The sensor makes use of the property of chlorophyll to fluoresce when excited with blue light (~470 nm wavelength). The ease of measuring chlorophyll levels with in-situ fluorescence however comes at the cost of reduced accuracy compared to conventional extractive chlorophyll measurement methods, including spectrophotometric analysis and high performance liquid chromatography (HPLC) (Buzzelli et al. 2003, YSI Inc. 2006). Details on the FerryMon program can be found in Buzzelli et al. (2003) and in Paerl et al. (2009a). Note that the FerryMon program is operated and maintained by the Paerl Lab at the University of North Carolina at Chapel Hill-Institute of Marine Science (UNC-IMS).
FerryMon data has recently been used to assess the capabilities of the MERIS satellite in predicting chlorophyll concentrations in the Neuse and APES (Sokoletsy et al. 2011).

Table 11: The ferry routes that are used to collect the FerryMon water quality monitoring data
Adapted from (Buzzelli et al. 2003)

<table>
<thead>
<tr>
<th>Route #</th>
<th>Ferry</th>
<th>Route</th>
<th>Average data points/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Floyd Lupton</td>
<td>Minnesota Beach to Cherry Branch</td>
<td>300</td>
</tr>
<tr>
<td>2</td>
<td>Carteret</td>
<td>Cedar Island to Ocracoke Inlet</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>Silver Lake</td>
<td>Ocracoke to Swan Quarter</td>
<td>200-300</td>
</tr>
</tbody>
</table>

Given the high sampling frequency of FerryMon, several in-situ measurements matched with a single concurrent satellite reading, a process that can bias the results. As such, we identified the in-situ measurements that had the smallest temporal difference with the satellite data for each scene. A 3 km x 3 km spatial window was then constructed around each of the selected in-situ measurement and the median of all in-situ chlorophyll-measurements within that window was then used for satellite matching. We also excluded in-situ measurements with a coefficient of variation (CV) greater than 0.15. This step ensures that we only retain in-situ data that are consistent over the space defined by the 3x 3 km window. Large variations in the CV may indicate high measurement errors and/or a ferry travelling through an algal front. This same criterion is often used with satellite data (Bailey and Werdell 2006).
The second in-situ monitoring database that we used came from the SeaWiFS Bio-optical Archive & Storage System (SeaBASS). This database is maintained by NASA for calibration purposes and for the future development of ocean color algorithms (Werdell and Bailey 2002, Werdell et al. 2003). A total of 1,312 chlorophyll-samples were available for the Pamlico (Figure 37). These samples were contributed by Dr. Rick Stumpf and spanned between 2001 and 2006. The SeaBASS chlorophyll data were all based on conventional fluorometric analysis. They were collected over a shallow depth profile (depths between 0 and 5 m), and as such depth-averaged concentrations were determined and used in this study to correspond with the satellite data. Note that we attempted to also use data from the ModMon program that operates in the Neuse Estuary; however, the coarse spatial resolution of SeaWiFS ruled out any chance for satellite match-up.

Collected in-situ data were geographically and temporally referenced in order to allow for easy comparison with SeaWiFS satellite images. The sampling sites for each of the three monitoring programs are shown in Figure 37. The combined datasets provided a wide range of diverse conditions (both spatially and temporally) that allowed us to effectively assess the performance of different empirical chlorophyll models. In addition as shown in Figure 38, the two datasets provided a relatively wide range of observed chlorophyll concentrations (between 0.1 and 744.0 mg/m$^3$). The interquartile range (IQR) of chlorophyll levels for the SeaBASS was between 2.0 and 16.2 mg/m$^3$; the IQR for the Ferrymon data ranged between 4.6 and 14.2 mg/m$^3$. However, following the
satellite matching (section 6.2.1.2) the range of the data decreased for both datasets (IQR for SeaBASS = 0.1 and 5.9 mg/m$^3$; IQR for FerryMon = 1.80 and 4.81 mg/m$^3$).

Figure 38: The distribution of chlorophyll concentrations for the combined data from FerryMon and SeaBASS. The top row shows the distributions for the entire dataset before satellite matchup. The second row shows the corresponding distributions following satellite matchup. The red vertical lines represent the mean chlorophyll value in each of the panels. Note the lognormal distribution for the observed chlorophyll concentrations in the Pamlico Sound.
6.2.1.2 Satellite Processing

Satellite images from the SeaWiFS sensor were downloaded from NASA’s Ocean Color website. The SeaWiFS satellite has been in operation since August 1, 1997. SeaWiFS collects data from 8 spectral bands that span the visible and infrared spectrum. Table 12 presents a summary of these 8 bands along with a brief description of the main uses associated with each band. The satellite images have a spatial resolution of ~ 1 km at nadir and the temporal recurrence is daily (Antonenko et al. 2001). Data from SeaWiFS are maintained, processed, and distributed by NASA Goddard’s Ocean Biology Processing Group (OBPG). Images were selected to overlap with the in-situ sampling time frame and the spatial extent of the study area. A total of 3,749 Level-1A image files were downloaded from NASA’s Ocean Color browser to cover the period from 2000 to 2007. Level-1A files, which hold the reflectance information, were then processed to Level-2 files that store the predicted chlorophyll concentrations based on a set of predefined NASA OC algorithms. This processing step was conducted through the use of the l2gen tool under the SeaDAS 6.1 software.
Table 12: SeaWiFS spectral channels along with their typical uses

<table>
<thead>
<tr>
<th>Band</th>
<th>Band width (nm)</th>
<th>Central wavelength (nm)</th>
<th>Color</th>
<th>Band usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>402-422</td>
<td>412</td>
<td>Blue</td>
<td>Yellow pigment/phytoplankton</td>
</tr>
<tr>
<td>2</td>
<td>433-453</td>
<td>443</td>
<td>Blue</td>
<td>Chlorophyll</td>
</tr>
<tr>
<td>3</td>
<td>480-500</td>
<td>490</td>
<td>Blue-green</td>
<td>Chlorophyll</td>
</tr>
<tr>
<td>4</td>
<td>500-520</td>
<td>510</td>
<td>Green</td>
<td>Chlorophyll</td>
</tr>
<tr>
<td>5</td>
<td>545-565</td>
<td>555</td>
<td>Red</td>
<td>Yellow pigment/phytoplankton</td>
</tr>
<tr>
<td>6</td>
<td>660-680</td>
<td>670</td>
<td>Red</td>
<td>Chlorophyll</td>
</tr>
<tr>
<td>7</td>
<td>745-785</td>
<td>765</td>
<td>Near-Infrared</td>
<td>Land-water contact, atmospheric correction, vegetation</td>
</tr>
<tr>
<td>8</td>
<td>845-885</td>
<td>865</td>
<td>Near-infrared</td>
<td>Land-water contact, atmospheric correction, vegetation</td>
</tr>
</tbody>
</table>

A process of in-situ to satellite matching was then conducted, whereby in-situ measurements were matched with concurrent satellite data. A match was defined when the overlap between in-situ and satellite data fell within a ± 3 hours temporal window and within a 3 x 3 pixel spatial window that was centered on the in-situ measurement. Typically used spatial windows range in size between 5 x 5 and 3 x 3 pixels (Harding et al. 2005, Werdell et al. 2009). We decided to adopt the latter so as to reduce the effects of possible patchiness in the chlorophyll levels in the PS. The adoption of a constrained spatio-temporal window for matching ensures a high level of confidence that both the in-situ and satellite data are describing the same ambient conditions (Werdell et al. 2009). Evidently in a highly variable system like the PS, one should expect that this
exclusion criterion will only minimize the effects of algal fronts and hydrodynamic forcings.

Following the spatio-temporal matching, we conducted a series of quality control steps for each of the matched data pairs. This allowed us to flag and remove irregular values in the satellite data. The standard operational pixel-masking scheme for SeaWiFS was implemented, yet similar to Werdell et al. (2009) we did not exclude pixels with stray light contamination. This allowed us to include data from the near-shore areas. Furthermore, we applied the exclusion criteria proposed by Bailey and Werdell (2006) for processing Level-1A and Level-2 files. These quality control criteria have since been accepted by researchers in the wider ocean color community for validating in-situ measurements with satellite data. Removing aberrant pixels ensures that only high quality satellite estimates are retained (Bailey and Werdell 2006, Werdell et al. 2009).

Following the process of quality control and spatio-temporal matching, we were left with about 280 matches, 245 of which were from the FerryMon program and the rest form the SeaBASS program. This signifies an ultimate matchup rate of around 0.026 % and 2.36 % for FerryMon and the SeaBASS, respectively. Such low matchup rates are expected, especially for the FerryMon data, as that program was not developed with the intention of maximizing matchups with the satellite orbiting schedule. Relaxing some of the adopted quality control measures or increasing the size of the adopted spatio-temporal windows can increase the number of retained matches, but at the same time it could degrade the quality of the data.
6.2.2 Empirical Ocean Color Algorithms

Empirical ocean color bio-optical algorithms attempt to map reflectance values measured by satellites to geophysical parameters of interest (O'Reilly et al. 1998, Slade et al. 2003). Empirical OC algorithms capitalize on the ability of the chlorophyll pigment to absorb light in the blue and red spectrum, as part of the photosynthetic process, and to reflect light in the green (O'Reilly et al. 1998). As such, a shift in reflectance towards green is expected as the concentration of phytoplankton in the water increases. Gordon and Morel (1983) recognized this property early on and used the shift from blue towards green to assess chlorophyll concentrations in the early 1980s. Their approach has since become the default method used to link reflectance to chlorophyll concentrations. Currently, existing global empirical algorithms are based on the Blue:Green optical ratio. Over the years many empirical algorithms have been developed. Readers interested in an in-depth survey on the developments in the OC empirical algorithms are encouraged to refer to O'Reilly et al. (1998, 2001), Patt et al. (2003), Gregg and Casey (2004), and to the International Ocean-Colour Coordinating Group report (2006).

6.2.2.1 Global Blue:Green Based Empirical Algorithms

The most commonly adopted empirical OC algorithms make use of variations in the Blue:Green ratios to predict chlorophyll levels. The default algorithm for SeaWiFS is currently the OC4S-V6, which is an updated version of the widely used OC4V4 (O'Reilly et al. 2001, Werdell et al. 2009 and references therein). The OC4 algorithms make use of
four optical bands to estimate chlorophyll levels. Three bands are in the blue range \( \left( R_{rs}(443), R_{rs}(490), R_{rs}(510) \right) \) and one is in the green \( (R_{rs}(555)) \). In addition to the default OC4S-V6 algorithm, NASA distributes chlorophyll products from two commonly used empirical algorithms namely the OC3S-V6 and OC2S-V6. The former uses two blue bands \( (R_{rs}(443), R_{rs}(490)) \) and one green \( (R_{rs}(555)) \); the latter only uses the longer frequency blue band of SeaWiFS \( (R_{rs}(490)) \) along with the green band \( (R_{rs}(555)) \).

Both OC4S-V6 and OC3S-V6 adopt band switching to compensate for saturation in the blue wavelengths as pigment concentrations increase. Thus, band switching occurs with the goal of determining the maximum band ratio (MBR) between the different SeaWiFS blue bands and the green band. The change from one reflectance ratio to another is directly associated with the concentrations of chlorophyll in the water, with higher reflectance wavelengths used for waters with higher chlorophyll concentrations. Switching towards higher blue wavelengths minimizes saturation.

All three global Blue:Green empirical algorithms predict chlorophyll concentrations from the Blue:Green ratio using a fourth order polynomial function. Table 13 summarizes these global algorithms and tabulates their globally calibrated coefficients. These global algorithms have proved to be very effective in Case-I waters, yet their performance in Case-II waters has been less promising. This lack in performance has been attributed both to signal contamination by aerosols as well as to the effect of CDOM and suspended solids that artificially lower the Blue:Green optical ratio. This often leads to overestimating chlorophyll levels in coastal waters (De Cauwer
It is worth noting that the stated accuracy goal for these algorithms has been set to 35% for chlorophyll concentrations in an operational range defined between 0.05 and 50 mg/m³.

**Table 13: SeaWiFS global Blue:Green based empirical algorithms.**
The $R_{rs}(\lambda)$ is the reflectance at the specified wavelength, $\lambda$.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC4S-V6 :</td>
<td>$R_{rs}(\text{blue}) = \text{Max}(R_{rs}(443), R_{rs}(490), R_{rs}(510))$</td>
</tr>
<tr>
<td>$Chla = 10^{c_0 + c_1 \times R_{45} + c_2 \times R_{55}^2 + c_3 \times R_{55}^3 + c_4 \times R_{55}^4}$</td>
<td>$R_{45} = \log_{10}\left(\frac{R_{rs}(\text{blue})}{R_{rs}(555)}\right)$</td>
</tr>
<tr>
<td>$c_0 = 0.3272$; $c_1 = -2.9940$; $c_2 = 2.7218$; $c_3 = -1.2259$; $c_4 = -0.5683$</td>
<td></td>
</tr>
</tbody>
</table>

| OC3S-V6 :   | $R_{rs}(\text{blue}) = \text{Max}(R_{rs}(443), R_{rs}(490))$ |
| $Chla = 10^{c_0 + c_1 \times R_{45} + c_2 \times R_{55}^2 + c_3 \times R_{55}^3 + c_4 \times R_{55}^4}$ | $R_{35} = \log_{10}\left(\frac{R_{rs}(\text{blue})}{R_{rs}(555)}\right)$ |
| $c_0 = 0.2511$; $c_1 = -2.3798$; $c_2 = 1.5823$; $c_3 = -0.6372$; $c_4 = -0.5692$ |                              |

| OC2S-V6 :   | $R_{25} = \log_{10}\left(\frac{R_{rs}(490)}{R_{rs}(555)}\right)$ |
| $Chla = 10^{c_0 + c_1 \times R_{25} + c_2 \times R_{25}^2 + c_3 \times R_{25}^3 + c_4 \times R_{25}^4}$ |                              |
| $c_0 = 0.2511$; $c_1 = -2.0853$; $c_2 = 1.5035$; $c_3 = -3.1747$; $c_4 = 0.3383$ |                              |

### 6.2.2.2 Regional Blue:Green Based Empirical Algorithms

The optical properties of Case-II waters are complicated by the presence of organic and inorganic particles that tend to interact with the spectral signature of chlorophyll as explained above. This added complexity limits the ability to directly adopt and use NASA’s standard global empirical products in Case II waters. (Bukata et al. 1991, Bukata et al. 1995, Darecki et al. 2003, Cota et al. 2004, De Cauwer 2004). This limitation was recognized early on by O’Reilly et al. (1998) in an ocean color chlorophyll algorithm
comparison study that followed the launch of SeaWiFS in 1997. The authors acknowledged that the use of global algorithms for Case-II waters will likely generate results that are less accurate than those generated by local or regionalized algorithms.

A myriad of regional radiometric SeaWiFS based chlorophyll algorithms has been developed in the past two decades across the globe. These include single band algorithms, band ratio algorithms, semi-empirical models, and neural networks. These algorithms tend to be site, season, and/or satellite specific (Arst 2003). Yet, the level of effort and reprocessing involved in developing these regional algorithms is high and sometimes prohibitive. Moreover, embarking on such an endeavor requires an abundance of high quality site specific in-situ chlorophyll data (Cota et al. 2004, Werdell et al. 2007).

In this study, we focused on assessing the performance of three previously regionalized Blue:Green algorithms, which were developed for the coastal waters of the Southeastern US (Table 14). The first algorithm is a regionalized version of OC3S that uses a log-linear representation of the Blue:Green MBR to predict chlorophyll levels. The second algorithm adds an additional quadratic term to the equation. Finally, the Southeast Coastal algorithm developed by Stumpf et al. (2000) is a regionalized and constrained version of the OC2S global algorithm.
Table 14: Regional Blue:Green based chlorophyll algorithms used with the SeaWiFS sensor

<table>
<thead>
<tr>
<th>Regional algorithms</th>
<th>Chesapeake Bay: OC3CB2 (Werdell et al. 2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Chl_a = 10^{c_0+c_1 \times R_{3S}}; )</td>
<td></td>
</tr>
<tr>
<td>( R_{R5}(\text{blue}) = \max \left( R_{R5}(443), R_{R5}(490) \right); )</td>
<td></td>
</tr>
<tr>
<td>( R_{3S} = \log_{10} \left( \frac{R_{R5}(\text{blue})}{R_{R5}(555)} \right); )</td>
<td></td>
</tr>
<tr>
<td>( c_0 = -0.115; c_1 = -3.678 )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chesapeake Bay: OC3CB (Werdell et al. 2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Chl_a = 10^{c_0+c_1 \times R_{3S}+c_2 \times R_{3S}^2} )</td>
</tr>
<tr>
<td>( R_{R5}(\text{blue}) = \max \left( R_{R5}(443), R_{R5}(490) \right); )</td>
</tr>
<tr>
<td>( R_{3S} = \log_{10} \left( \frac{R_{R5}(\text{blue})}{R_{R5}(555)} \right); )</td>
</tr>
<tr>
<td>( c_0 = -0.0208; c_1 = -3.1263; c_2 = 10.9575 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Southeast Coastal (Stumpf et al. 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Chl_{SE} = 10^{-2.5S \times R_{S}}; R_{S} = \log_{10} \left( \frac{R_{R5}(490)}{R_{R5}(555)} \right); )</td>
</tr>
</tbody>
</table>
| \( \begin{matrix} 
\text{Chl} & = & Chl_{SE} \\
\text{OC} & \text{if} & Chl_{SE} \geq 0.5, \\
\text{if} & 0.1 < Chl_{SE} < 0.5, & Chl = 10^{ \left( \frac{\log_{10}(Chl_{SE}) \times \log_{10}(Chl_{SE})}{\log_{10}(Chl_{SE}) + \log_{10}(Chl_{SE})} \right) + \frac{\log_{10}(Chl_{SE}) \times \log_{10}(Chl_{SE})}{\log_{10}(Chl_{SE}) + \log_{10}(Chl_{SE})}} \end{matrix} \)  |
| \( \text{OC2V6 (Refer to Table 13)} \)  |  

We also attempted to calibrate a regionalized OC4 algorithm specifically for the PS using the in-situ data. Regionalization was conducted through a regression-based least squares optimization. Re-parameterizing the OC algorithms based on in-situ data has been successfully conducted previously, most notably for the Chesapeake Bay (Werdell et al. 2007, Werdell et al. 2009) and for the arctic ocean (Cota et al. 2004).

6.2.2.3 Near-Infrared:Red Based Empirical Algorithm

In turbid productive waters there is a high reflectance in the near-infrared (NIR) spectrum, which permits the use of the red and NIR reflectance to detect the chlorophyll absorption peak at the 675 nm (Ruddick et al. 2001, De Cauwer 2004). Based on that property, an approach has been developed that uses changes in the NIR:Red
ratio \( \left( \frac{R_{rs}(765)}{R_{rs}(670)} \right) \) to predict chlorophyll concentration in highly productive waters. The use of the NIR:Red ratio also reduces the effects of absorption by CDOM and non-algal particles and limits particulate scattering, both of which can affect chlorophyll estimates generated from Blue:Green based algorithms (Gilerson et al. 2010).

Predictions based on the NIR:Red approach have shown encouraging results in highly productive complex coastal waters and inland lakes (Gin et al. 2002, Gitelson et al. 2007, Gitelson et al. 2008, Moses et al. 2009, Gilerson et al. 2010). With the NIR:Red based algorithms, one would expect that as chlorophyll concentrations increase, the reflectance ratio will increase concurrently. Equation 21 shows the functional form for the NIR:Red empirical model proposed by Gitelson et al. (2007) for use with SeaWiFS. It should be noted that Gitelson et al.’s (2007) model did not use a logarithmic transformation on chlorophyll concentrations. Nevertheless, We have opted to use log transformation given that chlorophyll levels are known to be log normally distributed (Ott 1995, Alameddine et al. 2010, Qian 2010). Failing to transform the chlorophyll data will violate one of the assumptions of linear regression models, namely that the residuals are assumed to be normally distributed with a constant variance.

\[
Log_{10}(Chla) = \alpha + \beta \times \frac{R_{rs}(765)}{R_{rs}(670)} + \varepsilon; \quad \varepsilon \sim N \left( 0, \sigma^2 \right) \quad \text{Equation 21}
\]
6.2.2.1 Red:Green Empirical Algorithm

Another approach that has shown promise in the Chesapeake Bay uses the reflectance ratio between the red and green bands \( \frac{R_{rs}(670)}{R_{rs}(555)} \) to estimate chlorophyll concentration in complex waters (Tzortziou et al. 2007). Tzortziou et al. (2007) used chlorophyll fluorescence at \(~\)670 nm to construct a two band SeaWiFS algorithm that predicts chlorophyll levels from variations in the Red:Green optical ratio. As expected with higher concentrations of chlorophyll more fluorescence will occur, making the relationship between the Red:Green ratio and chlorophyll levels positive. Equation 22 presents the functional form of the model based on the SeaWiFS bands (Tzortziou et al. 2007). Note that the positive relationship between chlorophyll and sea surface radianace at the red band (670 nm) was recognized as early as 1980 by Gordon and Clark (1981).

\[
\log_{10}(Chla) = \alpha + \beta \times \log_{10} \left( \frac{R_{rs}(670)}{R_{rs}(555)} \right) + \varepsilon; \ varepsilon \sim \text{Norm}(0, \sigma^2) \tag{Equation 22}
\]

6.2.2.1 MARS Model

MARS is an adaptive non-parametric regression method that allows for the incorporation of non-linearities and interactions between the data. It can be viewed as a generalization of both stepwise regression and recursive partitioning techniques (Breiman et al. 1984, Hastie et al. 2001). MARS adopts a modified recursive partitioning
method to generate a model from a set of additive functions over the space of the predictive variables. This results in overlapping instead of disjoint functions (Lewis and Stevens 1991). The resulting final model will thus be made up of a set of separate splines (also known as basis functions). Each basis is linear or piecewise-linear, with a knot. The introduction of hinge functions at knots (i.e. breakpoints) allows the model to deal efficiently with potential non-linear behaviors. MARS models are flexible and allow for variable interactions through multiplying two or more hinge functions.

MARS models are usually fit through a forward pass that introduces new hinge functions to reduce the sum of squared residual errors. This is followed by a backward pass that prunes the model by removing hinges or one side of the hinge based on calculating the generalized cross validation score (GCV) (Friedman 1991, Friedman and Roosen 1995, Stockton 1998). Typical MARS models are of the form presented in Equation 23. In this analysis we used the Earth package (Milborrow 2011) in R (Ihaka and Gentleman 1996, R Development Core Team 2010) to fit a MARS model to the reflectance data in our study area.
\[ y_i = f(x_i) + \varepsilon_i; \quad \varepsilon_i = N(0, \sigma^2) \]

\[ f(x_i) = \sum_{i=1}^{k} c_i B_i(x); \]

\[ B_i(x) = \begin{cases} 1; & i = 1 \\ \prod_{j=1}^{J_i} \left[ s_{j,i} \left( x_{w_i} - t_{ji} \right) \right]; & i > 1 \end{cases} \]

Equation 23

Where:

- \( c_i \) = are the model coefficients
- \( B_i(x) \) = are the set of possible basis functions
- \( J_i \) = is the degree of the basis interaction
- \([\cdot]_+\) = is the function \( \max(0, \cdot) \)
- \( s_{j,i} \) = is a sign indicator equal to \( \pm 1 \) and is used to select the correct side of the hinge
- \( w_{ji} \) = gives the index of the predictor variable being split (from 1 to \( p \))
- \( t_{ji} \) = are the knot points

MARS has been shown to be capable of reliably tracking complex data structures that may be obscured in the high-dimensional data space. MARS models have also been shown to outperform Neural Networks, which are “black boxes”, while retaining a functional form that can be easily presented and explained (Francis 2003). Additionally, the generated models have shown high predictive accuracies when used as a forecasting
regression tool (Steinberg et al. 1999). To our knowledge this is the first attempt to use MARS in developing an empirical ocean color algorithm for predicting chlorophyll levels in complex coastal waters.

6.2.2.2 Model Assessment

The different empirical OC models were assessed through a set of commonly used statistical metrics adopted by the wider ocean color community. These included the $R^2$, the median ratio (MR) of the predicted to observed chlorophyll concentrations, and the median absolute percent difference (APD) (Equation 24). The MR and APD were calculated for all models; the $R^2$ was only calculated for models that were locally calibrated to the PS data. The MR gives an idea about the model bias and can be used to assess over-prediction, while the APD and the $R^2$ assess model fit. An APD close to 0 and an $R^2$ and MR values close to 1 are desirable metrics. It should be noted that since chlorophyll follows a log normal distribution the mean and the median do not coincide. The mean is influenced by the heavy right tail of the log-normal distribution. As such, for all locally calibrated models we will be reporting the median rather than the mean predicted values, unless otherwise stated.

$$APD = \text{Median} \left( 100 \times \left| \frac{\text{Chl}_{sat} - \text{Chl}_{in situ}}{\text{Chl}_{sat}} \right| \right)$$  

Equation 24
6.3 Results

6.3.1 Blue:Green Based Algorithms

As expected, the global algorithms performed poorly when it came to predicting in-situ chlorophyll concentrations in the study area. It is clear from Figure 39 that the model predictions for the most part overestimated chlorophyll concentrations in the PS, especially towards the lower range of measured chlorophyll concentrations. The MD of predicted to observed chlorophyll concentrations was closer to 2 for the three global algorithms (Figure 39). This indicates that the predicted concentrations were roughly two times larger than their in-situ counterparts. This positive bias has been previously documented in the Chesapeake Bay and is a direct result of the presence of CDOM, which shares many optical properties with chlorophyll (Harding et al. 2005, Werdell et al. 2009). The median APDs were consistently above 90% for the three global algorithms. These findings are consistent with the previous work conducted in Massachusetts Bay (Hyde et al. 2007) and in the Chesapeake Bay (Harding et al. 2005, Signorini et al. 2005, Werdell et al. 2007), where the MD ratios were found to range between 1.24 and 1.97, and the APDs ranged between 31 and 118%. Additionally, we observed that the in-situ concentrations had a wider spread than that of the predicted concentrations, especially for the FerryMon data. This indicates that these global algorithms were incapable of capturing important elements of chlorophyll dynamics in the PS.
The regional Blue:Green algorithms outperformed their global counterparts. Their MD ratios ranged between 1 and 1.5 and the APDs were up to 50% smaller than those for the global algorithms. The Southeast Coastal (OC\textsubscript{coastal}) algorithm developed by Stumpf et al. (2000) showed the most promise with an MD of 1 and an APD of 41.6 %. Despite this, all three regional algorithms had limited capabilities of accurately predicting chlorophyll concentrations in the 1 to 5 mg/m\textsuperscript{3} range. Unfortunately, around 30 % of the in-situ data collected in the PS were in that range.
Figure 39: SeaWiFS global and regional Blue:Green based OC chlorophyll predictions versus in-situ data. Validation statistics are reported in the lower left corner of each panel. For reference the 1:1 line is drawn in each panel. APD is the median absolute percent differences and the MR is the median ratio.
The ordinary least squares PS regionalized Blue:Green OC4 algorithm showed some promising results; yet the improvements were not substantially different from the other regionalized Blue:Green based models (Figure 40). We first attempted to fit a fourth order polynomial model (OC4PS
\textsubscript{Poly}); the model results showed strong collinearities between the four predictor terms. This is to be expected given that polynomial regressions are prone to overfitting and suffer from collinearity. Only the coefficients on the first and second degree power terms were found to be statistically significant at the 0.1 level. Nevertheless, all five model coefficients were retained in order to keep the functional form of the OC4S-V6. The parameterization of the OC4PS
\textsubscript{Poly} algorithm is summarized in Table 14. The residual standard error of the model was found to be 0.34 in \text{Log}_{10} units. The $R^2$ was 0.18.

Another version of the model (OC4PS
\textsubscript{Lin}), which only uses a linear representation of the Blue:Green ratio instead of the polynomial equation, was also assessed. The model coefficients for that model are also summarized in Table 14. For the OC4PS
\textsubscript{Lin} the model coefficient on the blue:green ratio was significant at the 0.05 level. The model results for both the OC4PS
\textsubscript{Poly} and the OC4PS
\textsubscript{Lin} models are shown in Figure 40. As can be seen, the polynomial model did not result in any significant improvements over the simpler linear model.
Figure 40: (a) The relationship between the MBR Blue:Green ratio and the OC4PS\textsubscript{Poly} predicted chlorophyll levels; the red line represents the median predicted concentrations from model and the dashed lines represent the 95% confidence intervals; also shown is the model’s $R^2$ value. (b) The median OC4PS\textsubscript{Poly} predicted chlorophyll concentrations are plotted against the in-situ data; validation statistics are reported. (c) The relationship between the MBR Blue:Green ratio and the OC4PS\textsubscript{Poly} predicted chlorophyll levels; the red line represents the median predicted concentrations from the OC4PC\textsubscript{Lin} model and the dashed lines represent the 95% confidence intervals; also shown is the model’s $R^2$. (d) The median OC4PS\textsubscript{Lin} predicted chlorophyll concentrations are plotted against the in-situ data; validation statistics are also reported. APD is the median absolute percent differences and the MR is the median ratio. For reference the 1:1 line is also drawn in panels (b) and (d).
Table 15: Ordinary least squared regression based regionalization of the OC4 algorithm to the Pamlico Sound

<table>
<thead>
<tr>
<th>OC4PS (Poly):</th>
<th>( R_{rs}(\text{blue}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log_{10}(Chla) = c_0 + c_1 \times R_{4S} + c_2 \times R_{4S}^2 + c_3 \times R_{4S}^3 + c_4 \times R_{4S}^4 + \epsilon )</td>
<td>( = \max(R_{rs}(443), R_{rs}(490), R_{rs}(510)) )</td>
</tr>
<tr>
<td>( \epsilon \sim \text{Norm}(0, \sigma^2) )</td>
<td>( R_{4S} = \log_{10}\left(\frac{R_{rs}(\text{blue})}{R_{rs}(555)}\right) )</td>
</tr>
<tr>
<td>( c_0 = -0.35 \pm 0.26 )</td>
<td></td>
</tr>
<tr>
<td>( c_1 = -5.78 \pm 3.27 )</td>
<td></td>
</tr>
<tr>
<td>( c_2 = 17.49 \pm 8.89 )</td>
<td></td>
</tr>
<tr>
<td>( c_3 = -133.94 \pm 157.61 )</td>
<td></td>
</tr>
<tr>
<td>( c_4 = 277.4 \pm 376.02 )</td>
<td></td>
</tr>
<tr>
<td>( R^2 = 0.18 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma = 0.34 )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OC4PS (Lin):</th>
<th>( R_{rs}(\text{blue}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log_{10}(Chla) = c_0 + c_1 \times R_{4S} + \epsilon )</td>
<td>( = \max(R_{rs}(443), R_{rs}(490), R_{rs}(510)) )</td>
</tr>
<tr>
<td>( c_0 = -0.092 \pm 0.08 )</td>
<td>( R_{4S} = \log_{10}\left(\frac{R_{rs}(\text{blue})}{R_{rs}(555)}\right) )</td>
</tr>
<tr>
<td>( c_1 = -4.27 \pm 0.58 )</td>
<td></td>
</tr>
<tr>
<td>( R^2 = 0.16 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma = 0.34 )</td>
<td></td>
</tr>
</tbody>
</table>

6.3.1 Infrared:Red Based Algorithm

We initially expected that the 2-band SeaWiFS NIR:Red \( \left(\frac{R_{rs}(765)}{R_{rs}(670)}\right) \) algorithm would perform satisfactorily in the study area; yet its performance was not consistent across the FerryMon and SeaBASS datasets. The model showed a positive relationship with the SeaBASS data with an \( R^2 \) of 0.42, yet that strong relationship was lost when the model was optimized using the observed measurements from FerryMon (Figure 41). This indicates that the relationship between chlorophyll and the NIR:Red ratio was possibly not robust. When optimized with the combined data, the model showed a poor fit \( (R^2 = 0.15) \), although both the intercept and slope were significant at the 0.05 level (Equation 25).
\[ \log_{10}(Chl) = -0.79 + 5.65 \times \left( \frac{R_{rr(765)}}{R_{rr(670)}} \right), \varepsilon \sim \text{Norm}(0, 0.35^2) \]

Equation 25

The lack of model fit can be explained by the relatively wide range of chlorophyll concentrations observed in the study area (between 0.1 and 61 mg/m\(^3\)), with the majority of the in-situ measurements falling below 10 mg/m\(^3\) (Figure 38). Chlorophyll concentrations below 10 mg/m\(^3\) tend to have dampened absorption at 700 nm. This can affect the relationship between chlorophyll levels and the NIR:Red ratio, which makes the model potentially unreliable for estimating low to moderate chlorophyll concentrations (Gilerson et al. 2010). At these low chlorophyll levels, backscattering and absorption by CDOM and inorganic particles may also obscure chlorophyll absorption in the red region of the spectrum \((R_{rr(670)})\) (Dall'Olmo et al. 2005, Gitelson et al. 2008).

Previous work with the NIR:Red algorithms in lakes and reservoirs with chlorophyll concentrations below 5 mg/m\(^3\) have shown that the model predicted values were up to six times larger than in-situ measurements (Gitelson et al. 2008). In our case, the model did not show a bias towards over prediction \((MR \approx 1)\). Another possible limitation hampering the NIR:Red ratio model is the absence of an operational SeaWiFS atmospheric correction algorithm that can be satisfactorily used with the red and NIR bands (Gitelson et al. 2008).
6.3.2 Red:Green Based Algorithm

A regionalized linear regression model was constructed to predict chlorophyll levels from the Red:Green \( \left( \frac{R_{(670)}}{R_{(554)}} \right) \) ratio in the PS (Equation 26). Both model coefficients were found to be significant at the 0.05 level. The Red:Green based chlorophyll model was also found to be robust. The values on the model coefficients did not change appreciably when the model was fit separately with the FerryMon and the SeaBASS datasets (Figure 42). The model had an APD of 41.6\% and an MR of 0.9. Both are comparable to the results of the other Blue:Green regionalized models. Yet, the model unlike the Blue:Green based models behaved well across the entire observed range of Red:Green ratios (Figure 42 and Figure 43).
Equation 26

\[ \log_{10}(Chl) = 0.94 + 1.50 \times \log_{10} \left( \frac{R_{rs}(670)}{R_{rs}(554)} \right), \epsilon \sim \text{Norm}(0, 0.32^2) \]

**Figure 42**: The Red:Green model based chlorophyll concentrations plotted against the SeaBASS data (left panel), and the FerryMon data (right panel). The red lines represent the median predicted chlorophyll values as a function of the Red:Green ratio. The dashed lines show the 95% confidence intervals.
Figure 43: The Red:Green chlorophyll model results for the combined data. In the left panel the red line shows the median predicted chlorophyll values as a function of the Red:Green ratio. The dashed lines show the 95% confidence intervals. In the right panel, the median predicted chlorophyll concentrations are plotted against the in-situ data. Validation statistics are also reported. APD is the median absolute percent differences and the MR is the median ratio. For reference the 1:1 line is also drawn in the panel to the right.

The coefficient of determination for the model was around 0.25, which is significantly better than that for the OC4PS models and the NIR:Red based model. This improvement is supported by the findings of Tzortziou et al. (2007) in the Chesapeake Bay and the results reported by Lunetta et al. (2009) for APES. The positive relationship between chlorophyll levels and the Red:Green ratio might at first be a bit surprising, as one might expect that chlorophyll absorption in the red band (670 ± 10 nm) would increase with increased chlorophyll concentrations. Yet, the positive relationship between the Red:Green ratio and chlorophyll concentrations can be attributed to the
effect of chlorophyll fluorescence, which has a Gaussian spectral profile centered at 685 nm with a width at half maximum of 25 nm (Mobley 1994, Gower et al. 2004). Fluorescence seems to overwhelm any chlorophyll absorption at the red wavelength. Moreover, Gitelson et al. (2000) have previously shown that light scattering by the phytoplankton cell walls offsets any red absorption.

In an effort to ascertain that the positive relationship between the Red:Green ratio \( \frac{R_{670}}{R_{554}} \) and chlorophyll was primarily due to chlorophyll fluorescence, we compared the satellite generated Red:Green ratios to the fluorescence data measured by the YSI-MA 6600 sensor onboard the three ferries in FerryMon. The relationship was positive and all model coefficients were significant at the 0.05 level (Figure 44 and Table 16). Note that by excluding the potential outliers in the fluorescence data (fluoresce > 4 %FS), the observed positive relationship between the Red:Green ratio and fluorescence is even stronger and more significant (Table 16). A more formal outlier detection approach, similar to the one outlined by Alameddine et al. (2010), may be required to better assess and identify outliers in the dataset.
Figure 44: Relationship between in-situ fluorescence measurements collected from the FerryMon program and the Red:Green ratio from SeaWiFS (with the potential outliers)

Table 16: The linear regression model describing the relationship between the in situ fluorescence measurements from the FerryMon program and the Red:Green ratio from SeaWiFS

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All data</td>
<td>$\sqrt{\text{Fluorescence (FS)}} = 1.3 + 1.3 \times \left( \frac{R_{rr}(670)}{R_{rr}(554)} \right) + \varepsilon$</td>
</tr>
<tr>
<td></td>
<td>$\varepsilon \sim \text{Norm}(0, 0.36^2)$</td>
</tr>
<tr>
<td>Without outliers</td>
<td>$\sqrt{\text{Fluorescence (FS)}} = 1.2 + 1.2 \times \left( \frac{R_{rr}(670)}{R_{rr}(554)} \right) + \varepsilon$</td>
</tr>
<tr>
<td>(Outliers with fluorescence &gt; 4 %FS)</td>
<td>$\varepsilon \sim \text{Norm}(0, 0.22^2)$</td>
</tr>
</tbody>
</table>
6.3.3 The MARS Based Hybrid Algorithm

The adopted MARS modeling framework generated a hybrid algorithm that uses both the $\log_{10}(Blue: Green)$ and $\log_{10}(Red: Green)$ ratios for predicting chlorophyll levels in the PS (Equation 27). The hybrid model captured the negative relationship linking the Blue:Green ratio to chlorophyll concentrations (Figure 45). This relationship is consistent with our understanding of higher absorbance in the blue spectrum as chlorophyll levels increase. However, the results indicated that the relationship in the Blue:Green range is non-linear over the entire range of the ratio. It is piecewise linear with a hinge introduced at $\log_{10}(Blue: Green) = -0.103$ (Equation 27). Past the identified hinge (breakpoint), the Blue:Green ratio loses its predictive power with respect to chlorophyll concentrations. This loss in predictive power is possibly due to the increased contribution of CDOM absorption in the blue spectrum at these low levels of chlorophyll. Under these conditions, chlorophyll is no longer the most prominent optical component absorbing in the blue range. Note that the introduction of the hinge also counteracts the high leverage exhibited by two potential outliers.

On the other hand, the positive linear relationship that we previously observed between chlorophyll concentrations and the Red:Green ratio (section 6.3.2) is also found to be significant in the MARS hybrid model (Figure 45). That relationship is continuous over the entire Red:Green range, as there was no statistical support for the introduction of a hinge. Together, the two ratios provide a means of predicting chlorophyll concentrations in the PS under a variety of optical conditions.
Log$_{10}$(Chl) = 0.75 + 1.27 \times \text{Log}_{10}\left(\frac{R_{rr(670)}}{R_{rr(554)}}\right) + 4.0 \times \left(-0.103 - \text{Log}_{10}\left(\frac{\text{Blue}}{\text{Green}}\right)\right) + \varepsilon \quad \text{Equation 27}

\varepsilon \sim \text{Norm}(0, 0.31^2)

Figure 45: The results from the MARS generated hybrid model. In the left panel, the line shows the predicted median chlorophyll concentrations as a function of the Blue:Green ratio while holding the Log$_{10}$(Red:Green) at its median value. In the right panel, the line shows the predicted median chlorophyll concentrations as a function of the Red:Green ratio while holding the Log$_{10}$(Blue:Green) ratio at its median value. In both panels rugs are added above the x-axis in order to show the distribution of the data with respect to the predictor.

The MARS chlorophyll model surpassed the other models in terms of model fit and showed reliability for operating under the diverse in-situ conditions. Yet, there still remains a significant amount of unexplained variability ($R^2 = 0.32$) in the observed in-situ chlorophyll concentrations. The APD and the MR for the MARS based model were also encouraging (Figure 46). The MR was close to 1 indicating no systematic over-
prediction and the APD was around 40%, which is similar in magnitude to the APDs of the other regionalized OC empirical models.

In order to evaluate whether the model suffer from overfitting, we conducted a four-fold cross validation. Cross validation proceeded by fitting the model with 75% of the data and then measuring its performance on the remaining 25%. Performance was measured by calculating the mean of the $R^2$ from the four cross validations. The mean cross validated $R^2$ was 0.30, which was very close to the model’s $R^2$ of 0.32. Similarly, the GCV statistic was calculated to be 0.09. The GCV is the average-squared residual of the model fit divided by a penalty term that accounts for increasing model complexity (Friedman 1991). As such, simple models that fit the data well have low GCV (close to zero). Together, the GCV statistic and the cross validation results indicate that the model does not show any signs of overfitting. Reassuringly, the model structure remained largely unchanged when data from SeaBASS and FerryMon were used separately to parameterize the model. This gives us confidence that the generated MARS hybrid model is robust.
We also explored fitting a MARS model that allows for the predictors to interact. The resulting model showed marginal improvement in terms of model fit, yet this improvement and added model complexity came at the expense of the model’s predictive power. This is a clear sign of model overfitting. Cross validation tests also showed a deterioration in the cross validation metrics. The inclusion of other optical ratios was also attempted, but the resulting models did not show promise.
6.4 Discussion

6.4.1 Model Comparison

The default global algorithms were found to be inferior to regional Blue:Green ratio models as well as to models based on the Red:Green ratio. The performance of the NIR:Red based algorithm was not satisfactory in the case of the PS. This may be due to the relatively low in-situ chlorophyll concentrations in the study area. The NIR:Red based algorithms are known to perform well in hyper-productive waters, which are more typical in inland lakes and some highly productive coastal waters (Dall’Olmo et al. 2005, Dall'Olmo 2005, Gitelson et al. 2008).

Both the global Blue:Green and the NIR:Red based OC algorithms showed a more constrained dynamic range as compared to the in-situ data, particularly in the chlorophyll range from 2 to 10 mg/m$^3$. This could be attributed to a more pronounced effect of CDOMs on reflectance in that range.

The Red:Green optical ratio algorithm, showed the most promise in the study area. This improvement was achieved by linking the observed chlorophyll levels to fluorescence in the red band ($R_{rs}(670 \text{ nm})$) rather than to absorbance in the blue bands ($R_{rs}(443), R_{rs}(490), R_{rs}(510)$). Given that peak fluorescence occurs at 677 nm, we expect that the use of MODIS-Aqua’s 667 nm red band will result in further noise reduction and a better fit. This improvement was reported for the Chesapeake Bay (Tzortziou et al. 2007).
On the other hand, the MARS-based hybrid algorithm, which uses both the Blue:Green and the Red:Green optical ratios, was found to be the most promising model for the PS. It showed robustness, limited bias, operability over a wide range of chlorophyll and optical band ratios, and a reasonable correspondence with in-situ data. Yet even with the MARS hybrid model, there remains a significant amount of unexplained variation in the measured chlorophyll levels in the Sound.

6.4.2 Model Fit

The relatively large variance associated with the satellite-predicted chlorophyll values in the PS is to be expected given the measurement uncertainties (both in-situ and satellite), the temporal lag, and the limitations in the existing atmospheric correction algorithms. While higher model performance results have been reported in other studies, those were largely achieved either by constraining the calibration over a limited time period (which in turn limits variability) and/or through concurrently profiling radiometric measurements with shipboard hyperspectral radiometers that were used alongside in-situ chlorophyll sampling. Using shipboard radiance reflectance measurements instead of satellite based measurements isolates the negative effects that aerosols and the atmosphere have on light reflectance. This increases the signal-to-noise ratio and allows for better fit. Additionally, using radiometric data from radiometers bypass the Level-2 spatio-temporal match-up process, as sampling is done concurrently. In this study, we have chosen to use actual satellite data rather than in-
situ radiometric measurements and to work over an extensive temporal period (over 7 years) that covers a wide variety of environmental conditions. We also used two chlorophyll datasets that implemented different in-situ chlorophyll sampling/measurement methods. We believe that our approach like that of Werdell et al. (2009) and Sokoletsky et al. (2011) provides a more realistic assessment of the adequacy and limitations associated with the use of SeaWiFS’s rich time series data on chlorophyll dynamics in the coastal waters.

6.4.3 Limitations and Recommendations

The spatial distribution of the in-situ data was not homogenous over the entire PS as seen in Figure 37. This may limit the accuracy of the model predictions outside of the spatial extent used for model calibration. Similarly, the temporal distribution of the matched data was not uniformly distributed across seasons, with more match-ups occurring in the winter (30% of total) as compared to the summer (19% of total). This is largely due to the disproportionate effect of cloud cover over the year. In addition, the satellite data exclusion criteria we adopted (Bailey and Werdell 2006) biased the in-situ data towards lower observed chlorophyll concentrations (Figure 38). This unintended bias may result in poor performance when it comes to predicting elevated chlorophyll concentrations.

Currently, there are many expectations that the use of the MERIS satellite will result in major improvements to the accuracy of satellite-based chlorophyll prediction in
coastal waters. These expectations are attributed to MERIS’s high spectral (15 spectral bands in the 390-1040 nm range) and spatial (300 m) resolutions. Sokoletsky et al. (2011) have recently used MERIS to predict chlorophyll levels in APES. Their locally calibrated power-law based model achieved an $R^2$ of 0.45. These results indicate that the use of MERIS in coastal waters may not bring about the major improvements that were initially expected. One of the main sources of uncertainty in coastal areas remains the lack of adequate atmospheric correction algorithms. Research aimed towards developing more efficient atmospheric correction models will surely result in major improvements to satellite-based chlorophyll predictions in complex waters. There is reason to be encouraged as new atmospheric correction models, such as that proposed by Wang and Shi (2007), are adopted and implemented by NASA. Nevertheless, a recent field assessment study conducted by Werdell et al. (2010) in the Chesapeake Bay showed that Wang and Shi’s atmospheric correction method did not yield the anticipated improvements.

Given these limitations, the role of in-situ modeling will remain an important source of information in optically complex waters for the foreseeable future. Yet, integrating in-situ data with satellite-based estimates is likely to become more important. Moreover, several models from different satellites can be integrated to generate blended chlorophyll product. One way to accomplish this integration is through the use of Bayesian model averaging. This approach combines the results from several plausible models/satellites to generate model-weighted chlorophyll predictions.
This is done by weighing the contribution of each model by its associated posterior model probability (Gibbons et al. 2008). This process of model averaging acknowledges that there is no one “true” model nor one preferred satellite. The use of a Bayesian approach to predict chlorophyll concentrations from satellite imagery is attractive given the flexibility it provides when it comes to the effortless integration of new data. This is an area that we are interested in exploring more closely in the near future. For that, we will be expanding our work to encompass data from the Modis-Aqua and the MERIS sensors.
7. Conclusions

7.1 Overview of Results

This dissertation presented an integrated approach towards better understanding the effectiveness of the Neuse nitrogen TMDL program in reducing nitrogen delivery to the Neuse Estuary and its impacts on chlorophyll dynamics in the Neuse Estuary. This assessment was achieved through the development of multiple novel water quality models that operate at different spatio-temporal scales.

In Chapter 2, I developed a Bayesian temporally dynamic SPARROW model to assess nitrogen delivery in the Neuse River basin between 1990 and 2001. The developed modeling framework adopted a novel regionalization approach that nested the model within a regional SPARROW model and used Bayesian updating over time to allow SPARROW to predict nitrogen loads on an annual basis. The model results indicated an overall drop in the annual TN loading rates between 1990 and 2001 in most sub-basins in the Neuse. Another key result of this work was the ability to quantify model parameter correlations, assess model equifinality, as well as measure the value of additional information resulting from additional monitoring activities.

In chapter 3, I developed a new Bayesian flow-concentration model that allows for greater flexibility when it comes to characterizing the relationship between measured river flows and pollutant concentrations. This was achieved through the incorporation of a temporal changepoint along with a flow threshold within the model.
structure. The results indicated that there is strong evidence to suggest that TON loading to the Neuse Estuary changed in the year 1999, as a result of major successes in reducing point source TON loadings in the basin. The results also showed that post 1999 TON levels reaching the Neuse Estuary were reduced on average by 32% as compared to pre-1999 load estimates. Yet, the model results highlighted that load reductions were a function of the observed flow regime in the Neuse basin, with most reductions occurring at low flow conditions. These findings indicate that future management and mitigation measures aiming at reducing nitrogen inputs should target non-point source pollution sources in the basin.

In chapter 4, I assessed chlorophyll dynamics in the UNRE post 2000 through the development of a BN model that was constructed using automated constraint-based structure learning algorithms. The model results showed that chlorophyll dynamics were largely governed by the physical environment (river flows and water temperature). The link between nitrogen species and chlorophyll levels was weak and not supported by the available data. One of the reasons for the lack of a connection between nitrogen and chlorophyll levels in the UNRE was attributed to the overall temporal stability of the nitrogen concentrations in that section of the estuary. Moreover, the results showed the importance of using a hybrid structure learning approach in building BN model topology. This work also provided a set of recommendations for using BN in environmental/ecological applications.
In Chapters 5 and 6, I focused on assessing the existing monitoring programs in the Neuse Estuary and Pamlico Sound. I developed a Bayesian Hierarchical spatio-temporal model for the Neuse Estuary that incorporated three entropy-based optimization criteria, which accounted for model uncertainty as well as DO and chlorophyll standard violations. The final optimized monitoring solution was generated by weighting the three decision criteria based on expert elicitation and the use of a multiplicative AHP approach. The results indicated that the southeastern shoreline of the Neuse Estuary would benefit the most from additional water quality monitoring. In Chapter 6, I focused on assessing the capability of using the SeaWiFS remote sensor to complement existing in-situ chlorophyll monitoring programs in the Neuse Estuary and Pamlico Sound. The results indicated that the coarse spatial resolution of the sensor made it unusable for the Neuse Estuary; but the sensor did provide useful information on chlorophyll levels in the Pamlico Sound. The results from the Pamlico Sound indicated that typically used Blue:Green based algorithms performed poorly in the study area as compared to a Red:Green based algorithm. As such, I developed a MARS based hybrid algorithm that made use of the Red:Green ratio to capture chlorophyll dynamics in the sound. The developed algorithm links fluorescence levels to chlorophyll levels.

### 7.2 Future Directions

In the future, I intend to expand the Bayesian temporally dynamic SPARROW model that I developed (Chapter 2) to a fully multilevel/hierarchical Bayesian model.
This would allow for better understanding regional (and/or basin) differences in pollutant export. The hierarchical model will also permit for load predictions in basins/regions with few monitoring stations by borrowing strength from the rest of the regions. In addition, I am interested in incorporating the concept of entropy within the Bayesian SPARROW framework in order to better identify locations within the river basin that will benefit from extra monitoring efforts. Moreover, I am currently exploring the development of a Chinese application for the Bayesian SPARROW model.

I also plan to build upon the flow-concentration model that I developed for assessing changes in TON loading to the Neuse estuary (Chapter 3). I will use the developed model framework to assess TN load reductions in the river basin in light of the recent work by Lebo et al. (In Press), who reported that organic nitrogen sources have increased over time in the Neuse thus diminishing the achieved nitrogen reductions in terms of inorganic nitrogen. Moreover, I am interested in assessing the sensitivity of different nitrogen reduction estimates with respect to the adopted load estimation methods. This assessment will help resolve some of the uncertainties associated with TMDL evaluation.

I also plan to expand the developed BN for the UNRE to cover the entire Neuse Estuary. This will be achieved through the adoption of an object oriented dynamic BN model that efficiently links the different segments of the estuary. Furthermore, I plan to further develop on the chlorophyll remote sensing work I did in Chapter 6 by including radiance measurements from the MODIS-Aqua satellite. The integration of both MODIS
and SeaWiFS data will generate a hybrid product that is expected to be more robust and less constrained by cloud cover.

Finally, I am interested in interlinking the different modeling components that were presented in chapters 2 through 6 into a coupled system. This integration will improve information propagation, allow for transparent uncertainty quantification, and ultimately lead to informed decisions through the adoption of adaptive management. Moreover, the coupled system will improve the capability to update and improve model forecasts over time as new data become available, which ultimately will result in better management decisions that aim at achieving compliance with water quality standards. For this model integration to be successful, I will be exploring efficient approaches for statistical downscaling (both spatial and temporal) that aim to bridge the gap between observations and model predictions made at different spatio-temporal scales.
Appendix A: Neuse Dynamic Bayesian SPARROW (WinBUGS Code)

model{

    mm[1]<-0

    for (k in 1:2){ #k is the number of years in each slice
        for (i in 1:stat.numb){
            logload.matrix.sim[i,k] ~ dnorm(MM[(k-1)*N+i], tau);
            MM[(k-1)*N+i] <- log(mm[(k-1)*N+i+1]);
            mm[(k-1)*N+i+1]<-sum(BS[(k-1)*N+i, 1:J[i]]);#This is for station i since
            mm[1] is zero ==> mm should have a length that is +1 of the length of the
            # of stations
            for (j in 1:J[i]){  
                decay1[(k-1)*N+i,j] <- exp(-ks.b*time.travel.b[(k-1)*N+i,j]
                -ks.m*time.travel.m[(k-1)*N+i,j]-ks.u*time.travel.u[(k-1)*N+i,j]));
                decay2[(k-1)*N+i,j] <- pow((1+kr*qload[(k-1)*N+i,j]),-1);
                deliver[(k-1)*N+i,j] <- exp(-alpha*soil[(k-1)*N+i,j]);
                new[(k-1)*N+i,j] <- beta.point*point[(k-1)*N+i,j] + 
                (beta.1*ag[(k-1)*N+i,j]+ beta.2*non.ag.source[(k-1)*N+i,j])*deliver[(k-1)*N+i,j];
            }
        }
    }
}
BS[(k-1)*N+i,j] <- (mm[upland.1[(k-1)*N+i,j]] + mm[upland.2[(k-1)*N+i,j]]
+ mm[upland.3[(k-1)*N+i,j]] + new[(k-1)*N+i,j])*decay1[(k-
1)*N+i,j]*decay2[(k-1)*N+i,j];

}
}

##########################################################################
## priors from previous SPARROW run ##
##########################################################################
ks.b ~ dnorm(ks.b.mu, ks.b.tau)[0,);
ks.m ~ dnorm(ks.m.mu, ks.m.tau)[0,);
ks.u ~ dnorm(ks.u.mu, ks.u.tau)[0,);
kr ~ dnorm(kr.mu, kr.tau)[0,);
beta.point ~ dnorm(point.mu, point.tau)[0,);
tau ~ dgamma(tau.r,tau.mu);
alpha~dnorm(alpha.mu,alpha.tau)[0,);
beta.1~dnorm(beta.1.mu, beta.1.tau)[0,];#Here beta.1 is for Atm+Fert+Manure
beta.2~dnorm(beta.2.mu, beta.2.tau)[0,);

}
Appendix B: Flow-Concentration Model (WinBUGS Code)

model{
    # n= number of observations = 2430
    for (i in 1:n){
        y[i]~dnorm(y.hat[i],tau.y[J[i],K[i]])  # The regression model
        y.hat[i]<-beta.0[J[i]]+beta.T*tempc[i] +step(year[i]-cp1)*(beta.1[J[i]]
        +delta*step(x[i]-x.change[i]))* (x[i]-x.change[i])+
        (1-step(year[i]-cp1))beta.1[J[i]]*x[i]
        x.change[i]<- (1-pp)*low+pp*hi  #The flow threshold
        J[i]<- step(year[i]-cp1)+1  # J = An indicator for pre and
        # post changepoint  
        # (either 1 or 2)
        K[i]<- step(x[i]-x.change[i])+1  # K = An indicator for pre and
        # post flow threshold  
        # (either 1 or 2)
    }
    delta~dnorm(0.0,1.0E-3)  #The prior on delta. Delta is
    #the change in the slope on log
    #flow after flow threshold for
    #the post-1999 submodel
    pp~dbeta(1,1)  #The prior on the location of
    #the flow threshold  
    # (between 0 and 1)
    beta.T~dnorm(0.0,1.0E-3)  #The prior on the coefficient on
    #water temperature
    for (p in 1:2){  
        #p=2 is the number sub-
        #models
        beta.0[p]~dnorm(0.0,1.0E-3)  #The priors on the model
        #intercepts. The intercept is
        #the log conc at the flow
        #threshold for the post-1999
        #sub-model, while it is the
        #regular intercept(i.e at
        #log(flow)=0) for the pre-1999
        #submodel
        beta.1[p]~dnorm(0.0,1.0E-3)  #The priors on the model
        #slopes on log flow for before
        #flow threshold for the post-
### 1999 submodel and the regular slope on log flow for the pre-1999 submodel

```r
tau.y[1,1] <- pow(sigma.y[1], -2)  # The precision for the pre-1999 model
sigma.y[1] ~ dunif(0, 4)  # The prior on the standard deviation of model residuals

tau.y[1, 2] <- tau.y[1, 1]  # The precision for the pre-1999 post flow threshold model

tau.y[2, 1] <- pow(sigma.y[2], -2)  # The precision for the post-1999 submodel
sigma.y[2] ~ dunif(0, 4)  # The prior on the standard deviation of model residuals

tau.y[2, 2] <- pow(sigma.y[3], -2)  # The precision for the post-1999 post flow threshold submodel
sigma.y[3] ~ dunif(0, 4)
```

```r
for(k in 1:gridi) {
  priort[k] <- punifi[k] / sum(punifi[])  # The probability for each year being the changepoint year
}
```

```r
cp1 <- uniyear[cp]  # The year of the observation

cp ~ dcat(priort[])  # The prior on a certain changepoint year
```

```r
}
```
Appendix C: Flow: Concentration Model: Adjusted Sampling

Data collection at the Fort Barnwell station between 1979 and 2007 passed through several sampling frequencies. Prior to 1996 the sampling effort ranged between monthly and bimonthly. The sampling frequency substantially increased to an almost daily effort between 1997 and 2002, as a direct response to the increased interest in the frequent fish kills and algal blooms that were common during that period. Following 2002, as public interest waned and fish kills became less frequent the sampling effort was reduced to weekly sampling. Changes in sampling frequency is a common occurrence when it comes to environmental monitoring as the sampling frequency is often dictated by available funds. Sampling bias can in some cases result in biases and erroneous model results, if not properly accounted for. In that respect, we tried to test the validity of our model results under a more balanced sampling effort. This was done through the generation of balanced sub-datasets whereby we randomly sampled 4 samples for each month that had in excess of 4 reported DON concentrations. Twenty balanced datasets were generated in total. Note that while the generated balanced datasets still has more samples post 1996, we strongly feel that reducing the temporal resolution further would result in a significant loss of data. We then tried to fit the Bayesian changepoint-threshold model (Equation 12) using each of the balanced dataset. Model results were saved and compared to the results presented in the paper in Table 6 (using the whole dataset). Model convergence was achieved in ten out of the twenty
model runs. The remaining ten runs did not converge within the specified number of iterations (17,000 iterations per chain, number of chains = 3, length of burn = 8,000).

Examining the non-convergent models, we observed that two of the three chains managed to converge on the original model results, while the third chain often got caught in a local minima. In most cases that local minima involved the identification of year 1996 as the changepoint year. We believe that this local minima is most probably either an artifact of the imbalance in sampling (< 54 samples before 1996 versus 54 samples post 1996) or to the model picking up the effects of hurricane Fran or possibly to both. We believe that increasing the number of runs may lead to the convergence of some of the non-converging models. Yet some might not converge within a realistic run time. Changing the distribution on the changepoint years from discrete to a continuous distribution may also help the Metropolis-Hasting to converge within WinBugs. Figure C-1 presents the kernel density estimates for the 11 model parameters. The black density kernel represents the results generated after running the model using the whole dataset, while the dashed grey density kernels estimate the distribution of each of the 11 parameters using the posterior marginal distributions from the 10 models that converged when using the adjusted balanced datasets. All density kernels were drawn using a Gaussian smoothing kernel using the density function in the stats package in the R software. It is evident that using the entire dataset (without adjustment) results in a less dispersed distribution as compared to the adjusted-dataset densities. Moreover,
the results show a reasonable degree of overlap indicating that the model results are not a product of sampling bias.

Figure C-1: (a) density kernels for the pre-1999 intercept ($\alpha_1$) (smoothing bandwidth set to 0.5); (b) density kernels for the log TON concentration at the identified flow threshold for the post-1999 model ($\alpha_2$) (smoothing bandwidth set to 0.2); (c) density kernels for pre-1999 slope on the natural logarithm of flow ($\beta_{1,1}$) (smoothing bandwidth set to 0.2); (d) density kernels for the post-1999 slope on the natural logarithm of flow before the flow threshold ($\beta_{2,1}$) (smoothing bandwidth set to 0.2). The black density kernel represents the results generated after running the model on the whole dataset, dashed grey density kernels correspond to the results generated from the adjusted datasets.
Figure C-1 (cntd.): (i) density kernels of the error term ($\sigma_3$) for the post-flow threshold & post-1999 model (smoothing bandwidth set to 0.07); (j) density kernels of the flow threshold in log m$^3$/sec (smoothing bandwidth set to 0.5); (k) density kernels of Changepoint year with clear centering around the year 1999 (smoothing bandwidth set to 0.5). The black density kernel represents the results generated after running the model on the whole dataset, dashed grey density kernels correspond to the results generated from the adjusted datasets.
Appendix D: TON-TN Relationship in the Neuse River

The Neuse rivers shows a strong linear relationship between TON and TN, which allows us to extend our TON-model results to infer about the status of TN dynamics in the river. Figure D1 shows the linear relationship between TON and TN both in the non-transformed as well as the logarithmic scale. This linearity is to be expected as TON is the major constituent of TN on most river systems. Note the relatively high values of TN values recorded (> 5 mg/L) that will have high leverage on linear regression models used to fit the data. These points should be individually checked to ascertain if they are outliers or not through the use of either univariate or multi-variate robust outlier detection methods. Furthermore, as indicated by Dodds (2003) the linear relationship between TON and TN weakens at low TON concentrations as evident in Figure D1.
Figure D1: Total Nitrogen (TN)-Total Oxidized Nitrogen (TON) relationship in the Neuse River in non-transformed and log-transformed scales. The relationship between the two is very close to linear.
Appendix E: Monitoring Optimization: Expert Elicitation

The following questionnaire was prepared in order to collect the experts' opinion on the relative importance of three criteria/attributes selected for monitoring optimization. The three criteria represent different aspects of uncertainty:

- **Criterion 1** - Model uncertainty is a measure of model uncertainty. Uncertainty in this case is being assessed as the joint uncertainty associated with both Bottom DO and surface Chl-a levels.

- **Criterion 2** - It is the uncertainty associated with violating the Chl-a standard of 40 μg/L. Locations that we are sure will exceed the standard will have low standard violation uncertainty. Similarly stations that we are sure that they will be in compliance will have low standard violation uncertainty. Uncertainty will be largest when there is an equal probability that a given station will be in violation or in compliance with the Chl-a standard.

- **Criterion 3** - It is the uncertainty associated with violating the DO standard of 4 mg/L. Locations that we are sure will violate the standard will have low standard violation uncertainty. Similarly stations that we are sure that they will be in compliance with the standard will have low standard violation uncertainty. Uncertainty will be largest when there is an equal probability that a given station will be in violation or in compliance with the DO standard.
We evaluated how strongly an expert prefers one criterion with respect to the other using an AHP approach. Experts were asked to rank their preferences using a scale ranging between 1 and 9 (Table F1 and Figure F1) following the conventional method of Analytic Hierarchy Process (AHP) (Saaty 1977, Saaty 1987, Saaty 1990):

Table F1: The semantic scale used in AHP (adapted from Saaty (1987))

<table>
<thead>
<tr>
<th>Weight</th>
<th>Preference Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Equally preferred</td>
</tr>
<tr>
<td>3</td>
<td>Moderately preferred</td>
</tr>
<tr>
<td>5</td>
<td>Strongly preferred</td>
</tr>
<tr>
<td>7</td>
<td>Very strongly preferred</td>
</tr>
<tr>
<td>9</td>
<td>Extremely preferred</td>
</tr>
</tbody>
</table>
1. Do you believe that Model Uncertainty is a more important/appropriate criterion to use as compared to the uncertainty around violating the Chl-a Standard for deciding the location of a new monitoring station?
   - Yes ☐
   - No ☐

   a. If your answer to Q.1 is YES, please indicate your preference using the preference scale between 1 to 9.
   1: indicates that you equally prefer Model Uncertainty to Chl-a Standard Violation uncertainty.
   9: indicates that you find that Model Uncertainty is extremely more important than the Chl-a Standard Violation uncertainty.

   b. If your answer to Q.1 is NO, then how much do you prefer Chl-a Standard Violation uncertainty to Model Uncertainty? Please indicate your perceived preference level using the preference scale between 1 to 9.
   1: indicates that you equally prefer Model Uncertainty to Chl-a Standard Violation uncertainty.
   9: indicates that you find that Chl-a Standard Violation uncertainty is extremely more important than Model Uncertainty.

2. Do you believe that Model Uncertainty is a more important/appropriate criterion to use as compared to the uncertainty around violating the DO Standard for deciding the location of a new monitoring station?
   - Yes ☐
   - No ☐

   a. If your answer to Q.2 is YES, please indicate your preference using the preference scale between 1 to 9.
   1: indicates that you equally prefer Model Uncertainty to DO Standard Violation uncertainty.
   9: indicates that you find that Model Uncertainty is extremely more important than the DO Standard Violation uncertainty.

   b. If your answer to Q.2 is NO, how much do you prefer DO Standard Violation uncertainty to Model Uncertainty? Please indicate your perceived preference level using the preference scale between 1 to 9.
   1: indicates that you equally prefer Model Uncertainty to DO Standard Violation uncertainty.
   9: indicates that you find that DO Standard Violation uncertainty is extremely more important than Model Uncertainty.

3. Do you believe that the uncertainty around violating the Chl-a Standard is a more important/appropriate criterion to use as compared to the uncertainty around violating the DO Standard for deciding the location of a new monitoring station?
   - Yes ☐
   - No ☐

   a. If your answer to Q.3 is YES, please indicate your preference using the preference scale between 1 to 9.
   1: indicates that you equally prefer Chl-a standard violation uncertainty and DO Standard Violation uncertainty.
   9: indicates that you find that Chl-a Standard Violation uncertainty is extremely more important than the uncertainty around DO Standard Violation.

   b. If your answer to Q.3 is NO, how much do you prefer DO Standard Violation uncertainty to Chl-a standard violation uncertainty? Please indicate your perceived preference level using the preference scale between 1 to 9.
   1: indicates that you equally prefer Chl-a standard violation uncertainty and DO Standard Violation uncertainty.
   9: indicates that you find the uncertainty around DO Standard Violation is extremely more important than the Chl-a Standard Violation uncertainty.

Name: __________________________

Notes: ____________________________

Figure F1: Expert-based questionnaire for assessing monitoring criteria weights
Appendix F: Analytical Hierarchical Process

The following steps were followed for conducting AHP (Saaty 1980; Yoon and Hwang 1995; Honert and Lootsma 1996; Huang et al. 2009) for multi attribute decision making:

Step I: Setting criteria and alternatives - construct the decision hierarchy – We first identified the criteria, subcriteria, and candidate alternatives. These were used to construct the decision hierarchy shown in Figure E1.

Step II: Performing the pairwise comparisons for attributes – A Pairwise Comparison Matrix (PCM) was developed for each of the three attributes/criteria based on the experts' opinions. If \( w_{ki} \) and \( w_{kj} \) represent the weights provided by expert \( k \) for attributes \( i \) and \( j \), respectively, then \( a_{kij} = w_{ki}/w_{kj} \) represents the pairwise judgment between the two attributes (in Saaty’s scale which ranges between 1/9 to 9). As such, a-
\( k_{ij} \) quantifies the importance of attribute \( i \) over attribute \( j \) based on the judgment of individual \( k \). The pairwise comparison information can thus be concisely represented in a matrix, \( A_k \), whose elements in row \( i \) and column \( j \) are populated with the ratios of row \( i \) and column \( j \). The element in the transpose position, i.e., \( a_{kji} \) are assigned a value of \( 1/a_{kij} \). Consistency tests were performed on the generated PCMs in order to ensure the consistency of the experts' judgment. Following Saaty (1977), a PCM with positive entries is deemed consistent iff \( \lambda_{\text{max}} = n \), where \( \lambda_{\text{max}} \) is the principal eigenvalue and \( n \) is the number of attributes being compared. The eigenvalue of matrix \( A_k \) were derived by solving the equation: \((A-\lambda I)x = 0\), where \( \lambda \) is the eigenvalue, \( I \) is the identity matrix, and \( x \) is the eigenvector.

The deviation from consistency is termed the Consistency Index (CI) = \([\lambda_{\text{max}} - n]/(n-1)\). The ratio of CI to the Average Random Index (ARI) is called as Consistency Ratio (CR). The PCM is deemed "consistent" if its CR value is found to be less than or equal to 0.1 (Saaty 1990, Saaty and Vargas 1991). Note that the ARI values are sampled from randomly generated reciprocal matrices. The ARI for value of \( n \) between 1 and 15 and between 20 and 25 were obtained from Saaty (1990) and Stein and Mizzi (2007), respectively.

Step III: Obtaining the decision weights or priorities for the attributes - Following the consistency check, the decision priorities for each PCM were obtained. The final decision priority (\( P_{ki} \)) for attribute \( i \) for expert \( k \) is aggregated through the normalization of the geometric mean of the pairwise judgment values between two attributes:
Step IV: Performing pairwise comparisons for alternatives - The PCM between alternatives for a given criterion were derived in a similar manner to the PCMs generated for the attributes. Since the attribute data (i.e., total system entropy, Chl-a and DO standards violation entropies) were computed by the developed Bayesian spatio-temporal model, a \((n \times n)\) PCM (where \(n = 25\) and represents the number of ungauged sites) for each attribute was derived numerically. As a first step, a \(n \times n\) ratio matrix was formed by taking the ratio of the normalized attribute values for each pair of alternatives. For example, the ratio for the total system entropy between station 1 and 2 was placed in the \((1 \times 2)\) position, whereas the reciprocal value was placed in the \((2 \times 1)\) position within the ratio matrix. Ratios values greater than or equal to 1 were linearly interpolated to values between 1 and 9. Values were rounded off to obtain integer numbers ranging between 1 and 9. The integer values were placed in their corresponding positions. Their reciprocals were placed in the location of their transpose in the matrix. This process allowed us to generate the PCM for the 25 ungauged locations for the first attribute (i.e., total system entropy). The same procedure was followed to generate the PCMs for the other two criteria/attributes. The consistency of the matrices was checked following the same procedure explained in Step II.

Step V: Obtaining the decision weights or priorities for alternatives – A similar procedure as that explained in Step III was followed to determine the priorities between
the 25 proposed ungauged locations. This was conducted for each of the three
criteria/attributes considered.

Step VI: Aggregation of group decisions - Aggregation of priorities to obtain the
overall priority matrix was conducted through the use of two methods: (1) the first
method followed the conventional AHP (Saaty 1990), (2) the second method followed
the multiplicative AHP for group decision making. In multiplicative AHP, the PCMs of the
three proposed attributes were not aggregated. The priority \( P_{ak} \) for attribute \( a \) based
on expert \( k \) was directly combined with the priority \( P_{Si a} \) for location \( Si \) and attribute \( a \)
through the following equation:

\[
p_{Si}^k = \prod_{a=1}^{3} (P_{Si a})^{P_{ak}}; \quad \forall \, Si = 1, 2, ..., 25 \text{ and } \forall \, k = 1, ..., 4. \tag{Equation E2}
\]

where \( P_{Si}^k \) is the priority of location \( Si \) based on expert \( k \).

The final decision priority \( P_{Si} \) for location \( Si \) was calculated by aggregating the
normalized geometric means of the rows form the generated matrix, i.e.,

\[
P_{Si} = (\prod_{k=1}^{4} P_{Si}^k)^{1/4} / \sum_{Si=1}^{25} (\prod_{k=1}^{4} P_{Si}^k)^{1/4}; \quad \forall \, Si = 1, ..., 25. \tag{Equation E3}
\]

The rankings of the proposed locations were derived based on the final \( P_{Si} \)
values.
References


regional climate change impact studies. Journal of Great Lakes research 17:461-469.


Deamer, N. 2009. Neuse river basinwide water quality plan. NC Department of Environment and Natural Resources: Division of Water Quality, Planning Section – Basinwide Planning Unit, Raleigh, NC.


255


StataCorp LP. 2009. Stata: Data analysis and statistical software. College Station, TX.


Biography

Ibrahim Alameddine was born in Beirut, Lebanon on May 31 1978. He completed his Bachelor of Science in Biology at the American University of Beirut (AUB) in June 1999. He then went on to receive his Masters of Science in Environmental Technology from the Civil and Environmental Engineering Department at AUB in June 2002. Following his graduation, he worked as an environmental consultant on several World Bank, European Union, United Nations, and United States Agency for International Development funded project across the Middle East and North Africa.

He joined the Nicholas School of the Environment in 2005 to pursue his doctoral studies in water quality modeling and analysis under the guidance of Dr. Kenneth H. Reckhow. His research has focused on developing a set of statistical models to better link watershed pollutant loading and estuarine eutrophication in the Neuse Estuary.

During his doctoral studies, he was involved in several research projects that included monitoring optimization in the Neuse Estuary, water quality remote sensing in coastal North Carolina, predicting harmful algal bloom dynamics in the Gulf of Maine, modeling the effect of urbanization on aquatic ecosystems across the US, quantifying bacterial contamination in coastal North Carolina, assessing outlier detection techniques for environmental data, developing a virtual water tool for counties in California, and reassessing nutrient limitations in the Saginaw Bay, several Fennish lakes, and the Neuse Estuary.
Ibrahim was awarded a Quantitative Environmental Analysis, LLC (QEA) Graduate Scholarship in 2008. He also received several Duke University, Nicholas School of the Environment Conference Scholarship. Additionally, he achieved the Honors List at AUB for the years 2000 and 2001. He has presented his work at many national and international conferences. Additionally, he has published more than 30 peer-reviewed articles and technical reports, which include the following:

**Peer refereed journal articles**

**In Review**  

**Accepted**  

**2011**  

**2011**  

**2010**  


**Articles in peer-refereed conference proceedings**


**Scientific and Technical reports**


2006  **Advisory Services to Support Water Resources Management and Remediation of Wastewater and Other Pollution in the Upper Litani River and Qaraoun Basin**: *USAID*. Beirut, Lebanon.


2004  **Pre-design EIA for Lebanon’s Solid Waste Management Plan**: *Lebanese Council for Development and Reconstruction*, Beirut, Lebanon.


2003  **Environmental Impact Assessment Process/Environmental Management Plans**: *Consolidated Contractors Company (CCC)*, Athens, Greece


2003  **Characterization and optimization of the Chekka domestic wastewater outfall to achieve the guidelines set forth by the Lebanese Ministry of Environment and MED POL II**, Beirut, Lebanon: *Council for Development*


2002 El-Fadel, M., Alameddine, I., Kazopoulo, M. EIA of the Al-Qasr Wastewater Treatment Plant, Al Qasr, Lebanon: USAID, Cooperative Housing Foundation (CHF), Municipality of Al-Qasr, Lebanon.


2002 El-Fadel, M., Alameddine, I., Kazopoulo, M. EIA of the Shebaa Dam and Water Lake: USAID, Mercy Corps Lebanon, Municipality of Shebaa, Lebanon


2002 Environmental Compliance and Phasing Program for the Food Industry (Olive Pressing), Lebanese Ministry of the Environment, Beirut, Lebanon.