

ABSTRACT

Title of Document: USE OF MACROINVERTEBRATE
PREDICTIVE MODELS TO EVALUATE THE
STREAM RESTORATION EFFECT

Yin-Phan Tsang, Ph.D., 2008

Directed By: Associate Professor Gary K. Felton,
Environmental Science and Technology

Multivariate analysis was used to build macroinvertebrate predictive models for stream assessment in Britain, Australia, and the west coast of the United States. The philosophy behind these predictive models was similar, but variations exist and have been adapted for different regions. The macroinvertebrate predictive model in Maryland has been improved using Region-style models, including the Assessment by Nearest Neighbour Analysis (ANNA), the Burn's Region of Influence (BROI), and the New Datum Region of Influence (NROI) predictive schemes. For better prediction precision, different parameter selection methods (stepwise AIC, exhaustive AIC, and exhaustive BIC) and rational multiple regression function checking have been used to prevent overfitting. Root mean squared error (RMSE) was used to select the final best model. The calibration results from the Region-Style models are better than those from previously built River InVertebrate Prediction And Classification System (RIVPACS)-style model. The different parameter selection criteria along with

rational regression function checking discourage overfitting and improve the prediction results. Region-style methods can be alternative methods for building predictive model. GISHydro2000 is a GIS-based program for performing hydrologic analysis in Maryland. This tool was used to determine numerous hydrologic characteristics as potential predictors to be used in the macroinvertebrate predictive model. The best performing ANNA, BROI, and NROI predictive models can be automated in the GISHydro2000 environment. Theses multivariate analyses (i.e., Observed/Expected (O/E) scores), as well as multimetric analysis (i.e., Benthic Index of Biotic Integrity (IBI) metrics), were applied to evaluate the stream restoration sites in Montgomery County, Maryland. The evaluation results show most stream habitat conditions were still degraded after stream restoration projects. The environmental stressors at the stream site were not immediately alleviated by the restoration design, or the stressors overshadowed the restoration efforts. At many sites, the stream condition starts to recover at the 3rd - or 4th - year post-restoration. More time may be needed for monitoring the recovery of stream ecosystems. The benthic IBI metrics response to not only environmental stressor, but also other natural variances. The results suggested that O/E scores from multivariate analysis provides valuable supplemental information for evaluating stream health.

USE OF MACROINVERTEBRATE PREDICTIVE MODELS TO EVALUATE
THE STREAM RESTORATION EFFECT

By

Yin-Phan Tsang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Associate Professor Gary K. Felton, Chair
Professor and Associate Chair Adel Shirmohammadi
Associate Professor Glenn E. Moglen
Associate Professor William O. Lamp
Associate Professor Hubert Montas

© Copyright by
Yin-Phan Tsang
2008

Dedication

To Grandma, Huang-Yang, Shan,
Who now is in heaven looking after me and my mom.

To Grandpa, Huang, Fu-Ju.
I did it, granpa. I'm going hiking with you again, as soon as I can.

To Mom, Huang, Hsunshu.
I did it as you believe I will.

Acknowledgements

I would like to show my acknowledgments to the Maryland Biological Stream Survey conducted by Department of Natural Resource, Maryland, and the Department of Environmental Protection in Montgomery County, Maryland, especially the help from Keith Van Ness, Alicia Youmans, and Daniel Harper. This research can not be done without their valuable data.

I want to first address my gratitude to my advisor, Dr. Gary Felton, who has been very supportive to my research interests and provided any possible help and advice. I cherished and grew up from this study experience. I know not many students have the opportunity to study their interests like I had. I very appreciate his support. I want to thank Dr. Glen Moglen, who spent time discussing the research with me and taught me all he knows about GIS. I wish I will become a good researcher like him one day. I want to thank Dr. Adel Shirmohammadi, who advised and help me with my research and academic career. I want to thanks Dr. William Lamp, who I learned all the knowledge about biomonitoring and macroinvertebrate from. I want to thank Dr. Hubert Montas, who pointed out my blind point in developing models and the interpretation of results. I am blessed to have these committees during my study and I am grateful to have them.

This research could not be done without a lot of helps. Mike Paul taught me everything about the RIVPACS model, shared the data, and discussed the results with me. I am so glad to see him in my defense. I want to thank Mu Zhu, who answered my question in parameter selection though e-mail, and John Van Sickle, who

explained to me how to use RMSE to select the best model. I want to thank Shu-Chih Yang, who provided her work station for me and thus speeded up my model iteration process. I appreciate her support not just in my research but in my life through the years. I want to thank Sean Smith who discussed the stream restoration with me and point out the deficiency of my models. I want to thank Raymond Morgan, who gave me suggestion in the result interpretation and provided reference.

I want to thank Aisha Sexton, who has been my mentor since the first day I came in the office. I know her support in my life and study will never end. I am blessed to have a sister and friend like her. I want to thank my graduate school friends, Shaya, Melissa, and Preston. I enjoyed the time with you all. Preston is taking off to “play” the army in Afghanistan soon. May God bless him. I want to thank all my fellow graduate students. It’s been a pleasure to know you.

I want to thank my husband, Kuolin Fang. It’s not easy to have long distance relationship. We did it and finally get to be together. I want to thank my mom, who has been supportive and believes I can do this since long time ago. I want to thank my landlady friend, Dona Karlyne, who makes me feel like home. We had fun time together. And finally I want to remember Harley, Rose, Pantera, and Baby, who made my life in Chickadee Lane joyful.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	v
List of Tables	viii
List of Figures	xi
CHAPTER 1: INTRODUCTION	1
1.1 JUSTIFICATION	1
1.2 LITERATURE REVIEW	5
1.2.1 Biological monitoring	5
Multimetric Analysis	6
Multivariate Analysis	9
1.2.2 RIVPACS-Style Model	10
1.2.3 Assessment by Nearest Neighbour Analysis (ANNA) Scheme	14
1.2.4 Region of Influence (ROI) Scheme	15
1.2.5 Maryland Biological Stream Survey (MBSS) Data	16
1.2.6 Stream Restoration	17
1.3 OBJECTIVES	18
CHAPTER 2: REGION-STYLE MODEL DEVELOPMENT – METHOD	20
2.1 THE FIVE GENERAL STEPS	20
Step 1. Convert reference site biodiversity into multidimensional scaling.	20
Non-metric Multidimensional Scaling (NMDS)	21
Bray-Curtis Distance	22
Step 2. Weight the predictor variables using a multivariate approach	23
Step 3. Calculate the weighted Euclidian distance from a test site to the reference sites based on the environmental predictors.	24
Step 4. Predict the faunal composition based on the nearest reference sites and predictors	25
Step 5. Calculate an observed/expected (O/E) ratio analogous to RIVPACS/AUSRIVAS	25
2.2 FROM STEP 2 – THE POTENTIAL OVERFITTING ISSUE	27
2.2.1 Cross-Validation	30
2.2.2 Rational Multiple Regression Function Checking	32
2.2.3 Stepwise AIC method	34
2.2.4 Exhaustive AIC and Exhaustive BIC	34
2.3 FROM STEP 4 – DIFFERENT SCHEMES FOR SELECTION REGION	38
2.3.1 Assessment by Nearest Neighbour Analysis (ANNA) Scheme	38
2.3.2 Burn's Region of Influence (BROI) Scheme	39
2.3.3 New Datum Region of Influence (NROI) Scheme	42
2.3.4 Tasker's Region of Influence (TROI) Scheme	43
CHAPTER 3: REGION-STYLE MODELS FOR MARYLAND – METHOD	46
3.1 BIODIVERSITY DATA	46
3.2 CANDIDATE PREDICTORS LIST	47
3.2.1 Data Transformation	50

3.2.2 Soil data	50
3.2.3 Two sets of data	51
CHAPTER 4: MODEL PARAMETERS FOR MARYLAND – RESULTS	53
4.1 SUBSET PREDICTORS AND OVERFITTING ANALYSIS RESULT	53
4.1.1 Cross-Validation	53
4.1.2 Exhaustive AIC, Exhaustive BIC	57
4.1.3 Rational Regression Checking	64
4.2 WEIGHTING FUNCTION PARAMETERS AND MODEL SELECTION	70
4.2.1 For the ANNA scheme.....	73
4.2.2 For the BROI and NROI schemes	73
CHAPTER 5: REGION-STYLE MODEL RESULTS	76
5.1 MODELS WITH ONLY GISHYDRO2000 PREDICTORS.....	79
5.1.1 Family Level	79
5.1.2 Genus Level	84
5.2 MODELS WITH ALL PREDICTORS	89
5.2.1 Family Level	89
5.2.2 Genus Level	94
CHAPTER 6: DISCUSSION AND CONCLUSIONS FROM THE MODEL RESULTS	99
6.1 COMPARING THE REGION-STYLES	99
6.1.1 Family Level	99
6.1.2 Genus Level	104
6.1.3 Individual validation reference site comparison	108
6.1.4 Residual Analyses	112
6.2 COMPARING THE REGION-STYLES AND RIVPACS-STYLE MODEL RESULTS	122
6.2.1 Family Level	122
6.2.2 Genus Level	126
6.3 CONCLUSIONS	130
CHAPTER 7: APPLYING THE PREDICTIVE MODEL TO STREAM RESTORATION DATA WITH GIS TOOL	132
7.1 MONTGOMERY COUNTY AND STREAM RESTORATION PROJECTS BACKGROUND	132
7.1.1 Montgomery County.....	132
7.1.2 Stream Restoration Projects by the Montgomery County DEP.....	133
7.2 APPLICATION METHODS	142
7.2.1 GISHydro2000 Applications	142
Building Region-Style Models in GISHydro2000.....	143
Manual for using Region-Style Models in GISHydro2000	146
Predicting macroinvertebrates at the stream restoration site with Region-Style Models in GISHydro2000.....	152
7.2.2 Application of Stream Restoration Biomonitoring Data	153
Identifying the observed taxon names in the family level	153
Assigning the O according to E	154
CHAPTER 8: STREAM RESTORATION SITE APPLICATION RESULTS	155
8.1 MACROINVERTEBRATE PREDICTIVE RESULTS FROM DIFFERENT REGION-STYLE MODELS	155
8.1.1 The ANNA Scheme	156

8.1.2 The BROI Scheme	160
8.1.3 The NROI Scheme	164
8.2 THE MACROINVERTEBRATE PREDICTIVE RESULTS AT STREAM RESTORATION SITES.....	167
8.2.1 Family Level	167
8.2.2 Genus Level	177
8.3 MULTIMETRIC RESULTS FROM STREAM RESTORATION SITES	185
CHAPTER 9: DISCUSSION AND CONCLUSIONS	204
9.1 DISCUSSION	204
9.1.1 The Multivariate Analysis.....	204
9.1.2 The Multimetric Analysis	206
9.1.3 The Comparison.....	207
9.2 CONCLUSIONS	213
9.3 FUTURE RESEARCH	215
Appendix A-1. Potential Predictor Distribution	221
Appendix A-2. Transformed Potential Predictor Distribution.....	224
Appendix B. R-Code for Region-Style Models.....	227
Appendix C-1. Avenue Code for GISHydro2000.....	275
Appendix C-2. FORTRAN-Code for GISHydro2000.....	295
Appendix D. 27 Validation Reference Sites	355
Appendix E. Original Predicted Taxa Probabilities At Stream Restoration Sites	382
Appendix F. The BIBI and O/E Comparison at Restoration Stream Sites	419
References.....	423

List of Tables

Table 1. Quantitative criteria for identifying reference sites in Maryland (Roth et al. 1997)	7
Table 2. Summary table for candidate predictors	49
Table 3. Two sets of parameters used in this study	52
Table 4. The best combination for the family level with GISHydro2000 predictors using AIC_c and BIC criteria.	59
Table 5. The best combination for the family level with all predictors using AIC_c and BIC criteria.....	60
Table 6. The best combination for the genus level with GISHydro2000 predictors ..	61
Table 7. The best combination for the genus level with all predictors	62
Table 8. Summary parameter selection results from different criteria (AIC: Exhaustive AIC, BIC: Exhaustive BIC)	63
Table 9. Rational regression checking result for the family level with GISHydro2000 predictors.....	64
Table 10. Rational regression checking result for the family level with all predictors.	65
Table 11. Rational regression checking result for genus level with GISHydro2000 predictors.....	66
Table 12. Rational regression checking result for the genus level with all predictors.	67
Table 13. Summary parameter selection results from different criteria after rational regression checking. (AIC: Exhaustive AIC, BIC: Exhaustive BIC).....	68
Table 14. Weighting function parameter testing ranges in different region-style methods	74
Table 15. O, E values comparison between outliers and calibration, validation data.	78
Table 16. The O/E statistics at $P_c=0.5$ from all the Region-style models for the case in family level with GISHydro2000 predictors.....	80
Table 17. The O/E statistics at $P_c=0.5$ from all the Region-style models for the case in genus level with GISHydro2000 predictors.....	85
Table 18. The O/E statistics at $P_c=0.5$ from all the Region-style models for the case in the family level with all predictors	90
Table 19. The O/E statistics at $P_c=0.5$ from all the Region-style models for the case in genus level with all predictors	95
Table 20. Region-Style Model: number of reference sites for the family level using only GISHydro2000 predictors.....	102
Table 21. Region-Style Model: number of neighbors' selection for the family level using both GISHydro2000 and MBSS predictors.....	103
Table 22. Region-Style Model: number of reference sites at the genus level with GISHydro2000 predictors.....	106
Table 23. Region-Style Model: number of neighbors' selection at the genus level with all predictors.....	107

Table 24. Region-Style number of neighbors' selection for validation sites in family level with GISHydro2000 predictors ($P_c = 0.5$)	110
Table 25. Region-Style number of neighbors' selection for validation sites in genus level with GISHydro2000 predictors ($P_c = 0.5$)	111
Table 26. Residual Analyses for Region-Style Models at the family level with only GISHydro2000 predictors.....	116
Table 27. Residual Analyses for Region-Style Models at the genus level with only GISHydro2000 predictors.....	116
Table 28. Residual Analyses for Region-Style Models at the family level with all predictors.....	121
Table 29. Residual Analyses for Region-Style Models at the genus level with only all predictors.....	121
Table 30. RIVPACS-style family level result with GISHydro2000 predictors (adapted from Moglen and Paul, 2005)	122
Table 31. RIVPACS-style family level results with All predictors (adapted from Paul, 2006 symposium presentation)	125
Table 32. RIVPACS-style genus level result with GISHydro2000 predictors (adapted from Moglen and Paul, 2005)	127
Table 33. RIVPACS-style genus level results with all predictors (adapted from Paul, 2006 symposium presentation)	128
Table 34. The summary table of stream restoration projects by the Montgomery County DEP. (Adapted from Montgomery County DEP data sheet) (! : First post-monitoring done in the same year as construction)	137
Table 35. The parameter averages of reference sites chosen by ANNA for LFLF 301B stream restoration site in the family level	158
Table 36. The parameter averages of reference sites chosen by ANNA for LFLF 301B stream restoration site in the genus level	158
Table 37. The reference sites chosen by BROI for LFLF 301B stream restoration site	160
Table 38. The parameter averages of reference sites chosen by BROI for LFLF 301B stream restoration site in the family level.....	161
Table 39. The parameter averages of reference sites chosen by BROI for LFLF 301B stream restoration site in the genus level.....	161
Table 40. The reference sites chosen by NROI for LFLF 301B stream restoration site	164
Table 41. The parameter averages of reference sites chosen by NROI for LRLR 301B stream restoration site in the Family level	165
Table 42. The parameter averages of reference sites chosen by NROI for LRLR 301B stream restoration site in the Genus level.....	165
Table 43. The summary table of Expected (E) and Observed (O) at the Family Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.).....	171
Table 44. The summary table of O/E scores at the Family Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.).....	175
Table 45. The summary table of Expected (E) and Observed (O) for the Genus Level at stream restoration sites in Montgomery County.	179

Table 46. The summary table of O/E scores for the Genus Level at stream restoration sites in Montgomery County.....	183
Table 47. The benthic metrics for the Piedmont region in Maryland. (Adapted from Southerland et al. 2005) Note that the new benthic IBI is intended for use with subsample sizes ranging from 80 to 120 individuals.....	186
Table 48. IBI score ranges and corresponding narrative ratings. (From Stribling et al. 1998).....	188
Table 49. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individual not within the IBI defined range) (Please refer to Table 34 for Site ID.).....	190
Table 50. The summary table of benthic IBI at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.).....	202

List of Figures

Figure 1. Illustration of sample size effect on showing true taxa number.....	2
Figure 2. 158 reference sites locations in Maryland.....	47
Figure 3. Family level cross validation results. (a)GISHydro2000 generated predictors (b)All predictors.....	55
Figure 4. Genus level cross validation results. (a)GISHydro2000 generated predictors (b)All predictors.....	56
Figure 5. Exhaustive AIC, BIC searches at the family level with GISHydro2000 predictors. Open circles denote axis 1. Open triangles denote axis 2. Crosses denote axis 3. The number after the hyphen is the model oder with lowest AIC or BIC at respective axis.....	59
Figure 6. Exhaustive AIC, BIC searches at the family level with all predictors. Open circles denote axis 1. Open triangles denote axis 2. Crosses denote axis 3. The number after the hyphen is the model oder with lowest AIC or BIC at respective axis.	60
Figure 7. Exhaustive AIC, BIC searches at the genus level with GISHydro2000 predictors. Open circles denote axis 1. Open triangles denote axis 2. Crosses denote axis 3. The number after the hyphen is the model oder with lowest AIC or BIC at respective axis.....	61
Figure 8. Exhaustive AIC, BIC searches at the genus level with all predictors. Open circles denote axis 1. Open triangles denote axis 2. Crosses denote axis 3. The number after the hyphen is the model oder with lowest AIC or BIC at respective axis.	62
Figure 9. The validation data outliers in the O, E plots. O/E histograms are on the top and O vs.E scatter plots are on the bottom. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are from the left to right.....	78
Figure 10. The O/E histograms and E vs.O scatter plots from best ANNA model at the family level with GISHydro2000 predictors. Exhaustive BIC method was used. Neighbor number is 21. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	81
Figure 11. The O/E histograms and E vs.O scatter plots from best BROI model at the family level with GISHydro2000 predictors. Stepwise AIC method was used. $n =$ 2.5 , $NST = 8$, $\Theta_L = 10\%$, $\Theta_U = 30\%$, $TP = 50\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.....	82
Figure 12. The O/E histograms and E vs.O scatter plots from best NROI model at the family level with GISHydro2000 predictors. Exhaustive AIC method was used. $n=2.5$, $NST = 5$, $\Theta_L = 10\%$, $\Theta_U = 40\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.....	83
Figure 13. The O/E histograms and E vs.O scatter plots from best ANNA model at the genus level with GISHydro2000 predictors. Exhaustive AIC method was used. Neighbor number = 16. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	86

Figure 14. The O/E histograms and E vs.O scatter plots from best BROI model at the genus level with GISHydro2000 predictors. Exhaustive AIC method was used. $n = 2$, $NST = 11$, $\Theta_L = 5\%$, $\Theta_U = 30\%$, $TP = 75\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.....	87
Figure 15. The O/E histograms and E vs.O scatter plots from best NROI model at the genus level with GISHydro2000 predictors. Exhaustive AIC method was used. $n = 4$, $\Theta_L = 6\%$, $\Theta_U = 40\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	88
Figure 16. The O/E histograms and E vs.O scatter plots from best ANNA model at the family level with All predictors. Exhaustive BIC method was used. Neighbor number = 34. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	91
Figure 17. The O/E histograms and E vs.O scatter plots from best BROI model at the family level with All predictors. Exhaustive AIC method was used. $n = 3.5$, $NST = 8$, $\Theta_L = 5\%$, $\Theta_U = 25\%$, $TP = 50\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	92
Figure 18. The O/E histograms and E vs.O scatter plots from best NROI model at the family level with All predictors. Exhaustive AIC method was used. $n = 3$, $NST = 6$, $\Theta_L = 6\%$, $\Theta_U = 40\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	93
Figure 19. The O/E histograms and E vs.O scatter plots from best ANNA model at the genus level with All predictors. Stepwise AIC method was used. Neighbor number = 25. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	96
Figure 20. The O/E histograms and E vs.O scatter plots from best BROI model at the genus level with All predictors. Stepwise AIC method was used. $n = 2.5$, $NST = 13$, $\Theta_L = 10\%$, $\Theta = 25\%$, $TP = 50\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	97
Figure 21. The O/E histograms and E vs.O scatter plots from best NROI model at the genus level with All predictors. Stepwise AIC method was used. $n = 4$, $NST = 7$, $\Theta_L = 10\%$, $\Theta_U = 40\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.	98
Figure 22. E vs. Residual at $P_c = 0.5$ for (a) ANNA, (b) BROI, and (c) NROI schemes at the family level with only GISHydro2000 predictors. The calibration cases are on the top, and the validation cases are on the bottom.	114
Figure 23. E vs. Residual at $P_c = 0.5$ for (a) ANNA, (b) BROI, and (c) NROI schemes at the genus level with only GISHydro2000 predictors. The calibration cases are on the top, and the validation cases are on the bottom.	115
Figure 24. E vs. Residual at $P_c = 0.5$ for (a) ANNA, (b) BROI, and (c) NROI schemes at the family level with all predictors. The calibration cases are on the top, and the validation cases are on the bottom.	119
Figure 25. E vs. Residual at $P_c = 0.5$ for (a) ANNA, (b) BROI, and (c) NROI schemes at the genus level with all predictors. The calibration cases are on the top, and the validation cases are on the bottom.	120

Figure 26. RIVPACS-style family level results using GISHydro2000 predictors, 9 groups, $P_c > 0.5$. (a) O/E scores frequency histogram, (b) E vs. O scatter plot. (Adapted from Moglen and Paul, 2005)	123
Figure 27. RIVPACS-style family level results using all predictors, 9 groups, $P_c > 0.5$. (a) O/E scores frequency histogram, (b) E vs. O scatter plot. (Adapted from Paul, 2006 symposium presentation)	125
Figure 28. RIVPACS-style genus level results using GISHydro2000 predictors, 7 groups, $P_c > 0.5$	127
Figure 29. RIVPACS-style genus level results using all predictors, 7 groups, $P_c > 0.5$. (a) O/E scores frequency histogram, (b) E vs. O scatter plot (Adapted from Paul, 2006 symposium presentation)	129
Figure 30. Location of Montgomery County in Maryland. (From Montgomery County Government website)	133
Figure 31. The Watersheds within Montgomery County. (From Montgomery County Government website)	134
Figure 32. The stream restoration sites in Montgomery County	141
Figure 33. The reference sites used by ANNA for Site LFLF 301B in Montgomery County	159
Figure 34. The reference sites used by BROI for site LFLF 301B in Montgomery County	163
Figure 35. The reference sites used by NROI scheme for site LFLF 301B in Montgomery County	166
Figure 36. The stream condition trends at restoration site PBGS 111 captured by different analysis methods	211
Figure 37. The stream condition trends at restoration site PBPB 305C captured by different analysis methods	212

CHAPTER 1: INTRODUCTION

1.1 JUSTIFICATION

The Maryland Biological Stream Survey (MBSS), executed by the Department of Natural Resources (DNR) in Maryland, has developed an extensive biological, chemical, and physical data base of streams within Maryland. The raw data from MBSS has been used in many studies and research projects. These projects are listed on the DNR website (http://www.dnr.state.md.us/streams/mbss/mbss_pubs.html). Invertebrate prediction models have been proven effective in and of themselves and these models are helpful in interpreting stream monitoring data (e.g., RIVPACS in the UK (Moss et al., 1987; *RIVPACS (River Invertebrate Prediction and Classification System)* ; Wright, 1995; Wright, 2000; Wright et al., 1984; Wright et al., 2000), AUSRIVAS in Australia (Wright, 2000), and RIVPACS-style in America's west coast (Hawkins et al., 2000), etc.). One of the key steps in building these models is to cluster invertebrate taxa sample at reference sites either at the genus or family level and then use discriminant function analysis to link the environmental features and biological data. Stream biodiversity at a specific testing site can then be predicted by using reference sites with similar environmental features. Moglen and Paul (2005) have constructed a RIVPACS-style model using the MBSS data base with the GISHydro2000 application. The results showed that although the genus level model predicted a larger percentage of predictable variance than at the family level, the family level model has better precision. But these models were still considered imprecise, relative to the models that were built in the UK and Australia (Moglen and

Paul, 2005). This unsatisfactory result may be due to the limitation in clustering reference sites. For the clustering step, the goal was to have as many clusters as possible to represent the continuous distribution, while at the same time avoiding small clusters (<5 sites) (Moglen and Paul, 2005). However, an ecosystem is a continuous system, and the clustering step treats the system discretely and is artificial and unnatural (Linke et al., 2005), so there may be an alternative way to describe the system. According to Dr. Michael Paul, the unsatisfactory result may be due to the MBSS data (Personal Communication, Paul, 2005). MBSS data are not designed for predictive modeling, so only 100 invertebrates were selected from each sample, while 500 individuals from each sample are preferred for building a model (Karr and Chu, 1999). One hundred invertebrates may not reveal the true variance of the fauna composition within the site. This concept, obtained from both consulting with Dr. Paul and from Karr and Chu (1999) is illustrated in Figure 1.

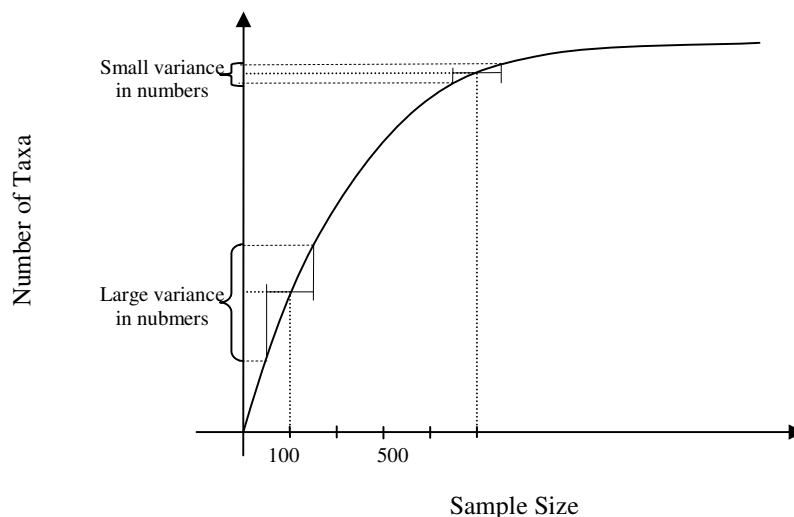


Figure 1. Illustration of sample size effect on showing true taxa number

On the other hand, because the clustering step is considered an artificial part of the model, Linke et al. (2005) suggested passing up the clustering step. They applied

nearest reference sites community composition to predict the faunal composition (Assessment by Nearest Neighbour Analysis, ANNA), thus treating the macroinvertebrate assemblage as a continuum instead of discrete groups. This alternative prediction scheme has been examined and compared with AUSRIVAS, but not the RIVPACS model. The results show that ANNA is potentially more robust for Observed/Expected (O/E) regression and more accurate on trace metal gradient sites. Hydrologists have been focusing on the issue of estimating the flood frequency at unmonitored locations. They have been using different methods to approach this problem (Acreman and Wiltshire, 1987; Burn, 1990a; Tasker et al., 1996; Tasker and Slade, 1994). This estimation issue and their approaches are really similar to the taxa prediction problem and the approach developed by biologists. One of their methods, Region of Influence (ROI) (Burn, 1990a, 1990b; Tasker et al., 1996; Tasker and Slade, 1994), is practically the same as the above mentioned ANNA scheme in concept. Both ANNA and ROI are studied to improve the macroinvertebrate predictive model. As for Maryland, in order to build a better predictive model, stream site sampling protocols should be improved because more individuals should be pulled out. At this time, ANNA and ROI may give us a better predictive model using the existing data. This research constructs a macroinvertebrate predictive model with ANNA and ROI using the same MBSS data base and GISHydro2000(Moglen, 2005) application, and examines and compares with the previous results from RIVPACS-style model.

The recognition of stream restoration as a Best Management Practice for many water quality benefits is increasing. However, in the last decade, stream restoration

related projects haven't been properly evaluated or monitored (Bernhardt et al., 2005). Based on National River Restoration Science Synthesis (NRRSS) results, Bernhardt et al. (2005) proposed that stream restoration projects should be described, the results measured, and all the information should be tracked in a readily accessible data base. Not only should data be collected, but the success of stream restoration projects should be evaluated to identify both more successful approaches and less successful approaches to guide and target future stream restoration expenditures. Meanwhile, little agreement exists on what is a successful river restoration effort. Palmer et al. (2005) proposed five criteria as standards for ecologically successful restoration. One of the criteria is that the river's ecological condition must be measurably improved. Biomonitoring has been recognized as an important method to assess stream health, so it has the potential to evaluate the success of stream restoration.

Invertebrate modeling is a science which can predict the expected biodiversity for a site. This, in the future, can be used to review the effectiveness of stream restoration. Currently, multimetric analysis is the most popular approach in the United State when biomonitoring. It needs reference sites from other locations to evaluate the stream condition. Predictive models which use multivariate analysis on the other hand, can predict the reference biodiversity condition in any given site. This advantage is useful and helpful for most engineers when designing stream restoration at any site. With collected biological data before and after stream restoration from the Montgomery County Environmental Protection Department, this research will assess efforts in stream restoration within Montgomery County with the invertebrate

predictive model. The restoration design along with environmental features can be assessed from an engineering point of view with this tool in the future. Ultimately, if this project is successful, then the predictive model built with MBSS data should be a stream restoration evaluating tool for the entire state of Maryland.

1.2 LITERATURE REVIEW

1.2.1 Biological monitoring

Water quality monitoring started as mankind discovered that our activities pollute our water resources and affect our health. Initially, chemical monitoring was the focus. However, after decades of monitoring and remediation, water resources throughout the United States were still significantly degraded (U.S.EPA., 1995). With the understanding that humans are part of the biological world and that human life relies on biological systems, the water resources now are valued not just for human health but also for ecological integrity in aquatic habitats. Upon realizing aquatic life is the product of the aquatic environment and representative of the aquatic system, biological monitoring has been recognized and accepted as a valid water resources health indicator (Karr and Chu, 1999).

In general, there are two major analysis methods in biomonitoring: Multimetric and Multivariate Analysis. Multimetric analysis for benthic macroinvertebrates is the most widely used approach for water-quality assessment among the aquatic resource agencies in the United States (U.S.EPA., 1999). This study focuses on using multivariate analysis, more details on methods and development are described in later chapters. Multimetric analysis is simply explained here.

Multimetric Analysis

Multimetric analysis for biomonitoring assessment is an iterative process where the site classification and metric selections are revisited at various stages of the analysis. In general, it has two phases. The first phase is the selection and calibration of the metrics and subsequent aggregation of metrics into an index according to homogenous site classes; the second phase is the assessment of biological conditions at sites and a judgment of impairment. As for the first phase, it can be divided into five steps (Davis and Simon, 1994; U.S.EPA., 1999).

Step 1: Classify the Stream Resource

The purpose of classification is to partition the stream reference sites into relatively homogeneous groups with regard to physical, chemical, and biological attributes. Candidate reference sites that are based on minimally degraded physical habitat and water chemistry are used as the basis for stream classification. The quantitative criteria for identifying reference sites for Maryland (Roth et al., 1997) are presented below. A reference site must meet all 12 criteria.

Table 1. Quantitative criteria for identifying reference sites in Maryland (Roth et al. 1997)

Reference site must meet all the 12 criteria	
1.	pH \geq 6; if blackwater stream, then pH $<$ 6 and DOC \geq 8 mg/l
2.	ANC (Acid Neutralizing Capacity) \geq 50 μ eq/l
3.	DO \geq 4 ppm
4.	nitrate \leq 300 μ eq/l
5.	urban land use \leq 20% of catchment area
6.	forest land use \geq 25% of catchment area
7.	remoteness rating: optimal or suboptimal
8.	aesthetics rating: optimal or suboptimal
9.	instream habitat rating: optimal or suboptimal
10.	riparian buffer width \geq 15 m
11.	no channelization
12.	no point source discharges

Initially, sites are classified according to distinctive geographic, physical, or chemical attributes. Then, classification with biological data is used in order to minimize variance within groups and maximize variance among groups.

Step 2: Identify Potential Measures for Each Assemblage

Select the possible metrics, which are (1) ecologically relevant to the biological assemblage or community under study and to the specified program objectives; and (2) sensitive to stressors and provide a response that can be discriminated from natural variation. Those representative metrics should be selected from each of 4 primary categories: (1) richness measures for diversity or variety of the assemblage (e.g., total number of taxa, number of Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies) (EPT) taxa); (2) composition measures for identity and dominance (e.g., % community similarity, % EPT); (3) tolerance measures that represent sensitivity to perturbation (e.g., % tolerant diatoms, %

sensitive taxa) ; and (4) trophic or habit measures for information on feeding strategies and guilds (e.g., % Clingers, % Filterers, % Scrapers).

Step 3: Select Robust Measures

Robust measures are those specific metrics that can discriminate between good and poor quality ecological conditions regarding various stressors. Potential metrics at stream sites are collected from both good and poor sites. Here, “good site” means a reference site, while “poor site” means a site with known stressors. The metric values are plotted against various stressor levels, e.g., chemical concentrations, habitat condition, and other measured stressors. In Maryland, metrics whose values differed between reference and degraded sites ($p < 0.05$) by both criteria were retained for further analysis, whereas those metrics having similar medians and distributions in reference and degraded sites were not considered in subsequent analyses. Multiple metrics should be selected to provide a strong and predictable relationship with stream conditions.

Step 4: Determine the best aggregation of core measures for indicating status and change in ecological condition

The index developed from the integrated core metrics is designed to provide a means of aggregating the information from various biological measures. It is called IBI, which stands for Index of Biotic Integrity. Metrics vary in their scales. They can be integers, percentages, and dimensionless numbers. To integrate all different units of metrics, standardizing the metric to unitless scores is applied, assuming that each metric is weighted equally. In Maryland, metric values above the 50th percentile of reference sites data were scored as 5, metric values between and including the 10th

and 50th percentiles were scored as 3, and all metric values below the 10th percentile were scored as 1.

$$\begin{cases} \text{if } > 50\% \text{ reference site data} & \text{index} = 5 \\ \text{if } \geq 10\% \text{ and } \leq 50\% \text{ reference site data} & \text{index} = 3 \\ \text{if } < 10\% \text{ reference site data} & \text{index} = 1 \end{cases} \quad (1)$$

Those metrics that increase in response to perturbation (reverse metrics) were scored such that values below the median received a score of 5, values between and including the 50th and 90th percentiles were scored as 3, and values above the 90th percentile were scored as 1.

Step 5: Index Thresholds for Assessment and Biocriteria

Impaired and non-impaired stream conditions are compared to determine the thresholds. For example, a Wyoming benthic IBI used the 25th percentile of the reference distribution to determine the threshold for good and fair conditions. Other condition ratings (very good, poor, very poor) resulted from equidistant sectioning of the remaining index range. Once the metrics and IBI standard are set up, bioassessments can be conducted.

Multivariate Analysis

Multivariate analysis is more commonly implemented in Europe and Australia and now in the western United States. The River InVertebrate Prediction And Classification System (RIVPACS) was developed for assessing the biological quality of rivers within the UK. The Benthic Assessment of Sediment (BEAST) was developed for assessing sediment quality in the Great Lakes of North America. And

the AUStralian RIVer Assessment System (AUSRIVAS) adapted RIVPACS concepts to the Australian environment. This study is based on the concept of the RIVPACS model, which is also the primary multivariate analysis concept for all three of the above models. It is described below as a RIVPACS-style model.

1.2.2 RIVPACS-Style Model

The River InVertebrate Prediction And Classification System (RIVPACS) was developed by the Natural Environment Research Council and the Department of the Environment of the UK in 1977 (Wright et al., 2000). It has been gradually improved and is now fully implemented in the UK and Australia. Between 1977 and 1981, the standard field procedures were formulated in the UK and site classification based on fauna was produced with 268 sites. Using multiple discriminant analysis (MDA), a strong link between biological and environmental features was demonstrated. After that, from 1981 to 1984, the geographical coverage of reference sites in the UK increased to a total of 370 sites. A procedure for predicting the probability of taxon occurrence was developed. In the UK, biological indices were developed by the Biological Monitoring Working Party (BMWP). These indices were incorporated into RIVPACS between 1984 and 1988 and the RIVPACS I model was tested by water industry biologists. RIVPACS was extended to deep lowland rivers and small streams and a new classification/prediction system was built with 438 sites in 25 classification groups. Using these additional sites and classification groups, the RIVPACS II model was used in the 1990 River Quality Survey in the UK. With the summarized results of the 1990 River Quality Survey, the RIVPACS II model was tested comprehensively from 1991 to 1995 using an independent data set of high quality sites. The number of

reference sites was increased to 614 for Great Britain and 70 for Northern Ireland. Meanwhile, alternative procedures for site classification and prediction of fauna were investigated using environmental attributes (Wright et al., 2000, Chapter 1). Finally, a new classification and prediction system for Great Britain and Northern Ireland based on the enlarged dataset (RIVPACS III) was developed and used in the 1995 General Quality Assessment Survey.

On the other side of the earth, the 1992 Monitoring River Health Initiative (MRHI) in Australia was charged with reporting on the ecological condition of Australia's rivers in response to the extensive blue-green algal blooms in the Darling River. The initiative focused on establishing a national river bioassessment system using invertebrates and RIVPACS was recommended (Wright et al., 2000, Chapter 8) for the project. However, Australia is recognized as one of the world's most arid continents. It has substantial water resources, but these are concentrated largely around the northern, eastern and southern coastal fringes. Thus, a RIVPACS-style model was adopted and adapted for Australia.

From the initial prototype to the now-comprehensive software, the main concept behind the RIVPACS-style model remained the same. First of all, it needed a large number of reference sites, which should be considered dense enough to represent the region. There are few rivers unaffected by human activity to be utilized as reference sites. Thus, "the reference site" was defined as the minimally impacted and best example of their type. It was checked by the water quality criteria, physical criteria, preferred habitats,...etc. After that, the model used biological data from reference sites to determine the biological similarity between sites. Then classification

regarding the biodiversity was used in order to minimize variance within reference site groups and maximize variance among reference site groups. The environmental features were related to the biodiversity between sites with multiple discriminate analyses (MDA). The environment features were those that may influence species occurrence but were not influenced by human activities or environmental stress (e.g., longitude, latitude, soil type, etc.). The multiple discriminate analyses (MDA) derived the linear combinations of environmental features that will discriminate best between groups of reference sites. In order to predict the taxa at a test site, Moss et al. (1987) derived the following procedures.

- (i) If the number of discriminant functions is f , use the discriminant function coefficients and the environmental features for that test site to calculate the discriminant scores (x_1, x_2, \dots, x_f) .
- (ii) Calculate the Euclidean distance from the site's scores to the mean score of each classification group.

$$d_j^2 = \sum_{i=1}^f (x_i - m_{i,j})^2 \quad (2)$$

Where d_j^2 = the square of the distance from site to group j and

$m_{i,j}$ = the mean of function i for group j .

- (iii) Check if the test site is within the range of the classification. If let d_{\min}^2 be the minimum value of d_j^2 . Then d_{\min}^2 is a chi-squared variable with f degree of freedom. If it is significant at the 1% level, it implies that there is less than 1% probability that a genuine member of the

classification group would give discriminant scores far from the group mean.

- (iv) The probability that the test site is a member of each group can be calculated as proportional to the number of members of that group multiplied by the exponential of minus one half the square of the distances from the group mean.

$$p_j = q_j / \left(\sum_{j=1}^{\text{total number of groups}} q_j \right) \quad (3)$$

Where p_j = the probability that the test site is a member of group j ,

$$q_j = n_j \times \exp(-d_j^2/2), \text{ and}$$

n_j = the number of members in group j .

- (v) The probability of species k occurring at the test site can be calculated as the sum over all groups of the probability that the site belongs to groups j multiplied by the proportion of members of group j at which species k occurs.

$$r_k = \sum_{j=1}^{\text{Total number of groups}} p_j \cdot g_{j,k} \quad (4)$$

Where r_k = the probability of occurrence of species k ,

p_j is obtained from procedure (iv), and

$g_{j,k}$ = the proportion of members of group j which contain species k .

Then the frequency of occurrence of taxa finally predicts the expected taxa probabilities (E) at the given test site. The observed taxa (O) from the test site can then be compared with the expected taxa probabilities as Observed / Expected (O/E).

As explained in equations 3 and 4, the prediction for species k at each test site was not only from one group but from multiple groups. The prediction was based on the proportion of members in a group and the probability that the test site belongs to that group. This approach was later adapted by the non-grouping methods as well.

1.2.3 Assessment by Nearest Neighbour Analysis (ANNA) Scheme

The widely accepted multivariate predictive models for bioassessment, such as RIVPACS and AUSRIVAS, involve classification of the reference site into groups. Then, appropriate environmental features are chosen as predictor variables for the models and a subset of environmentally similar sites is selected from which to predict the macroinvertebrate community in reference condition at a given test site. By classifying or clustering sites based on similarities in their biota, the above predictive models assume that macroinvertebrate communities occur in discrete groups (Linke et al., 2005). However, the classification step is somewhat artificial because an ecosystem is a continuum. Species assemblages are commonly described as changing continuously. Some information may be lost if it is segregated. And the less similar members in the group may also add noise to the information and affect later predictions. Furthermore, the results of different cluster schemes give different group sizes and different prediction results (Moss et al., 1999). Using one classification method seems to be arbitrary.

The Assessment by Nearest Neighbour Analysis (ANNA) scheme tries to remove the above concerns but the predictions are based on the same philosophy as RIVPACS and AUSRIVAS. The difference in ANNA is that the choice of an appropriate set of reference sites is determined by the nearest neighbor sites, based on environmental predictor variables that are selected in a manner similar to the way they are selected for RIVPACS-style models. It avoids the classification and discriminant function analyses when matching the test site with reference sites. ANNA finds the reference sites that most resemble the test sites in their values for environmental predictor variables. It then predicts community composition of these test sites based on the community composition of those nearest neighbors, thus treating the macroinvertebrate assemblage as a continuum instead of discrete groups. Once the probability of a taxon occurring has been estimated, O/E scores are computed in the same way as in other RIVPACS-style models. The detailed steps of ANNA are presented in Chapter Two, the model development section.

1.2.4 Region of Influence (ROI) Scheme

While biologists have been working on using the multivariate analysis methods for bioassessment, hydrologists have used multivariate analysis to solve their own problems as well. Hydrologists have been trying to estimate flood frequency at unmonitored stream sites. Similar multivariate analysis development processes have been tried and found effective in estimating flood frequency at unmonitored stream sites and predicting biodiversity in stream sites where no sampling has been done (Burn, 1990a, 1990b). Hydrologists have used the regional method to classify unmonitored stream sites for developing the estimate. They also tried to let the

unmonitored site partially belong to more than one region to do the estimate. This is very similar to the RIVPACS-style model in our biodiversity prediction problem. Finally, they developed the Region of Influence (ROI) method to estimate the flood frequency based on only those sites with physical similarities to the unmonitored site. It was suggested (Molgen, Personal Communication, 2007) that the Region Of Influence (ROI) method may be an appropriate approach to solve the ANNA problem. This interesting similarity deserves further study of the region of influence scheme in hydrology. It can offer us more different ideas on using only neighbor sites to predict biodiversity. This study then extends the nearest neighbor method to the region of influence method and finds the better application for bioassessment.

1.2.5 Maryland Biological Stream Survey (MBSS) Data

The Maryland Biological Stream Survey (MBSS) was initiated as a response to questions raised by the Maryland Synoptic Stream Survey of 1987. It was designed to take periodic snapshots of Maryland streams, identify best and worst areas, find out what caused them to become bad or stay healthy, and help target streams and watersheds for protection and restoration (Roth et al., 2001). The MBSS data provides information critical to managing our aquatic resources, including the Chesapeake Bay. It uses random sampling to determine the status of wadeable streams and rivers. The questions it tries to answer are distributed throughout the following fields: Physical Characteristics (e.g., flow depth and velocity.), Water Chemistry (e.g., pH and conductivity.), Biological Resources (e.g., fish abundance), Landscape Characteristics (e.g., shading and adjacent land cover.), Resource-stressor Associations (e.g., Dissolved Oxygen (DO).), and Resource-landscape associations

(e.g., continuousness of buffer.). The final goal of the MBSS is to provide environmental managers and policymakers with the information they need to make effective decisions. Accordingly, it also provides massive data for stream research (<http://www.dnr.state.md.us/streams/mbss/>). Since 1994, water samples, biological, and habitat data have been collected for over 1000 streams in Maryland. The parameters are listed in <http://www.dnr.state.md.us/streams/mbss/parameters.html>. The taxa biodiversity data are based on MBSS data from 1995-2002. Some specific on-site habitat information collected at the same time at sampling sites was also used as predictors for building the macroinvertebrate predictive model.

1.2.6 Stream Restoration

Stream restoration is a growing area within both the hydraulic engineering field and environmental science encompassing a wide range of activities. The publications in related topics increased five fold in the past two decades. But the definitions for terms associated with river restoration are evolving along with developments in the literature and a scientific consensus does not necessarily exist for all terminology (Shields et al., 2003). In general, the ecologist is concerned with modifying a degraded stream corridor to regain diversity or abundance of biological populations, while the engineer is concerned with producing systems or structures that meet certain criteria, usually those specified by the client. Nevertheless, transfer of research results into general guidance for practice is hindered by the absence of universally applicable indicators or control variables (Shields et al., 2003). Public and private groups have spent more than \$10 billion on more than 30,000 U.S. projects. Billions of dollars will flow to thousands of U.S. river restoration projects in the next few

decades (Malakoff, 2004). However, from the results of the National River Restoration Science Synthesis (NRRSS) database, only 10% of project records in the entire United States indicated any form of assessment or monitoring occurred, and only 6% of project records in the Chesapeake Bay watershed are monitored or assessed. Most projects didn't even adequately record the rudimentary information on project actions and outcomes. It is apparent that many opportunities to learn from successes and even from failures, and thus to improve future practice, are being lost (Bernhardt et al., 2005). More efforts are expected to gather and disseminate the data on restoration methods and outcomes. Standardized methods for pre- and post assessments should be developed for restoration practitioners and managers to understand what types of activities can achieve their goals. So far, some have been done by engineers, geologists, and ecologists (Moerke et al., 2004; Smith and Prestegard, 2005; Walsh et al., 2005). They try to bridge the gap between basic and applied lotic ecology, which is exactly what is needed to clearly describe and properly evaluate stream restoration. On the other hand, it will be even better if many cases can be synthesized by a standardized method instead of just case by case data collection.

1.3 OBJECTIVES

- 1 Build alternative macroinvertebrate prediction models using ANNA, BROI and NROI predictive schemes for Maryland
- 2 Examine and compare the results within alternative models and with previously built RIVPACS-style model

- 3 Evaluate the Montgomery County, Maryland stream restoration efforts applying both multivariate analysis and multimetric analysis

CHAPTER 2: REGION-STYLE MODEL DEVELOPMENT –

METHOD

This study looks into those macroinvertebrate predictive models which are built without a classification step and categorizes them as the Region-style models. With slight variations, these models have generally similar steps. The following are the general steps based mainly on the ANNA method that was developed by Linke et al. (2005). In this study, the concerns of the potential overfitting issue at Step 2 and different neighbor selection schemes at Step 4 are described after the five general steps.

2.1 THE FIVE GENERAL STEPS

Compared with the RIVPACS-style macroinvertebrate predictive model, the Region-Style predictive model is developed in five general steps: 1. Convert reference site biodiversity into multidimensional scaling. 2. Weight the predictor variables using a multivariate approach. 3. Calculate the weighted Euclidian distance from a test site to the reference sites based on the environmental predictors. 4. Predict the faunal composition based on the nearest reference sites and predictors. 5. Calculate an Observed/Expected (O/E) ratio analogous to RIVPACS-style model for validation. Detailed descriptions of each step follow.

Step1. Convert reference site biodiversity into multidimensional scaling.

In RIVPACS-style models, reference sites were first clustered based on reference site biodiversity, and then discriminate function analysis was used to

identify predictor variables specific to each cluster of sites that would separate the clusters based on biodiversity. In Region-style models that have no classification step, a Bray-Curtis distance matrix was calculated between reference sites based on the presence/absence of taxa, which is similar to what the RIVPACS-style model does prior to the classification step. Then Non-metric Multidimensional Scaling (NMDS) was used to ordinate these reference sites based on the previous Bray-Curtis distance matrix and determine the number of dimensions are needed to explain these proximities. The detailed explanation of NMDS and the Bray-Curtis distance are in the following.

Non-metric Multidimensional Scaling (NMDS)

Multidimensional scaling is a set of mathematical techniques that enable a researcher to uncover the “hidden structure” of data bases (Kruskal and Wish, 1978). Suppose there is a map showing the locations of several cities in the United States. You are asked to construct a table of distances between these cities. It is a simple matter of just measuring the distance between the cities. Now, consider the reverse problem, where you are given the table of distances between the cities, your task is to rebuild the map. For this task, geometric procedures can solve this problem, but more effort will be needed. In essence, multidimensional scaling (MDS) is a method for solving this reverse problem. If we extend it to a more complicated problem, multidimensional scaling can use proximities among any set of objects as input distances. A proximity is a number which indicates how similar or how different two objects are. The main output is a spatial representation, consisting of a geometric configuration of points on a map. Each point in the configuration corresponds to one

of the objects. This configuration reflects the “hidden structure” in the data. The larger the dissimilarity (or the smaller the similarity) between the two objects, as shown by their proximity value, the further apart they should be in the spatial map. As for nonmetric multidimensional scaling (NMDS), suppose there are n objects with dissimilarities δ_{rs} . The procedure is to find a configuration of n points in a space, which is usually chosen to be Euclidean, so that each object is represented by a point in space. A configuration is sought so that distances between pairs of points d_{rs} in the space match “as well as possible” the original dissimilarities δ_{rs} (Cox and Cox, 2001).

Bray-Curtis Distance

In the case here, a Bray-Curtis distance matrix is used as the proximity. Bray Curtis distance, sometimes also called Sorensen distance, is a normalization method that is common in botany, ecology, and environmental science fields to measure the dissimilarity. Based on the presence/absence of taxa, the Bray-Curtis distance matrix was calculated between sites.

$$\text{Bray-Curtis distance} = 1 - (2W / (A+B)) \quad (5)$$

Where W is the sum of shared abundances, A and B are the sums of abundances of taxa found in each site (Bray and Curtis, 1957). The Bray-Curtis distance values range from 0 to 1. The closer the value is to 0, the more similar the biodiversity is between these two sites.

Using the Bray-Curtis distance formula builds the Bray-Curtis distance matrix between each site. Additionally, the Bray-Curtis distance matrix can be calculated

based on two different taxon levels – Family or Genus. The biodiversity distances between each site, just like the previous distance table between cities, are established. Then NMDS will be used to ordinate the sites in dimensions (Clarke, 1993). The dimensions here refer to the number of coordinate axes (axis 1, axis 2, axis 3, ...etc) in space. The values on the coordinate axes locate a site.

Goodness-of-fit is a very important consideration in deciding how many dimensions are appropriate. A measure of fit widely used in MDS is “stress,” which is the square root of normalized “residual sum of squares.” But, it is actually the larger values that indicate worse fit; it would be more descriptive to call stress a “Badness-of-fit” measure (Kruskal and Wish, 1978). Thus, here 25% was used as the upper limit of stress to find sufficient dimensions (Kruskal, 1964).

Therefore, the Bray-Curtis distance provides the numerical relationship between reference sites for NMDS in order to ordinate the biodiversity closeness in multiple dimensions.

Step 2. Weight the predictor variables using a multivariate approach

Environmental parameters, e.g., watershed area, soil types, flow depth...etc, which serve as potential predictors are also collected for the same reference sites. This step uses multiple regression to associate the potential predictors with the values in coordinates that are obtained from the previous NMDS step. The goal is to find the most useful multiple regression functions for later prediction. Stepwise multiple regression was used in previous literature (Linke et al., 2005). When using multiple

regressions, however, there is one concern that needs to be alleviated: the overfitting issue. This issue will be addressed later.

Step 3. Calculate the weighted Euclidian distance from a test site to the reference sites based on the environmental predictors.

Once the multiple regression functions are developed, the final distance between two sites can be calculated in multivariate space.

The modified Euclidean distance (d) between two sites i and j is calculated as:

$$d_{ij} = \sqrt{\sum_{k=1}^{\alpha} (q_{i,k} - q_{j,k})^2} \quad (6)$$

where α = the dimension resulting from Step 1 above and

$$q_{i,k} = \sum_{m=1}^M a_{km} X_{im} = \text{the predicted value for site } i \text{ on the NMDS axis } k \quad (7)$$

where a_{km} = the multiple regression coefficient for the environmental variable m on the NMDS axis k ; X_{im} = the value of the environmental variable m for site i .

Now the Euclidean distances between sites in each NMDS dimension are ready.

Step 4. Predict the faunal composition based on the nearest reference sites and predictors.

Based on different literature, Euclidean distances have been used differently in the weighting function (WF) for the Region-style prediction. ANNA, Burn's ROI, and Tasker's ROI predictive schemes were studied and a new weighting function ROI predictive scheme was developed in this study. All the above predictive schemes are described in more detail in Section 2.3. Once the weighting functions are obtained, the taxon composition can be predicted. For each taxon, the probability of a taxon occurring (p) was estimated as

$$p = \frac{\sum_{i=1}^n x_i \cdot WF_i}{\sum_{i=1}^n WF_i} \quad (8)$$

Where n = the number of reference sites,

$x_i = 0$ for absence at reference site i ,

$x_i = 1$ for presence at reference site i , and

WF_i = the distance weighting function from each reference site i .

The equation (8) is the alternative from equation (3) and (4).

Step 5. Calculate an observed/expected (O/E) ratio analogous to RIVPACS/AUSRIVAS

For any given site, each taxon occurrence probability can now be predicted. The predicted probability can range from 0 ~ 1. Zero means absence and one means occurrence. Assuming a fairly precise predictive model, the taxon with occurrence as

0.5 will have a better chance to be caught in the sample than those with 0.01 and 0.001 probabilities. Wright (2000) has defined a capture probability, p_c . They studied the result with $p_c = 0$, $p_c = 0.01$ and $p_c = 0.5$. Those species with $p > 0.5$ were referred to as “common” species. If the probability p is small (< 0.5), it seems to be hard to catch this taxon in the sample. For p larger than 0.5, there are better chances of capture in the observed sample (O). The use of $p_c = 0.5$ as cutoff line seems arbitrary. It does not mean that taxon with probability as 0.49 is much less common species than the probability as 0.5. The use of capture probability depends on the reference data quality used to develop the model. If the reference data are highly reliable, the p_c can be lower without adding more noise to the prediction. The use of $p_c = 0.5$ is thus not definite. In the literature, the predictive models built using $p_c = 0.5$ have better precision. All the later models used the capture probability as $p_c = 0.5$ to develop the model. This is also applied to the Region-style models here, so they can be compared to other models.

One index that concerns us when using these macroinvertebrate predictive models is the Observed/Expected ratio (O/E). Expected (E) is the sum of the taxa occurrence probability. And Observed data (O) can “almost” be considered as the sum of the taxa occurrence (presence =1, and absence = 0) in the sample site. The reason for “almost” is that not all taxa will be included in the O/E calculation. At each test site, the predicted occurrence probability of each taxon is considered. If the occurrence probability (p) of taxa is larger than p_c , then this probability p is

include to E and the presence or absence (1/0) is included in the calculation of O. The O/E ratio is then calculated at each given test site.

If an observed site is a reference site, the O/E ratio is supposed to be close to 1. Thus, O/E measures from the calibration and verification reference sites can be used to evaluate the prediction accuracy of the prediction model.

2.2 FROM STEP 2 – THE POTENTIAL OVERFITTING ISSUE

Previously in RIVPACS-style models, selecting the predictor to be included in the discriminate functions has been an important issue. Some literature suggested using stepwise discriminate function analysis to select the predictors. However, Van Sickle et al. (2006) showed the potential overfitting issue caused by this “stepwise” approach in the RIVPACS-style model. In the Region-style models, the stepwise method was one of the selection methods used to include the predictors in the multiple regression functions. The nearest neighbor method (Linke et al., 2005) used stepwise multiple regression with an entry- and removal criterion of $r^2 = 0.1$ to select the environmental predictors. In Region-style models, the overfitting issue is expected to exist in the models as well if the stepwise method for acquiring predictors is used to form the regression functions.

Overfitting means that some of the relationships that appear statistically significant are actually just noise. It occurs when the complexity of the statistical model is too great for the amount of data that you have. The result is that a model with overfitting does not replicate well and performs poorly when predicting future responses. In our multiple regression case, the overfitted model can be explained as

being so tuned to the calibration data set which is used to build the model that it performs badly when applied to the validation data.

Overfitting can occur in generalized linear models if either or both of the following conditions hold (Sarle, 2007):

1. The number of input variables is large with respect to the number of training cases. Typically it is preferred that there are at least 10 times as many training cases as input variables (Tanaka, 1987), but with noise-free targets, twice as many training cases as input variables would be adequate.

2. The input variables are highly correlated with each other. This condition is called "multicollinearity" in statistical literature. In multicollinearity, the collinear variables contain the same information about the dependent variable. The slope coefficients of those variables tend to be extremely large. The estimate of one variable's impact on the dependent variable tends to be less precise than if predictors were uncorrelated with one another.

Pairwise collinearity can be determined from viewing a correlation matrix of the independent variables. However, correlation matrices will not reveal higher order collinearity.

Van Sickle et al. (2006) used cross-validation techniques to detect overfitting in MDA. They had 14 candidate predictors in their model and they searched all the subsets' predictors for building models. From the Root Mean Square Error (RMSE) trend in calibrating and validating data along with the cross-validation technique, they have proven the potential for overfitting from using stepwise discriminant function

analysis. In this no-classification step study, the model development relies mainly on the multiple regression function to describe the reference site difference. Due to the NMDS step, there will be more than one multiple regression function used to predict the taxa occurrence. There is no evidence yet to prove that the model with better regression functions will give us better prediction results. The previously mentioned multicollinearity usually won't affect the prediction result. But since the prediction here is the results of combining several multiple regression functions, the overfitting effect from each function may accumulate increasing noise in the prediction. It will be interesting to find out and compare with the overfitting model if a not-overfitted model will result in a better prediction.

Cross-validation was first attempted to eliminate the overfitting and then different methods were studied and tried to eliminate the cause of multicollinearity in this study. Then rational multiple regression function checking (McCuen and Snyder, 1985) was used to improve the models. The detail checking procedures will be described later in its own section. During the checking process, highly correlated and irrational correlation sign predictors were removed, thereby reducing the input candidate predictors at the same time. The causes of overfitting can then be alleviated.

Besides, there is more than one method to choose predictors. Since the potential multicollinear predictors were reduced from the rational multiple regression function checking, the stepwise method, Stepwise AIC, was still kept to select the predictors. Additionally two other searching methods, exhaustive AIC and exhaustive BIC, were also added as options for the predictor selection. The difference between

selection methods will be described in later sections. The cross-validation technique to alleviate the overfitting problem is discussed below.

2.2.1 Cross-Validation

Van Sickle et al. (2006) applied cross validation (leave-one-out) to see the correct percentile of assigning reference sites to the classification group. That shows that when number of predictors in the discriminant functions goes up to a certain order (referred to as “model order” in the literature), the correct percentile stays the same, which implies the higher order models may be overfitting models. In their paper, there is no significant evidence that a model with higher classification accuracy has a smaller RMSE (better prediction) in the validation sites. This is because the prediction at the test site does not rely solely on the group that the test site belongs to, but is the result of being combined with other close reference groups. The biodiversity prediction concerns more than just the classification accuracy. The following is the cross validation definition:

Cross-validation – sometimes called rotation estimation, is the statistical practice of partitioning a sample of data into subsets such that the analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis.

The same leave-one-out cross-validation procedure (as Van Sickle et al. (2006)) was attempted to detect the overfitting in this study. However, it is easier for them to perform leave-one-out cross validation with 14 candidate predictors in total.

The model in this study not only has more predictors (23 and 30 parameters in different study cases), but the model also has more dimensions (3 dimensions) after NMDS (Section 2.1.Step1). So the combination is enormous ($(2^{30} = 1073741824) \times 3$ dimensions). Another difficulty is that some of these Region-style models have more than three parameters in weighting functions. It will be hard to investigate all the model performance.

The alternative method adopted here used the “leaps” library (Lumley, 2004) in R (R Development Core Team, 2007). With the “leaps” library, all the subset predictors in each model order were searched and their Residual Sum of Squares (RSS) were calculated. The RSS is the discrepancy between the observed data and the prediction. The smaller this discrepancy is, the better the estimation is. The five models with the smallest RSS in each model order were selected and applied in each dimension (coordinate). The best five RSS models in each order were then used to perform the leave-one-out cross-validation. There was no classification in Region-style models, so the observed sites were not assigned to a classification group to determine the correct percentile. The estimated prediction error instead of classification accuracy was used as a goodness-of-fit measurement. It was expected that the estimated prediction error would decrease as the model order increased. When the decreasing trend stops at a certain model order, it means the models with model order higher than that are most likely overfitting.

The RSS of each model was also used to calculate the Akaike’s Information Criterion (AIC) and Bayesian Information Criterion (BIC). Then the best result of

AIC and BIC in each order can be selected. There will be more detailed explanation for AIC and BIC later.

2.2.2 Rational Multiple Regression Function Checking

While searching for ways to alleviate the overfitting issue, one additional index was suggested that detects the important predictor as well as the irrational (overfitting) predictor from the regression analysis point of view. Assume our multiple regression of the biodiversity coordinate is

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots b_mX_m \quad (9)$$

Where Y is the biodiversity coordinate after NMDS,

X_1, X_2, \dots, X_m are the predictors,

b_0 is the intercept value of regression function, and

b_1, b_2, \dots, b_m are the slope coefficients for each predictor.

In order to check the rationality of the magnitude of a slope coefficient, since the slope coefficients are a function of the units of both Y and X_i , it is easier to interpret the rationality of the coefficient by converting a b_i value to the corresponding value of the standardized partial regression coefficient, t_i (McCuen and Snyder, 1985):

$$t_i = \frac{b_i S_{X_i}}{S_Y} \quad (10)$$

where S_{X_i} is the standard deviation of predictor X_i , and

S_Y is the standard deviation of coordinate Y .

The t_i values are dimensionless and independent of the units of Y and X_i , thus they are easier to interpret. A t_i value should be between 0 and 1 in absolute value. A value near 1 indicates an important predictor; a value near 0 indicates that the corresponding predictor variable is not important. Thus the larger the value, the more important a predictor variable is considered. Where the intercorrelations between the predictor variables are significant, the t_i values may exceed 1. This should be considered irrational because the t_i values are essentially correlations in the domain of standardized variables. If a t_i value is irrational, the corresponding b_i value should be considered irrational.

The slope coefficients sign should be rational as well. Since the slope coefficients represent the change in Y for a given change in the value of the predictor variable X_i , it is easy to check for rationality in the sign. If Y and X_i have a positive correlation, the b_i should be positive; if Y and X_i have a negative correlation, the b_i should be negative.

Based on the above, in order to avoid having the multiple regression function be irrational, the predictors were eliminated from the initial predictor set if the slope of the coefficient of regression function shows that the absolute t_i value is larger than 1 and the coefficient sign does not correspond to the correlation sign for coordinate Y . In this study, because the multiple regression is applied to each axis respectively, the predictor elimination is done for each axis as well. After eliminating the irrational elements, different initial predictor sets will be retained for later predictor selection.

2.2.3 Stepwise AIC method

For convenience, Stepwise AIC, the existing function in R (R Development Core Team, 2007), is used in this study to perform stepwise multiple regression step. The initial sets of potential predictors were obtained after rational multiple regression function checking. Then, at each step new predictor sets are formed by sequentially adding or dropping variables. The Akaike Information Criterion (AIC) is then computed for each new set, and the set that reduces AIC by the greatest amount is retained (Hastie, 1992). This stepwise procedure is continued until no improvements in AIC occur. More detail about the definition of AIC follows.

2.2.4 Exhaustive AIC and Exhaustive BIC

The Exhaustive AIC and Exhaustive BIC methods were presented by Zhu and Chipman (2006). They looked into different variable selection methods and compared their results between Stepwise AIC, Stepwise BIC, Exhaustive AIC, Exhaustive BIC and GA. The terms “Stepwise AIC” and “Stepwise BIC” refer to a stepwise search combining both forward selection and backward elimination using either the AIC (Akaike, 1973) or the BIC (Schwarz, 1978) as the objective function. The terms “Exhaustive AIC” and “Exhaustive BIC” refer to evaluating all possible subsets with either the AIC or the BIC. An exhaustive search over all possible subsets is feasible here using the leaps library (Lumley, 2004) in R (R Development Core Team, 2007). Their main research target was GA, the genetic algorithm. They have a true model with only three variables. Using the above selection methods, Exhaustive BIC gave the closest four variable selections out of 20 candidate variables, Exhaustive AIC picked seven variables, and Stepwise AIC selected up to 12 variables. It seems that

the Exhaustive BIC method was preferable. However, when I contacted Dr. Zu for more information, he told me that BIC is generally better than AIC when you are focusing on which variables are selected. When prediction error is the focus, then AIC gives the better prediction than BIC. It is expected that Region-style models have the fitted regression functions but also these functions are used for predictions. So besides the previously mentioned Stepwise AIC selection method, both Exhaustive AIC and Exhaustive BIC are applied in this study to look into the model results.

Akaike's Information Criterion (AIC) was first proposed in 1973 to estimate the Kullback-Leibler discrepancy (Akaike, 1973). In his derivation, he assumed that the true model belongs to the set of candidate models.

$$AIC = -2\ln(\text{likelihood}) + 2 \times \text{number of parameters} \quad (11)$$

where the *likelihood* is usually evaluated at the estimated parameters.

Assume that the model errors are normally and independently distributed and use the maximum likelihood estimates, then:

$$-2\ln(\text{likelihood}) = n\ln(2\pi) + n\ln(\hat{\sigma}^2) + n \quad (12)$$

where $\hat{\sigma}^2 = \frac{\sum \hat{\epsilon}_i^2}{n}$ is the maximum likelihood estimate of σ^2 ,

$\hat{\epsilon}_i$ are the estimated residuals for a particular candidate model, and

n is sample size.

Define the *number of parameters* in the equation as K , which is the total number of estimated regression parameters, including the intercept and σ^2 . Then the equation can be rewritten as

$$AIC = n \ln(2\pi) + n \ln(\hat{\sigma}^2) + n + 2 \times K \quad (13)$$

The constants $n \ln(2\pi) + n$ play no practical role in the model selection and can be ignored. Now,

$$AIC = n \cdot \ln(\hat{\sigma}^2) + 2K \quad (14)$$

However, AIC as derived by Akaike may perform poorly if there are too many parameters in relation to the size of the sample. Sugiura (1978) derived a second-order variant of AIC that he called c-AIC. Hurvich and Tsai (1989) further studied this small-sample bias adjustment criterion as AIC_c .

$$AIC_c = -2 \ln(\text{likelihood}) + 2K \left(\frac{n}{n - K - 1} \right) \quad (15)$$

where the penalty term is multiplied by the correction factor $n/(n - K - 1)$.

This can be rewritten as

$$AIC_c = -2 \ln(\text{likelihood}) + 2K + \frac{2K(K+1)}{n - K - 1} \quad (16)$$

or

$$AIC_c = AIC + \frac{2K(K+1)}{n - K - 1} \quad (17)$$

The second term in equation 17 will be referred to as a penalty term.

Since AIC_c converges to AIC as n gets large, AIC_c should be employed regardless of sample size. Both McQuarrie and Tsai (1998) and Burnhan and Anderson (2002) recommend the use of AIC_c . Thus, AIC_c was used in this study for the model predictor selection.

For Bayesian Information Criterion (BIC) also known as Schwarz Information Criterion (SIC), Schwarz's Bayesian Criterion (SBC) (Schwarz, 1978), it is defined:

$$BIC = -2 \ln(\text{likelihood}) + K \cdot \ln(n), \quad (18)$$

where K is the number of free parameters to be estimated and

n is the sample size.

Just as with AIC, it can be derived with the maximum likelihood estimates:

$$BIC = n \cdot \ln(\hat{\sigma}^2) + K \cdot \ln(n) \quad (19)$$

The maximum likelihood estimate σ^2 in both AIC (eq. 14), BIC (eq.19) can be derived as:

$$\hat{\sigma}^2 = \frac{\sum \hat{\epsilon}_i^2}{n} \quad (20)$$

The numerator here is actually the Residual Sum of Squares ($RSS = \sum \hat{\epsilon}_i^2$).

So both the AIC_c and BIC can be calculated from the Residual Sum of Squares (RSS).

$$AIC_c = n \cdot \ln\left(\frac{RSS}{n}\right) + \frac{2 \cdot n \cdot K}{n - K - 1}, \quad (21)$$

$$BIC = n \cdot \ln\left(\frac{RSS}{n}\right) + K \cdot \ln(n), \quad (22)$$

The second term in eq 22 will be referred to as a penalty term.

Both AIC_c and BIC are an increasing function of RSS and an increasing function of K . That is, unexplained variations and the number of explanatory variables increase the criterion values. Increasing the number of free parameters to be estimated improves the goodness of fit and lowers the RSS in the first term, while it also increases the total criterion value in the second term. Hence AIC_c and BIC not only reward goodness of fit, but also include a penalty that is an increasing function of the number of estimated parameters. This penalty discourages overfitting. The preferred model is the one with the lowest AIC_c or BIC value. The BIC penalizes free parameters more strongly than does the AIC_c .

2.3 FROM STEP 4 – DIFFERENT SCHEMES FOR SELECTION REGION

The goal for this study was to find an alternative method to build the taxa predictive model. There is already an alternative ANNA scheme for Region-style biodiversity prediction. Studying a similar case in hydrology gave us more different concepts to build the Region-style predictive models. The following describes the predictive schemes' origins and differences.

2.3.1 Assessment by Nearest Neighbour Analysis (ANNA) Scheme

The ANNA scheme uses the square root of the reciprocal environmental distance to weight the contribution. The weighting function can be expressed as:

$$WF = \frac{1}{\sqrt{d_{ij}}}, \quad (23)$$

where d_{ij} is the Euclidean distance between test site i and reference site j that were derived from general step 3 (Section 2.1.Step3).

The number of nearest neighbors that is included as reference sites for each test site can later be determined by iteration. Notice that sites with identical environmental characteristics to a test site (i.e. $d_{ij} = 0$) were removed from the analysis, because the weight of such a site would be infinite resulting in a relative weight of 1. This creates a disadvantage for the ANNA scheme when predicting sites that have reference sites with exactly the same physical parameter, which normally refers to the same location. The same location reference sites, which would provide exact prediction information, are taken out due to this limitation.

2.3.2 Burn's Region of Influence (BROI) Scheme

Hydrologists have been faced with the task of estimating the probability of exceedance associated with a selected flow value at a particular location on a river. Many researchers have explored various ways of delineating a set of monitoring stations that have sufficient homogeneity in extreme flow characteristics to be considered a region. Identification of homogeneous regions will often lead to an effective and efficient spatial transfer of information, while there are still problems in the cases where a station lies on the boundary between two homogeneous regions. It has been suggested by Acreman and Wiltshire (1987) that such a station could be regarded as being a partial member in both regions. Burn (1990a; 1990b) expanded

this concept leading to the realization that there is no need to define boundaries between regions but rather each site can have its own “region” consisting of those stations that are sufficiently similar to the site of interest. He explored three different strategies to define the Region of Influence (ROI).

Option 1

$$\begin{cases} \theta_i = \theta_L & \text{for } NS_i \geq NST \\ \theta_i = \theta_L + (\theta_U - \theta_L) \left(\frac{NST - NS_i}{NST} \right) & \text{for } NS_i < NST \end{cases}, \quad (24)$$

where θ_L is a lower threshold value defining a desired proximity for stations to be included in the ROI for site i , NS_i is the number of stations in the ROI for site i with the threshold at θ_L , NST is the target number of stations for a region of influence, and θ_U is an upper threshold value for sites with fewer than NST stations in the ROI.

The philosophy behind this option is that all stations within a distance corresponding to θ_L should be included in the ROI for the unmonitored site, and if the number of stations included is less than the desired number, NST , a less restrictive threshold should apply. The threshold for the latter case eventually is a function of an absolute upper boundary for the threshold, θ_U , and the number of stations with distance measures less than the lower threshold, as expressed above. The weighting function for this option is defined through

$$WF_{ij} = 1 - \left(\frac{d_{ij}}{TP} \right)^n, \quad (25)$$

where d_{ij} is the distance, Since it is adapted in the biodiversity prediction.

d_{ij} is the previously defined distance between an unmonitored site and reference site in general step 3 (Section 2.1.Step3). TP and n are parameters of the weighting function.

Option 1 requires the definition of an upper and lower threshold value, θ_U, θ_L , a target number of stations, NST , for the region of influence, and the parameters of the weighting function, TP, n . There are total of five parameters.

Option 2

This option has a constant threshold value which is given as

$$\theta_i = \theta_U, \quad (26)$$

where θ_U is the constant threshold value. The weighting function for this option is given as

$$\begin{cases} \eta_{ij} = 1 - \left(\frac{d_{ij} - \theta_L}{TN - \theta_L} \right)^n & \text{for } d_{ij} > \theta_L \\ \eta_{ij} = 1 & \text{for } d_{ij} \leq \theta_L \end{cases} \quad (27)$$

Where θ_L defines a lower threshold for the distance metric, and TN and n are parameters of the weighting function with TN defined through

$$TN = \text{Max}(TL_i, TPP) \quad (28)$$

With TPP a parameter of the weighting function, and TL_i given as

$$TL_i = \text{Max}_{\{j\}}(d_{ij}) \quad (29)$$

The information requirements for this option are the two threshold values, θ_L, θ_U , and the two parameter values of the weighting function, TPP, n .

Option 2 differs from option 1 by giving those stations larger than θ_L smaller weighting and keeping those smaller than θ_L a weighting of 1.

Option 3

The final option includes all sites in the region of influence such that the threshold is defined as

$$\theta_i = TL_i \quad (30)$$

The weighting function is defined in equations (27) – (29). This option requires the specification of a threshold parameter, θ_L , and two weighting function parameters, TPP, n .

In the preliminary study, Option 2 and Option 3 did not give better results when compared to Option 1 in our biodiversity prediction application. This may have been because Option 2 and Option 3 still include most of the monitoring stations, which bring more noise into the prediction results. Only option 1 was adapted for our study. It is referred to as the BROI scheme in later sections.

2.3.3 New Datum Region of Influence (NROI) Scheme

During the study, I developed another option for the region of influence method that is similar to Option 1.

$$\begin{cases} \theta_i = \theta_L & \text{for } NS_i \geq NST \\ \theta_i = \theta_L + (\theta_U - \theta_L) \left(\frac{NST - NS_i}{NST} \right) & \text{for } NS_i < NST \end{cases} \quad (31)$$

Where θ_L is a lower threshold value defining a desired proximity for stations to be included in the ROI for site i , NS_i is the number of stations in the region of influence for site i with the threshold at θ_L , NST is the target number of stations for a region of influence, and θ_U is an upper threshold value for sites with fewer than NST stations in the region of influence. The weighting function for this new scheme is

$$WF_{ij} = 1 - \left(\frac{d_{ij}}{\theta_i} \right)^n \quad (32)$$

The difference between this method and BROI option 1 is to use θ_i that was defined in equation (31) as the denominator of the weighting function. This means that each test station uses the most remote influence point as the new datum. Every test site that use BROI scheme share the same datum, TP, as region of influence. Each test station that use NROI scheme has its own datum. From the predictive point of view, it emphasizes that each test site has its own region of influence.

2.3.4 Tasker's Region of Influence (TROI) Scheme

Region of influence has evolved along different paths. In order to estimate stream flow characteristics, Tasker and Slade (1994) developed an interactive regional regression approach to estimate flood quantiles. A computer program has been developed which will estimate flood quantiles for an unmonitored site based on

data from monitoring stations with similar watershed characteristics. The user enters site location and watershed characteristics for an unmonitored site and the program selects, from a data base of monitoring stations, a subset of stations to be used in the regression analysis. The subset of stations is selected based on the similarity of their basin characteristics to the unmonitored site's basin characteristics.

There are differences between Burn's ROI and Tasker's ROI. Burn used weighting functions to define the region of influence and the weighting information from the monitoring station. Then regression analysis was used in the entire region. Tasker used the similarity of basin characteristics to decide the region of influence. Only the n most similar monitoring stations (e.g., $n = 20$) were used to conduct regression analysis and develop the unique regression function for each unmonitored station. How many monitoring stations will be needed when developing the function will be studied through iteration. When the region of influence method were applied to this biodiversity study, both Burn's and Tasker's ROI were used. However, some problems occurred during Tasker's ROI application. After picking out the reference sites with physical characteristic similarity, stepwise regression were used to find the regression functions. The regression result showed that either no candidate physical parameter is significant enough as a predictor or only intercept was obtained from the regression. The application of Tasker's ROI was discontinued and thus will not appear in this study. It was determined that this problem was caused by too few reference site data and too many predictors to perform the regression analysis, since three regression functions were needed for each test site. Tasker's ROI concept is

really unique. It would be an interesting future study can adapt this concept to predict biodiversity in the future.

CHAPTER 3: REGION-STYLE MODELS FOR MARYLAND –

METHOD

Predicting biodiversity is different from predicting stream flow characteristics. Site biodiversity relies on biological sampling and the laboratory work to identify the macroinvertebrate taxa. Thanks to the Maryland Biological Stream Survey (MBSS) program, Maryland has abundant biomonitoring data, which is sampled following the EPA rapid bioassessment protocols. Along with the biodiversity data base from the MBSS, environmental characteristic parameters (predictor variables) are necessary to develop the predictive model.

3.1 BIODIVERSITY DATA

The biodiversity data are from the MBSS for years 1995 to 2002. There are 1629 sampling sites. Both biodiversity data and environmental features were collected at each site except for some that were missing environmental data. The macroinvertebrate biodiversity data can be characterized either at the family or the genus level. In our data source, there are 124 taxa at the family level and 231 taxa at the genus level in the Maryland data. According to the previously mentioned 12 reference criteria, 158 sites of the total are reference sites. The reference site locations are shown in Figure 2. Twenty-nine sites were randomly selected and withheld for validation data. The remaining 129 reference sites became calibration data to build the macroinvertebrate predictive model. However, some of the environmental feature data were missing in the reference sites data. Only 124 reference sites were used for

calibration data, if environmental parameters from MBSS were used as candidate for the model predictors.

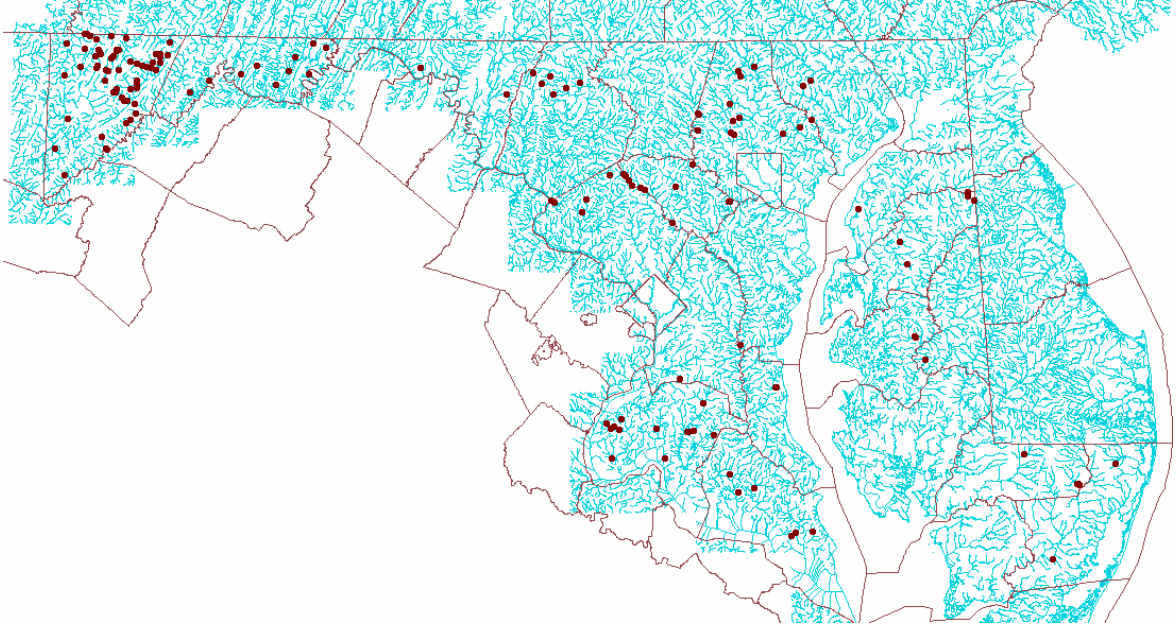


Figure 2. 158 reference sites locations in Maryland

3.2 CANDIDATE PREDICTORS LIST

Potential predictors can be obtained from two categories. For Maryland, the potential environmental predictors can be generated from GISHydro2000 or from the MBSS data set or both. The benefit of reliance on map-based predictors is that in the future the taxa prediction for any site alone, without the need for other field data is possible. On the other hand, developing a model using only map-based predictors might result in suboptimal performance because instream site-level attributes such as local habitat were not included in the list of candidates (Van Sickle et al., 2006). The candidate predictors obtained in this study are summarized in Table 2. There are 27 potential predictors obtained from GISHydro2000, the map-based predictors. Another

seven potential predictors were obtained from the on-site MBSS environmental characteristic data.

Table 2. Summary table for candidate predictors

Source	Predictor Variables Generated	Definition	Transformation
GISHydro2000	AREA	Watershed Area (mi ²)	Log ₁₀ (x+1)
	WSLOPE	Watershed Slope (ft/ft)	\sqrt{x}
	CSLOPE	Channel Slope (ft/mi)	Log ₁₀ (x+1)
	RELIEF	Basin Relief (ft) ¹	Log ₁₀ (x+1)
	LIME	Percent Limestone	Presence/Absence
	HUC14	14 digit hydrologic unit code, use only the last 6 digits of 14 digit code	
	HUC8	8 digit hydrologic unit code	
	PERIM	Watershed Perimeter Length (mi)	Log ₁₀ (x+1)
	LENGTH	Main Channel Length (mi)	Log ₁₀ (x+1)
	SSA, SSB, SSC, SSD	Percent Watershed as Hydrologic Soil Type A, B, C, D from STATSGO layer (%)	SSA and SSD use Arcsine(\sqrt{x})
	RSA, RSB, RSC, RSD	Percent Watershed as Hydrologic Soil Type A, B, C, D from Ragan soil layer (%)	RSA and RSD use Arcsine(\sqrt{x})
	ELEV	Sampling Point Elevation (m)	Log ₁₀ (x+1)
	SIN	Channel Sinuosity ²	
	P2	2-Year Precipitation (in)	
	P100	100-Year Precipitation (in)	
	HIGHELEV	Percent Watershed Above 2000 feet	Arcsine(\sqrt{x})
	HYPISO	Hypsometric Area Ratio ³	
	EASTING, X	Coordinate (m, MD StatePlane, NAD 1983) of sampling site	
	NORTHING, Y	Coordinate (m, MD StatePlane, NAD 1983) of sampling site (outlet)	
	LAT	Coordinate of sampling site (decimal degrees)	
LONG	Coordinate of sampling site (decimal degrees)		
MBSS	DATE_SPR	Actual Date Sampled - Spring	
	JDAY	Calculate Julian Period day	
	BOULDLT2	Presence/Absence	
	COBBLE	Presence/Absence	
	BEDROCK	Presence/Absence	
	MAXDEPTH	Maximum stream depth (cm)	Log ₁₀ (x)
	AVGWID	Average Wetted Width (m)	Log ₁₀ (x)

1 Basin Relief is the difference between the mean and outlet elevations of the drainage basin.

2 Channel sinuosity (P) is calculated as the ratio Lc/L, where Lc is channel length and L is straight-line valley length over a chosen distance.

3 The percentage hypsometric curve (area-altitude curve) relates horizontal cross-sectional area of a drainage basin to relative elevation above basin mouth (Bras, 1990). Hypsometric Area Ratio is calculated as the area below the percentage hypsometric curve.

3.2.1 Data Transformation

The potential predictors were intended to explain the variance of biodiversity between sites. In order to make the best possible explanation, each potential predictor variable histograms was drawn to preserve their original distributions (Appendix A-1). For those parameters whose distributions skewed to the right, different transformations were used to normalize the population distribution. The choices were log, 4th-root, square root, and no transformation (Clarke and Ainsworth, 1993). Earlier research (Moglen and Paul, 2005; Wright et al., 1984) has also utilized $\text{Log}_{10}(x+1)$, $\text{Log}_{10}(x+1)$ and $\text{Arcsine}(\sqrt{x})$ transformations. If the transformed data had a high correlation with NMDS values in axes space, these transformed data were better able to explain the biodiversity and were used as predictors (Appendix A-2). The transformation for each parameter is in Table 2.

The previous RIVPACS-style model for Maryland (Moglen and Paul, 2005) used transformed data in discriminate function analysis. Preliminary research demonstrated there were only small differences between using non-transformed and transformed data as predictors. The model prediction from transformed data was not necessarily better. For convenience, non-transformed predictors were used to do all of the following analysis. The model results with transformed predictors won't be shown in this dissertation.

3.2.2 Soil data

One set of the potential predictors derived from GISHydro2000 was the percentage of hydrologic soil type. For any given watershed outlet, GISHydro2000

can provide the watershed percentage as soil type A, B, C, and D. There are three different soil layer coverages, STATSGO, Ragan, and SSURGO. STATSGO is stored as a vector format and converted to raster 100-foot (30.48-meter) within GISHydro2000. The resolution of this coverage is the poorest of the three, but these data cover the entire region of GISHydro2000, not just what lies within Maryland. Ragan data were scanned from county soil maps and hand digitized. The native resolution is 400-foot by 500-foot (121.92-meter by 152.4-meter) per pixel, but has been resampled to 30-meter resolution. These data cover only the state of Maryland. SSURGO has very high quality and resolution. It was available only for selected counties within the state when the models was built, and now it is available everywhere in Maryland but Allegany County. Due to the limitations described above and to suit our purpose, SSURGO was not considered since the purpose is to make this predictive model usable statewide. The soil data obtained from STATSGO are called SSA, SSB, SSC, and SSD. The soil data obtained from Ragan are called RSA, RSB, RSC, and RSD.

It must be pointed out that these soil data were obtained from the GISHydro2000 package. All the regression functions that were developed for the predictive model were based on these data. Caution must be used if these regression functions are applied to data from any other source.

3.2.3 Two sets of data

At this point, all potential predictors have been developed. The predictors were gathered into two sets. One set is only the predictors from GISHydro2000. This will be convenient later when incorporating the model into this software package. The

other set is the predictors from GISHydro2000 and on-site measurements (It will be referred as “all predictors”). This data set will be used to evaluate how the model can be improved with more predictors included. The predictor variables are listed in Table 3.

The earlier RIVPACS-style models for Maryland (Moglen and Paul, 2005; Paul unpublished, 2006) was developed using soil data from the Ragan soil layer. To better compare results from different predictive styles, the same data sources were retained to build Region-style models. In this dissertation, all the Region-style macroinvertebrate predictive models were built based on Ragan soil data layer. When applying the models to GISHydro2000, this tool will be more convenient for users if the soil data layer is based on STATSGO. To increase the flexibility and applicability of the model, a predictive model based on the STATSGO data layer was developed to be incorporated into the GISHydro2000 software. Models built from both STATSGO and the Ragan soil layer will be available for the users to select.

Table 3. Two sets of parameters used in this study

GISHydro2000 Predictors	AREA, WSLOPE, CSLOPE, RELIEF, LIME, HUC14, HUC8, PERIM, LENGTH, RSA, RSB, RSC, RSD, ELEV, SIN, P2, P100, HIGHELEV, HYPSON, EASTING(X), NORTHING(Y), LAT, LONG
All Predictors	AREA, WSLOPE, CSLOPE, RELIEF, LIME, HUC14, HUC8, PERIM, LENGTH, RSA, RSB, RSC, RSD, ELEV, SIN, P2, P100, HIGHELEV, HYPSON, EASTING(X), NORTHING(Y), LAT, LONG, DATE_SPR, JDAY, BOULDLT2, COBBLE, BEDROCK, MAXDEPTH, AVGWID

All the acronyms are defined in Table 2.

CHAPTER 4: MODEL PARAMETERS FOR MARYLAND –

RESULTS

4.1 SUBSET PREDICTORS AND OVERFITTING ANALYSIS RESULT

Two candidate predictor sets are now ready to be used. As mentioned in Section 2.2, different techniques were applied to find out what parameters can serve as predictors in the model without overfitting. First, the cross-validation was used to analyze the different parameters for selection. The following are the results.

4.1.1 Cross-Validation

For the cross-validation results, in general it was expected that, as the model order increases, the estimated prediction error will decrease up to a particular model order, the estimated prediction error will remain stable afterward. The “leaps” library in R was used to find the best five models with the smallest RSS. These models were evaluated by cross-validation.

Figure 3 and Figure 4 represent estimated prediction error compared to model order. The point where the estimated prediction reaches a minimum is the point with the highest model order that remains stable and, hence, is the best model choice. Since there are three dimensions after the NMDS (Section 2.1.Step1), regression functions are at each axis (axis 1, axis 2, and axis 3).

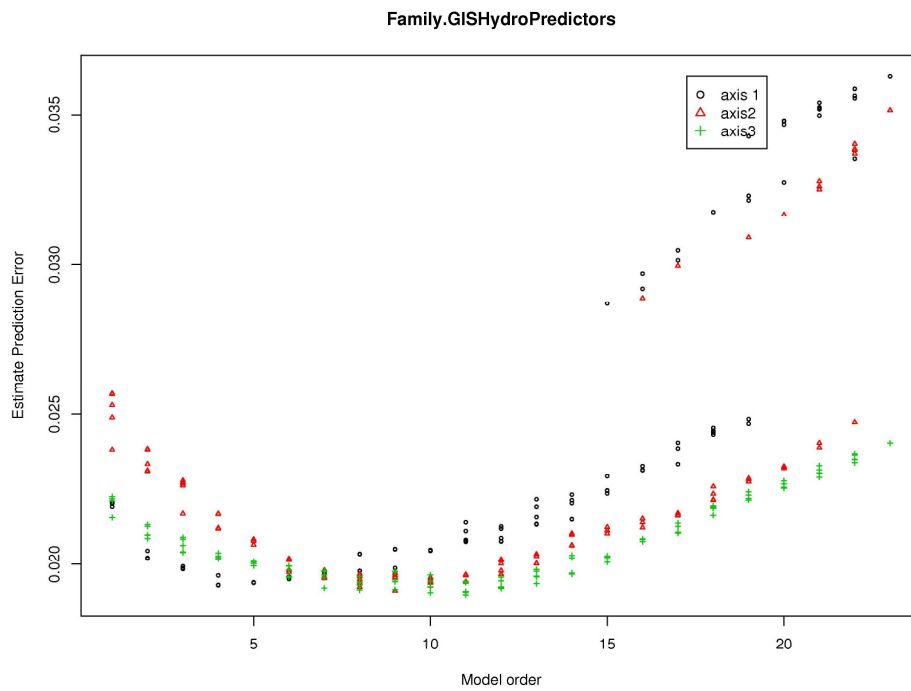
In the family level predictors from GISHydro2000, axis 1 reaches its lowest value at model order 4, axis 2 at model order 9, and axis 3 at model order 11. For the

family level using all predictors, axis 1 reaches its lowest value at model order 8, axis 2 at model order 9, and axis 3 at model order 11. The results are shown in Figure 3.

In the genus level predictors from GISHydro2000, axis 1 reaches its lowest value at model order 6, axis 2 at model order 8, and axis 3 at model order 5. For the genus level using all predictors, axis 1 reaches its lowest value at model order 12, axis 2 at model order 11, and axis 3 at model order 9. The results are shown in Figure 4.

The results shows that the estimated prediction initially decreased as the model order increased. After reaching its lowest estimated prediction error, however, some values of estimated prediction error start to increase again. These may due to the multicollinearity discussed in Section 2.2. More multicollinear predictors were included in the model as model order increased. But when more than one predictor tries to explain the same scenario, instead of more precision, more noise and error are added to the model. The lowest model order and its predictor combination at each axis may give us the best predictor sets for different taxa levels. But the estimated prediction increased after reaching the lowest point instead of staying stable, the results aren't convincing enough to show that these predictors are non-overfitting predictor sets. Therefore other methods to select the parameters were studied.

(a)



(b)

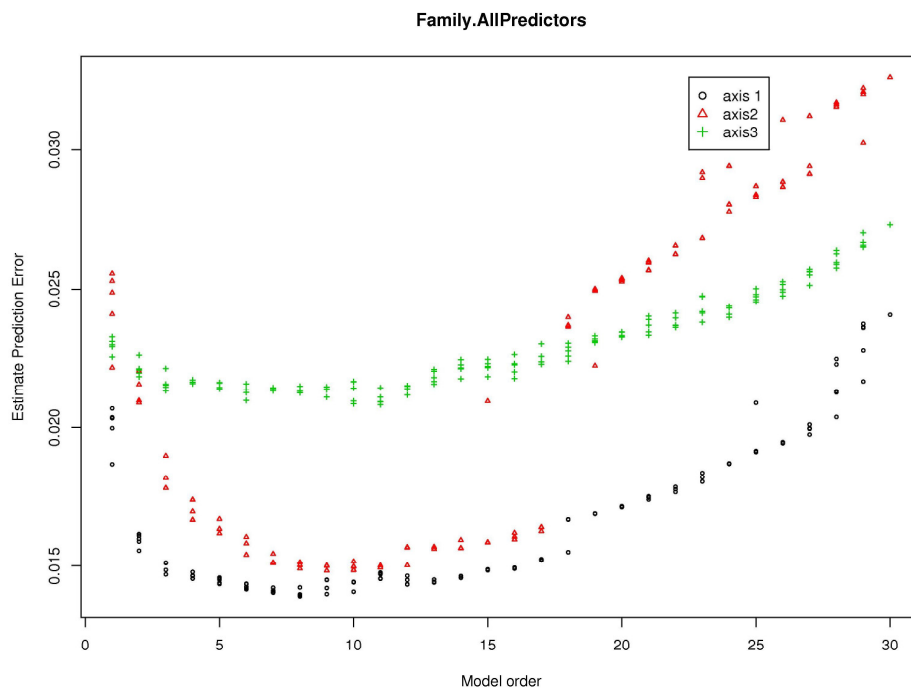
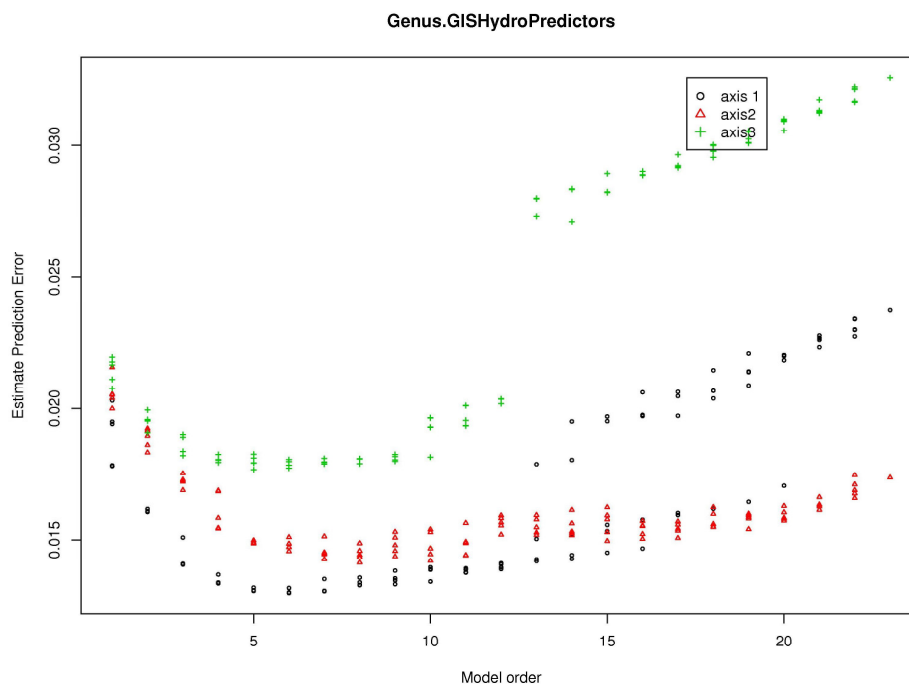


Figure 3. Family level cross validation results. (a)GISHydro2000 generated predictors (b)All predictors

(a)



(b)

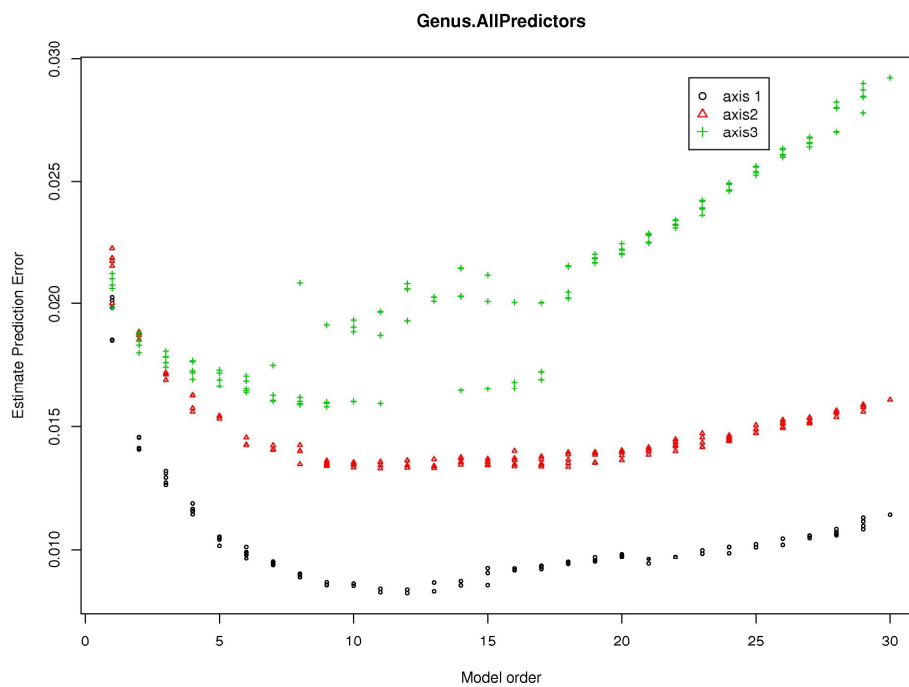


Figure 4. Genus level cross validation results. (a)GISHydro2000 generated predictors (b)All predictors

4.1.2 Exhaustive AIC, Exhaustive BIC

Cross-Validation does not prove that the models are not overfitting, so two different selection criteria to avoid overfitting and find the best predictor sets were studied and applied. As described in Section 2.2.4, the best five subsets with smallest RSS were identified by “leaps” library in R, which explored all the possible combinations of subset parameters in each order. AIC_c and BIC were calculated using RSS from each subset parameters and the subset parameters with smallest AIC_c and BIC at each axis were selected as the regression functions. The results for each macroinvertebrate determination level (family, genus) are shown in Figures 5, 6, 7, and 8, and the best parameter sets are listed in Tables 4, 5, 6, and 7.

As expected, the calculated AIC_c and BIC values decreased as the model order increased, and after reaching the minimum, both AIC_c and BIC values increased as the model order increased. This is because of the penalty from the second term in the equations 17 and 22. In general, both AIC_c and BIC reduce the number of selected predictors. BIC particularly selects limit parameters as predictors, because it has a stronger penalty than AIC_c .

The parameters selected by these two criteria were also collected. The results of parameter selection were summarized in Table 8, which shows that ELEV from GISHydro2000 was selected 13 times if all three axes were considered. The next most frequent selections were HYPSON, P100, SIN and HUC8, with 12, 10, 10, and 10 selections, respectively. These five parameters were selected 40% or more frequently, out of 24 selections. These are the best predictors of the biodiversity differences between sites. COBBLE from MBSS was selected 8 times out of 12. It is the best out

of all the predictors from MBSS. It is followed by DATE_SPR and BOULDLT2. Both were selected seven times. These results suggest ELEV, HYPISO, SIN, P100, HUC8, COBBLE, DATE_SPR, and BOULDLT2 are appropriate predictor variables for our models. It also shows that on-site parameters were not ignored in the selection process and do play important roles when building predictive models.

These criteria might help eliminate the less significant parameters as predictors. However, they do not necessarily prevent overfitting. It is necessary to eliminate those factors that cause overfitting to provide a better predictive model.

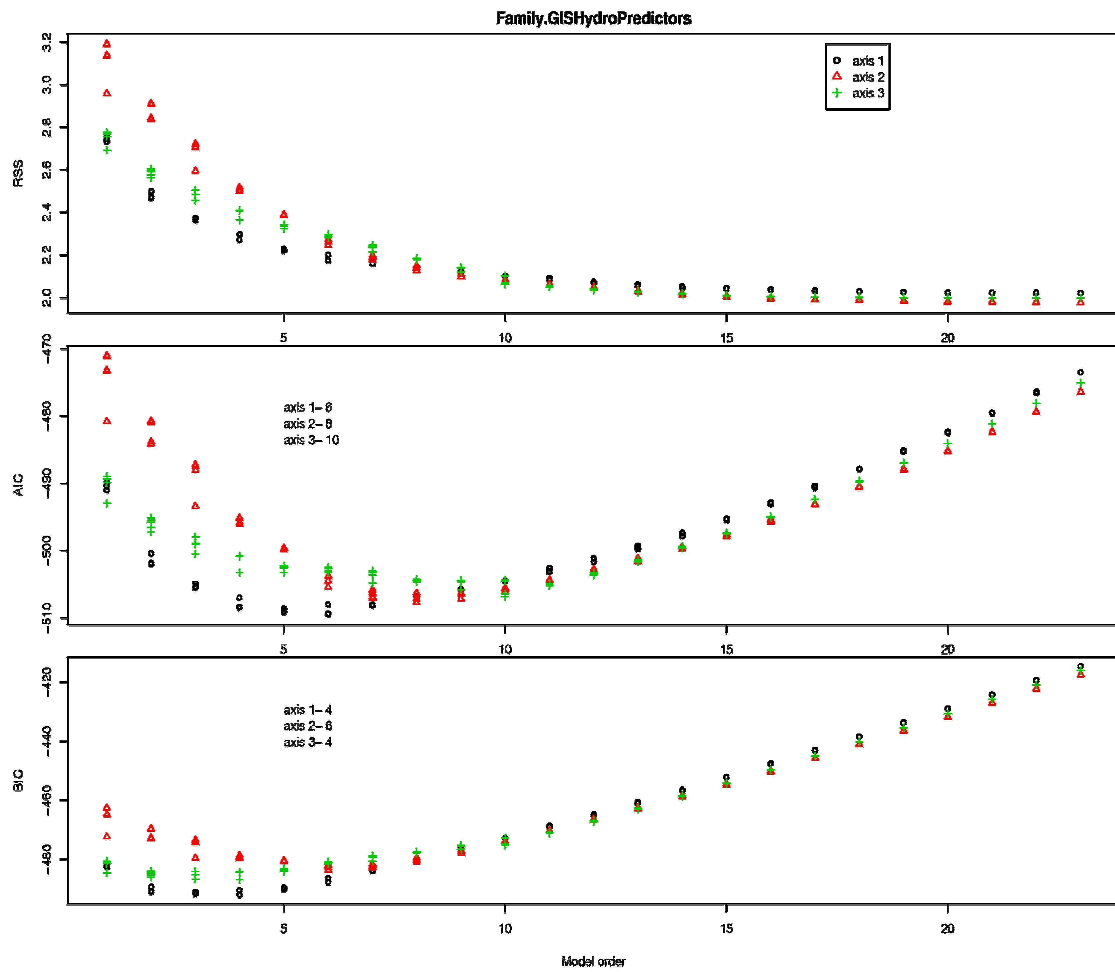


Figure 5. Exhaustive AIC, BIC searches at the family level with GISHydro2000 predictors. Open circles denote axis 1. Open triangles denote axis 2. Crosses denote axis 3. The number after the hyphen is the model order with lowest AIC or BIC at respective axis.

Table 4. The best combination for the family level with GISHydro2000 predictors using AIC_c and BIC criteria.

Criteria	Axis	Predictors
AIC _c	Axis 1 – 6	RSB, ELEV, SIN, HYP SO, LONG, Y
	Axis 2 – 8	WSLOPE, LIME, HUC14, HUC8, ELEV, P100, HIGHELEV, HYP SO
	Axis 3 – 10	CSLOPE, LIME, HUC8, RSB, RSC, P2, P100, HIGHELEV, LAT, Y
BIC	Axis 1 – 4	RSB, ELEV, LAT, LONG
	Axis 2 – 6	WSLOPE, HUC14, ELEV, P100, HIGHELEV, HYP SO
	Axis 3 – 4	SIN, P100, LAT, Y

All the acronym is defined in Table 2. The number after the hyphen is the model order with lowest AIC_c or BIC at respective axis.

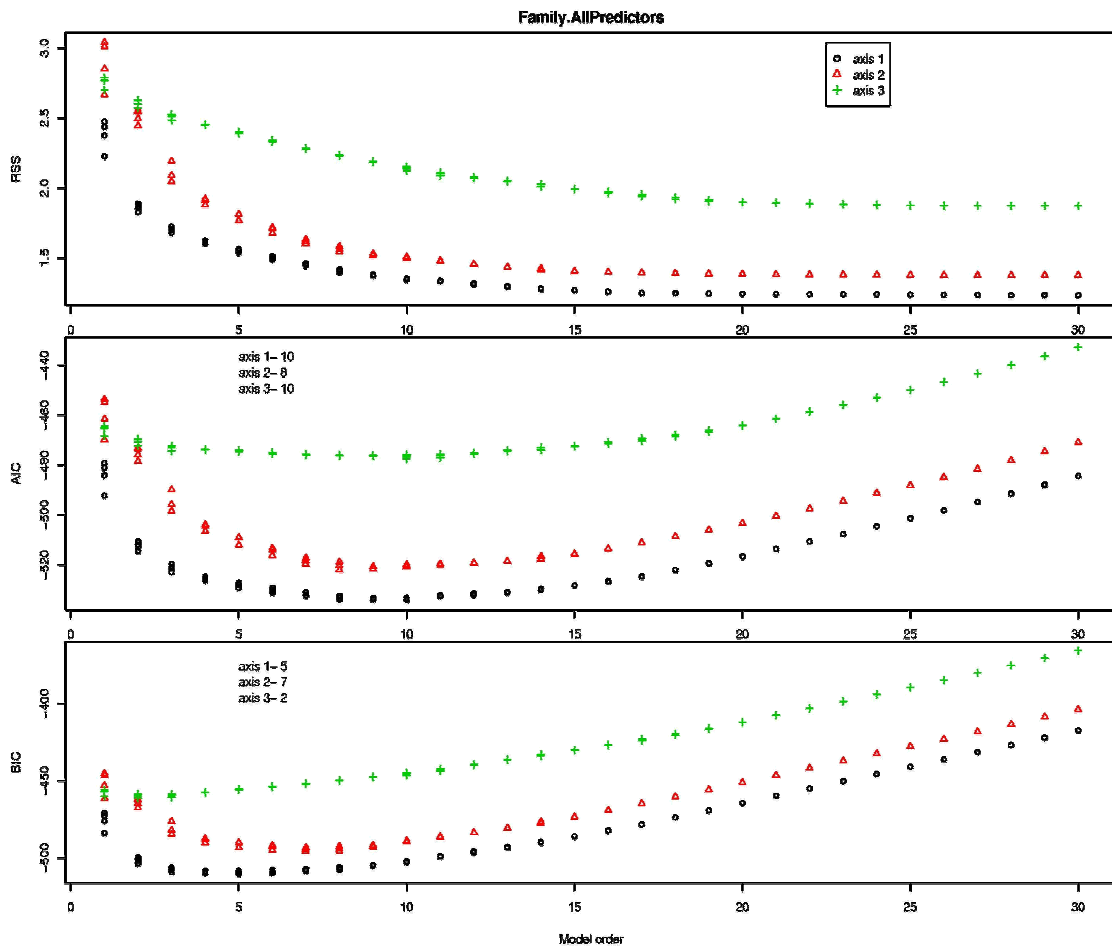


Figure 6. Exhaustive AIC, BIC searches at the family level with all predictors. Open circles denote axis 1. Open triangles denote axis 2. Crosses denote axis 3. The number after the hyphen is the model order with lowest AIC or BIC at respective axis.

Table 5. The best combination for the family level with all predictors using AIC_c and BIC criteria.

Criteria	Axis	Predictors
AIC _c	Axis 1 – 10	DATE_SPR, LIME, HUC14, RSB, ELEV, SIN, LONG, BOULDLT2, COBBLE, MAXDEPTH
	Axis 2 – 8	HUC14, HUC8, RSA, ELEV, SIN, HYPISO, COBBLE
	Axis 3 – 10	DATE_SPR, CSLOPE, HUC8, RSB, RSC, ELEV, P2, P100, HYPISO, BOULDLT2
BIC	Axis 1 – 5	RSB, P100, HYPISO, BOULDLT2, COBBLE
	Axis 2 – 7	HUC14, HUC8, ELEV, SIN, HYPISO, COBBLE
	Axis 3 – 2	RSB, SIN

All the acronym is defined in Table 2. The number after the hyphen is the model order with lowest AIC_c or BIC at respective axis.

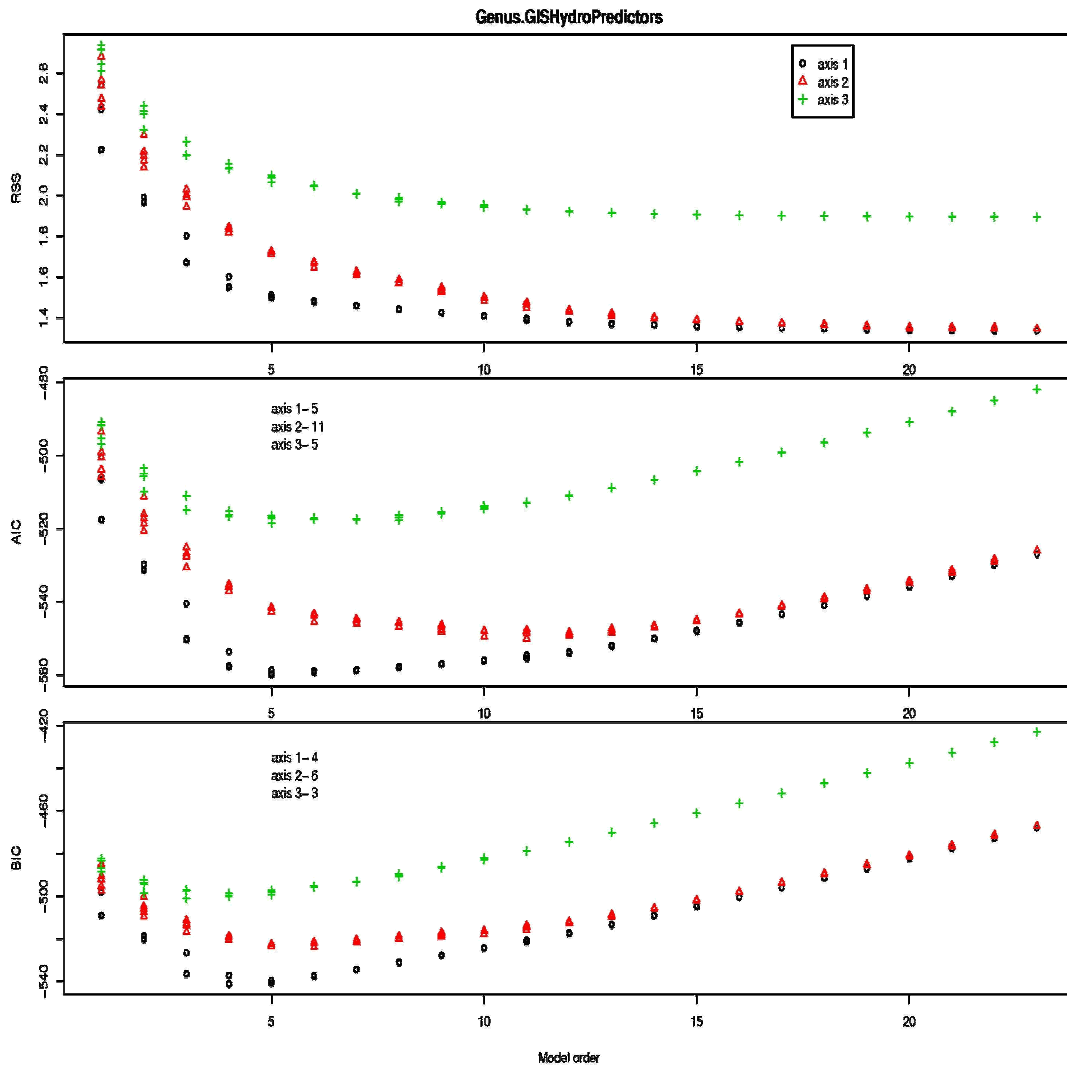


Figure 7. Exhaustive AIC, BIC searches at the genus level with GISHydro2000 predictors. Open circles denote axis 1. Open triangles denote axis 2. Crosses denote axis 3. The number after the hyphen is the model order with lowest AIC or BIC at respective axis.

Table 6. The best combination for the genus level with GISHydro2000 predictors

Criteria	Axis	Predictors
AIC _c	Axis 1 – 5	HUC8, SIN, HYPSO, LAT, LONG
	Axis 2 – 11	AREA, WSLOPE, LIME, HUC8, PERIM, P2, P100, LAT, LONG, X, Y
	Axis 3 – 5	WSLOPE, HUC14, HUC8, ELEV, HIGHELEV
BIC	Axis 1 - 4	SIN, HYPSO, LAT, X
	Axis 2 - 6	AREA, PERIM, ELEV, P100, LAT, Y
	Axis 3 - 3	WSLOPE, ELEV, HIGHELEV

All the acronym is defined in Table 2. The number after the hyphen is the model order with lowest AIC_c or BIC at respective axis.

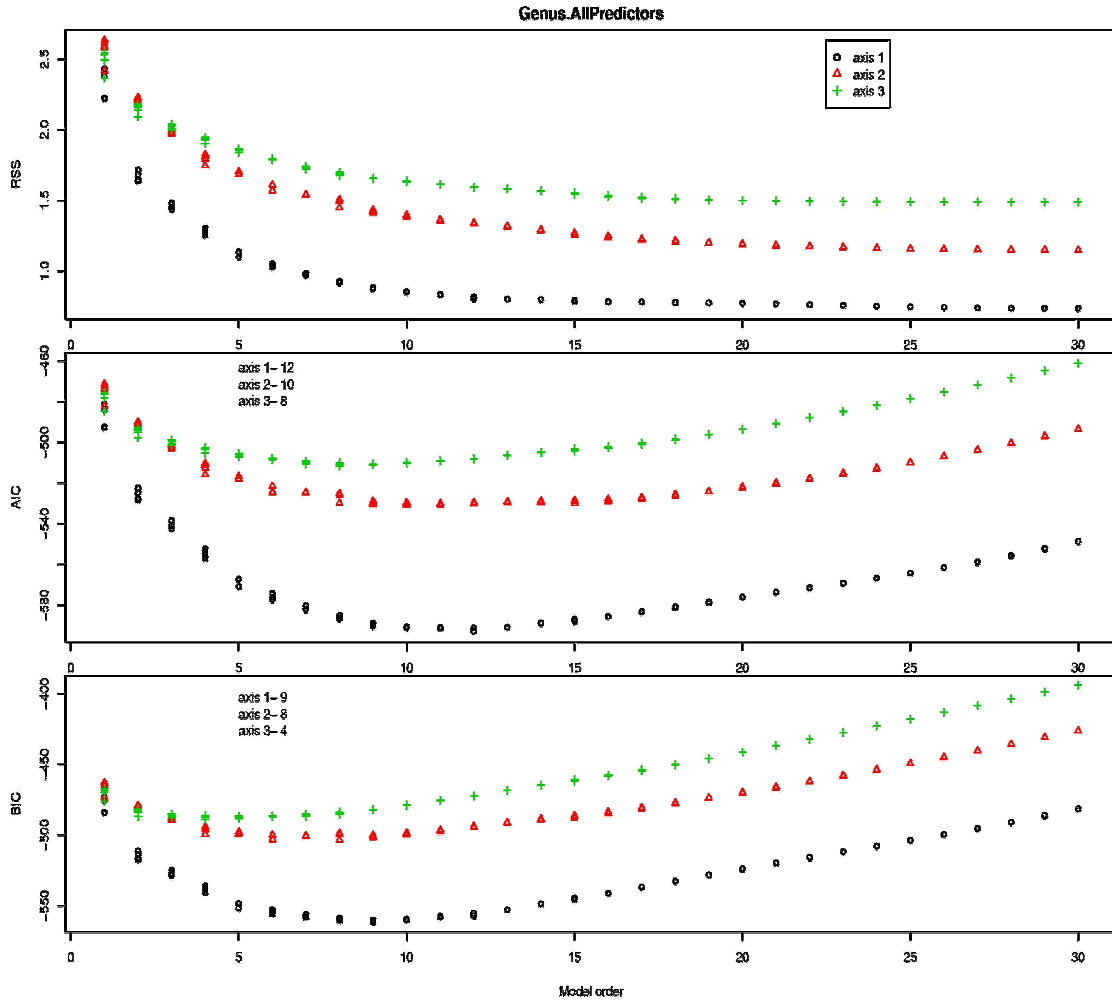


Figure 8. Exhaustive AIC, BIC searches at the genus level with all predictors. Open circles denote axis 1. Open triangles denote axis 2. Crosses denote axis 3. The number after the hyphen is the model order with lowest AIC or BIC at respective axis.

Table 7. The best combination for the genus level with all predictors

Criteria	Axis	Predictors
AIC _c	Axis 1 – 12	DATE_SPR, RELIEF, PERIM, LENGTH, SIN, HYP SO, LONG, BOULDLT2, COBBLE, MAXDEPTH, AVGWID, Y
	Axis 2 – 10	DATE_SPR, HUC8, PERIM, LENGTH, ELEV, P100, LAT, BOULDLT2, COBBLE, Y
	Axis 3 – 8	DATE_SPR, JDAY, AREA, HUC14, HUC8, LENGTH, HYP SO, COBBLE
BIC	Axis 1 – 9	DATE_SPR, RELIEF, SIN, HYP SO, BOULDLT2, COBBLE, MAXDEPTH, AVGWID, X
	Axis 2 – 8	DATE_SPR, PERIM, LENGTH, ELEV, P100, LAT, BOULDLT2, Y
	Axis 3 – 4	RELIEF, HUC8, HYP SO, COBBLE

All the acronym is defined in Table 2. The number after the hyphen is the model order with lowest AIC_c or BIC at respective axis.

Table 8. Summary parameter selection results from different criteria (AIC: Exhaustive AIC, BIC: Exhaustive BIC)

Predictor Set	Family												Genus												SUM	%	
	GISHydro2000						All						GISHydro2000						All								
	AIC			BIC			AIC			BIC			AIC			BIC			AIC			BIC					
Selection Method	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3			
GISHydro2000	AREA														X			X				X				3	12.5
	WSLOPE		X			X								X	X				X							5	20.8
	CSLOPE			X						X																2	8.3
	RELIEF																	X				X			X	3	12.5
	LIME		X	X					X						X											4	16.7
	HUC14		X			X			X	X			X			X			X			X				7	29.2
	HUC8		X	X					X	X			X		X	X									X	10	41.7
	PERIM														X			X	X					X		5	20.8
	LENGTH																	X	X					X		3	12.5
	RSA								X																	1	4.2
	RSB	X			X	X		X		X	X		X													7	29.2
	RSC			X						X																2	8.3
	RSD																									0	0.0
	ELEV	X	X		X	X		X	X	X			X			X		X	X			X		X		13	54.2
	SIN	X					X	X	X				X	X			X					X		X		10	41.7
	P2			X						X					X											3	12.5
	P100		X	X		X	X			X	X				X			X						X		10	41.7
	HIGHELEV		X	X		X										X										5	20.8
	HYPSON	X	X			X			X	X	X	X		X			X				X			X	X	12	50.0
	LAT			X	X		X							X	X		X	X			X			X		9	37.5
	LONG	X			X			X						X	X						X					6	25.0
	EASTING, X														X		X					X		X		3	12.5
	NORTHING, Y	X		X			X								X		X	X				X		X		8	33.3
	MBSS	DATE_SPR	-	-	-	-	-	-	X		X				-	-	-	-	-	-	X	X	X	X	X		7
JDAY		-	-	-	-	-	-							-	-	-	-	-	-			X				1	8.3
BOULDLT2		-	-	-	-	-	-	X		X	X			-	-	-	-	-	-	X	X		X	X		7	58.3
COBBLE		-	-	-	-	-	-	X	X		X	X		-	-	-	-	-	-	X	X		X		X	8	66.7
BEDROCK		-	-	-	-	-	-							-	-	-	-	-	-							0	0.0
MAXDEPTH		-	-	-	-	-	-	X						-	-	-	-	-	-	X			X			3	25.0
AVGWID		-	-	-	-	-	-							-	-	-	-	-	-	X			X			2	16.7

All the acronym is defined in Table 2. X denotes the parameter was selected using different criteria. The number under SUM column is the number of times each parameter was selected out of 24 (for GISHydro2000 predictor) and 12 (for MBSS data) potential selections. The number under % column is the selection percentage.

4.1.3 Rational Regression Checking

Rational regression checking, as explained in the Section 2.2.2, eliminates parameters that over-explain the variance, and thus pick out predictors with multicollinearity. After the rational regression checking, different selection methods, Stepwise AIC, Exhaustive AIC, and Exhaustive BIC, were still applied to select the predictors. The regression function results are listed in Tables 9, 10, 11, and 12 for different cases.

Table 9. Rational regression checking result for the family level with GISHydro2000 predictors.

Axis 1 deleted: ELEV,LAT, HIGHELEV

Axis 2 deleted: Y, X, ELEV

Axis 3 deleted: P2, LAT

Criteria	Axis	Regression Functions
Stepwise AIC	Axis 1 – 5	$0.103501403 + 0.060127473 \cdot (\text{SIN}) - 0.208011304 \cdot (\text{HYPSO}) + 4.87\text{E} - 07 \cdot (\text{X}) - 1.38\text{E} - 06 \cdot (\text{Y}) + 0.00142706 \cdot (\text{RSD})$
	Axis 2 – 7	$- 4.381554119 + 0.011999688 \cdot (\text{LIME}) - 9.15\text{E} - 07 \cdot (\text{HUC14}) - 2.92\text{E} - 08 \cdot (\text{HUC8}) - 0.090724613 \cdot (\text{SIN}) + 0.045777747 \cdot (\text{P100}) - 0.427729739 \cdot (\text{HYPSO}) + 0.115512656 \cdot (\text{LAT})$
	Axis 3 – 5	$0.374721272 - 2.67\text{E} - 08 \cdot (\text{HUC8}) - 0.00362953 \cdot (\text{LENGTH}) - 0.001279058 \cdot (\text{RSB}) - 0.083912161 \cdot (\text{SIN}) - 3.57\text{E} - 07 \cdot (\text{X})$
Exhaustive AIC	Axis 1 – 5	$0.103501403 + 0.060127473 \cdot (\text{SIN}) - 0.208011304 \cdot (\text{HYPSO}) + 4.87\text{E} - 07 \cdot (\text{X}) - 1.38\text{E} - 06 \cdot (\text{Y}) + 0.00142706 \cdot (\text{RSD})$
	Axis 2 – 6	$- 4.684527073 - 7.94\text{E} - 07 \cdot (\text{HUC14}) - 2.95\text{E} - 08 \cdot (\text{HUC8}) - 0.091748923 \cdot (\text{SIN}) + 0.04670853 \cdot (\text{P100}) - 0.466234625 \cdot (\text{HYPSO}) + 0.123474652 \cdot (\text{LAT})$
	Axis 3 – 5	$0.374721272 - 2.67\text{E} - 08 \cdot (\text{HUC8}) - 0.00362953 \cdot (\text{LENGTH}) - 0.001279058 \cdot (\text{RSB}) - 0.083912161 \cdot (\text{SIN}) - 3.57\text{E} - 07 \cdot (\text{X})$
Exhaustive BIC	Axis 1 – 3	$0.221544784 - 0.26303684 \cdot (\text{HYPSO}) + 5.47\text{E} - 07 \cdot (\text{X}) - 1.42\text{E} - 06 \cdot (\text{Y})$
	Axis 2 – 6	$- 4.684527073 - 7.94\text{E} - 07 \cdot (\text{HUC14}) - 2.95\text{E} - 08 \cdot (\text{HUC8}) - 0.091748923 \cdot (\text{SIN}) + 0.04670853 \cdot (\text{P100}) - 0.466234625 \cdot (\text{HYPSO}) + 0.123474652 \cdot (\text{LAT})$
	Axis 3 – 2	$0.209020012 - 0.001618405 \cdot (\text{RSB}) - 0.102125759 \cdot (\text{SIN})$

All the acronym is defined in Table 2. The number after the hyphen is the model order with lowest AIC_c or BIC at respective axis.

Table 10. Rational regression checking result for the family level with all predictors.

Axis 1 deleted: PERIM, HUC14, ELEV, LAT, HIGHELEV
 Axis 2 deleted: X, LONG, Y, HIGHELEV, ELEV, LENGTH
 Axis 3 deleted: LAT, P2, HIGHELEV, BOULDLTS

Criteria	Axis	Regression Functions
Stepwise AIC	Axis 1 – 8	$-0.84404 + 1.94E-05 \cdot (\text{DATE_SPR}) - 0.00101 \cdot (\text{RSB})$ $+ 0.065267 \cdot (\text{SIN}) + 0.051929 \cdot (\text{P100}) - 0.22827 \cdot (\text{HYPSO})$ $- 0.07838 \cdot (\text{BOULDLT2}) - 0.17484 \cdot (\text{COBBLE})$ $+ 0.000515 \cdot (\text{MAXDEPTH})$
	Axis 2 – 8	$-2.60776 - 6.72E-07 \cdot (\text{HUC14}) - 1.99E-08 \cdot (\text{HUC8})$ $+ 0.002152 \cdot (\text{RSA}) - 0.08862 \cdot (\text{SIN}) + 0.066663 \cdot (\text{P100})$ $- 0.60475 \cdot (\text{HYPSO}) + 0.063805 \cdot (\text{LAT}) + 0.18374 \cdot (\text{COBBLE})$
	Axis 3 – 6	$-0.51215 + 2.69E-08 \cdot (\text{HUC8}) + 0.001231 \cdot (\text{RSB}) + 0.089125 \cdot (\text{SIN})$ $+ 0.079726 \cdot (\text{COBBLE}) + 0.000705 \cdot (\text{MAXDEPTH}) + 4.80E-07 \cdot (\text{X})$
Exhaustive AIC	Axis 1 – 8	$-0.84404 + 1.94E-05 \cdot (\text{DATE_SPR}) - 0.00101 \cdot (\text{RSB})$ $+ 0.065267 \cdot (\text{SIN}) + 0.051929 \cdot (\text{P100}) - 0.22827 \cdot (\text{HYPSO})$ $- 0.07838 \cdot (\text{BOULDLT2}) - 0.17484 \cdot (\text{COBBLE})$ $+ 0.000515 \cdot (\text{MAXDEPTH})$
	Axis 2 – 8	$-2.60776 - 6.72E-07 \cdot (\text{HUC14}) - 1.99E-08 \cdot (\text{HUC8})$ $+ 0.002152 \cdot (\text{RSA}) - 0.08862 \cdot (\text{SIN}) + 0.066663 \cdot (\text{P100})$ $- 0.60475 \cdot (\text{HYPSO}) + 0.063805 \cdot (\text{LAT}) + 0.18374 \cdot (\text{COBBLE})$
	Axis 3 – 6	$-0.51215 + 2.69E-08 \cdot (\text{HUC8}) + 0.001231 \cdot (\text{RSB})$ $+ 0.089125 \cdot (\text{SIN}) + 0.079726 \cdot (\text{COBBLE})$ $+ 0.000705 \cdot (\text{MAXDEPTH}) + 4.80E-07 \cdot (\text{X})$
Exhaustive BIC	Axis 1 – 5	$-0.02575 - 0.001173236 \cdot (\text{RSB}) + 0.053028 \cdot (\text{P100})$ $- 0.22588 \cdot (\text{HYPSO}) - 0.07968 \cdot (\text{BOULDLT2}) - 0.17249 \cdot (\text{COBBLE})$
	Axis 2 – 5	$-0.11965 - 5.45E-07 \cdot (\text{HUC14}) - 0.09121 \cdot (\text{SIN})$ $+ 0.062243 \cdot (\text{P100}) - 0.62399 \cdot (\text{HYPSO}) + 0.196761 \cdot (\text{COBBLE})$
	Axis 3 – 2	$-0.20457 + 0.001480487 \cdot (\text{RSB}) + 0.102144 \cdot (\text{SIN})$

All the acronym is defined in Table 2. The number after the hyphen is the model order with lowest AIC_c or BIC at respective axis.

Table 11. Rational regression checking result for genus level with GISHydro2000 predictors.

Axis 1 deleted: LAT, LONG, HIGHELEV, HUC8, AREA
 Axis 2 deleted: LAT, LONG, P2, WSLOPE, AREA, ELEV
 Axis 3 deleted: HIGHELEV, Y, P2, X, LONG

Criteria	Axis	Regression Functions
Stepwise AIC	Axis 1 – 5	$0.115378 - 8.08E - 05 \cdot (\text{CSLOPE}) + 0.087613 \cdot (\text{SIN})$ $- 0.41304 \cdot (\text{HYP SO}) + 7.90E - 07 \cdot (\text{X}) - 1.44E - 06 \cdot (\text{Y})$
	Axis 2 – 7	$0.601372 - 0.014604154 \cdot (\text{LIME}) + 2.87E - 07 \cdot (\text{HUC14})$ $- 0.00218 \cdot (\text{PERIM}) + 0.00094 \cdot (\text{RSC}) - 0.00148 \cdot (\text{HIGHELEV})$ $- 9.95E - 07 \cdot (\text{X}) - 1.33E - 06 \cdot (\text{Y})$
	Axis 3 – 7	$3.151745 - 0.839735418 \cdot (\text{WSLOPE}) + 0.000152 \cdot (\text{RELIEF})$ $+ 4.28E - 07 \cdot (\text{HUC14}) + 3.14E - 08 \cdot (\text{HUC8}) - 0.04102 \cdot (\text{P100})$ $+ 0.238701 \cdot (\text{HYP SO}) - 0.07721 \cdot (\text{LAT})$
	Axis 1 – 5	$0.115378 - 8.08E - 05 \cdot (\text{CSLOPE}) + 0.087613 \cdot (\text{SIN})$ $- 0.41304 \cdot (\text{HYP SO}) + 7.90E - 07 \cdot (\text{X}) - 1.44E - 06 \cdot (\text{Y})$
	Axis 2 – 7	$0.601372 - 0.014604154 \cdot (\text{LIME}) + 2.87E - 07 \cdot (\text{HUC14})$ $- 0.00218 \cdot (\text{PERIM}) + 0.00094 \cdot (\text{RSC}) - 0.00148 \cdot (\text{HIGHELEV})$ $- 9.95E - 07 \cdot (\text{X}) - 1.33E - 06 \cdot (\text{Y})$
	Axis 3 – 7	$3.151745 - 0.839735418 \cdot (\text{WSLOPE}) + 0.000152 \cdot (\text{RELIEF})$ $+ 4.28E - 07 \cdot (\text{HUC14}) + 3.14E - 08 \cdot (\text{HUC8}) - 0.04102 \cdot (\text{P100})$ $+ 0.238701 \cdot (\text{HYP SO}) - 0.07721 \cdot (\text{LAT})$
Exhaustive BIC	Axis 1 – 4	$0.10415 + 0.098467415 \cdot (\text{SIN}) - 0.44065 \cdot (\text{HYP SO})$ $+ 8.23E - 07 \cdot (\text{X}) - 1.51E - 06 \cdot (\text{Y})$
	Axis 2 – 4	$0.81055 - 0.002427075 \cdot (\text{PERIM}) - 0.00169 \cdot (\text{HIGHELEV})$ $- 1.24E - 06 \cdot (\text{X}) - 1.64E - 06 \cdot (\text{Y})$
	Axis 3 – 2	$2.760969 + 4.00E - 08 \cdot (\text{HUC8}) - 0.07296 \cdot (\text{LAT})$

All the acronym is defined in Table 2. The number after the hyphen is the model order with lowest AIC_c or BIC at respective axis.

Table 12. Rational regression checking result for the genus level with all predictors.

Axis1 deleted: RSB, PERIM
 Axis2 deleted: X, Y, PERIM
 Axis3 deleted: Y, DATA_SPR, AREA, PERIM

Criteria	Axis	Regression Functions
Stepwise AIC	Axis 1 – 11	$2.208317 + 2.41E - 05 \cdot (\text{DATE_SPR}) - 0.00116 \cdot (\text{JDAY})$ $- 0.0002 \cdot (\text{RELIEF}) + 0.002734 \cdot (\text{LENGTH}) + 0.057021 \cdot (\text{SIN})$ $- 0.21832 \cdot (\text{HYPSO}) - 0.03577 \cdot (\text{LONG}) - 0.102879 \cdot (\text{BOULDLT2})$ $- 0.16579 \cdot (\text{COBBLE}) + 0.000992 \cdot (\text{MAXDEPTH})$ $+ 0.00774 \cdot (\text{AVGWID})$
	Axis 2 – 10	$- 6.29797 + 3.82E - 05 \cdot (\text{DATE_SPR}) - 0.9912 \cdot (\text{WSLOPE})$ $+ 0.000102 \cdot (\text{CSLOPE}) - 2.63E - 08 \cdot (\text{HUC8}) + 0.001893 \cdot (\text{RSA})$ $+ 0.001452 \cdot (\text{RSC}) - 0.05138 \cdot (\text{SIN}) + 0.0671 \cdot (\text{LONG})$ $- 0.12967 \cdot (\text{BOULDLT2}) - 0.00977 \cdot (\text{AVGWID})$
	Axis 3 – 7	$- 0.32875 + 0.002052 \cdot (\text{JDAY}) - 0.37383 \cdot (\text{WSLOPE})$ $+ 0.000197 \cdot (\text{RELIEF}) + 4.02E - 07 \cdot (\text{HUC14}) + 3.65E - 08 \cdot (\text{HUC8})$ $+ 0.266023 \cdot (\text{HYPSO}) - 0.15509 \cdot (\text{COBBLE})$
Exhaustive AIC	Axis 1 – 5	$2.545812 + 2.68E - 05 \cdot (\text{DATE_SPR}) - 0.00016 \cdot (\text{RELIEF})$ $+ 0.065035 \cdot (\text{SIN}) - 0.27909 \cdot (\text{HYPSO}) - 0.04164 \cdot (\text{LONG})$ $- 0.0969 \cdot (\text{BOULDLT2}) - 0.14609 \cdot (\text{COBBLE})$ $+ 0.00092 \cdot (\text{MAXDEPTH}) + 0.010656 \cdot (\text{AVGWID})$ $- 4.49E - 07 \cdot (\text{Y})$
	Axis 2 – 7	$- 5.93083 + 3.56E - 05 \cdot (\text{DATE_SPR}) - 0.72848 \cdot (\text{WSLOPE})$ $- 2.57E - 08 \cdot (\text{HUC8}) + 0.001243 \cdot (\text{RSC}) - 0.05239 \cdot (\text{SIN})$ $+ 0.063764 \cdot (\text{LONG}) - 0.13185 \cdot (\text{BOULDLT2}) - 0.011853 \cdot (\text{AVGWID})$
	Axis 3 – 7	$- 0.1974 + 4.94E - 07 \cdot (\text{HUC14}) + 2.60E - 08 \cdot (\text{HUC8})$ $+ 0.004644 \cdot (\text{LENGTH}) + 0.000992 \cdot (\text{HIGHELEV})$ $+ 0.306033 \cdot (\text{HYPSO}) - 0.16605 \cdot (\text{COBBLE})$
Exhaustive BIC	Axis 1 – 9	$- 1.13117 + 3.14E - 05 \cdot (\text{DATE_SPR}) - 0.00018 \cdot (\text{RELIEF})$ $+ 0.06067 \cdot (\text{SIN}) - 0.24455 \cdot (\text{HYPSO}) - 0.10686 \cdot (\text{BOULDLT2})$ $- 0.15115 \cdot (\text{COBBLE}) + 0.001018 \cdot (\text{MAXDEPTH})$ $+ 0.010364 \cdot (\text{AVGWID}) + 5.65E - 07 \cdot (\text{X})$
	Axis 2 – 5	$- 1.37334 + 3.49E - 05 \cdot (\text{DATE_SPR}) - 0.11101 \cdot (\text{LAT})$ $+ 0.059205 \cdot (\text{LONG}) - 0.11266 \cdot (\text{BOULDLT2}) - 0.01122 \cdot (\text{AVGWID})$
	Axis 3 – 4	$- 0.14227 + 0.000162 \cdot (\text{RELIEF}) + 3.85E - 08 \cdot (\text{HUC8})$ $+ 0.268452 \cdot (\text{HYPSO}) - 0.17866 \cdot (\text{COBBLE})$

All the acronym is defined in Table 2. The number after the hyphen is the model order with lowest AIC_c or BIC at respective axis.

Table 13. Summary parameter selection results from different criteria after rational regression checking. (AIC: Exhaustive AIC, BIC: Exhaustive BIC)

Predictor Set	Family												Genus												SUM	%						
	GISHydro2000						All						GISHydro2000						All													
	Stepwise AIC			AIC			BIC			Stepwise AIC			AIC			BIC			Stepwise AIC			AIC					BIC					
Selection Method	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3		
Axis	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3		
AREA																															0	0.0
WSLOPE																															5	13.9
CSLOPE																															3	8.3
RELIEF																															7	19.4
LIME		X																													3	8.3
HUC14		X			X			X			X			X																	11	30.6
HUC8		X	X		X	X		X			X	X		X																	17	47.2
PERIM																															3	8.3
LENGTH			X			X																									4	11.1
RSA																															3	8.3
RSB			X			X		X	X		X	X		X	X																9	25.0
RSC																															4	11.1
RSD	X				X																										2	5.6
ELEV																															0	0.0
SIN	X	X	X	X	X	X		X	X	X	X	X	X	X	X																24	66.7
P2																															0	0.0
P100		X						X	X	X	X		X	X																	10	27.8
HIGHELEV																															4	11.1
HYPISO	X	X		X	X		X	X		X	X		X	X		X	X														22	61.1
LAT		X			X			X			X			X																	9	25.0
LONG																															5	13.9
EASTING, X	X		X	X		X	X			X			X			X															14	38.9
NORTHING, Y	X			X			X			X			X			X															10	27.8
DATE_SPR	-	-	-	-	-	-	-	-	-	X			X																		8	44.4
JDAY	-	-	-	-	-	-	-	-	-																						2	11.1
BOULDLT2	-	-	-	-	-	-	-	-	-	X			X																		9	50.0
COBBLE	-	-	-	-	-	-	-	-	-	X	X	X	X	X	X	X	X	X													14	77.8
BEDROCK	-	-	-	-	-	-	-	-	-																						0	0.0
MAXDEPTH	-	-	-	-	-	-	-	-	-	X			X			X															7	38.9
AVGWID	-	-	-	-	-	-	-	-	-																						6	33.3

All the acronym is defined in Table 2. X denotes the parameter was selected using different criteria. The number under SUM column is the number of times each parameter was selected out of 24 (for GISHydro2000 predictor) and 12 (for MBSS data) potential selections. The number under % column is the selection percentage.

Rational regression checking eliminated substantial multicollinearity in the initial candidate predictor sets. Stepwise AIC, Exhaustive AIC_c, and Exhaustive BIC were then used to select the predictors. The regression functions listed in Tables 9, 10, 11, 12 were the final regression functions applied to the Region-style models in the study. By comparing Stepwise AIC and Exhaustive AIC_c, the difference between stepwise regression and exhaustive parameter searching can be evaluated.

The frequency of parameters selected suggested their importance in the model (Van Sickle et al., 2006). The parameters selected by these three different criteria methods were gathered and the results are summarized in Table 13. It shows that SIN from GISHydro2000 was selected for 24 times out of a possible 36. The next most frequent selections were HYPISO, HUC8, and EASTING, with 22, 17, and 14 selections respectively. These four parameters were selected as the best predictors for the biodiversity differences between sites. COBBLE was selected 14 times out of a possible 18. It is the best predictors of all from MBSS variables. The next most frequent selections were BOULDLT2 and DATE_SPR. They were selected nine and eight times respectively. These results suggest that SIN, HYPISO, HUC8, COBBLE, BOULDLT2 and DATE_SPR are good predictors for our models.

Of those parameters that were not selected, AREA and ELEV are not particularly good predictors. But, AREA was selected three times in Table 8 and ELEV was the best prediction variable in Table 8. This suggests that ELEV is highly correlated with other parameters. Other parameters, such as P100 and HIGHELEV, can better explain the biodiversity between each site. LAT and NORTHING are the same location parameter in different coordinate systems. LAT and NORTHING are

the same location parameter in different coordinate. They were both selected in the regression function within the same axis in Table 8, the no rational regression function checking case. Using both LAT and NORTHING as predictors in the function duplicates the explanation of dependent variable and thus causes overfitting model. This kind of selection did not happen at the case after doing rational regression checking shown in Table 13. The rational regression checking successfully avoided overfitting the regression function.

The selection results from Stepwise AIC and Exhaustive AIC show little difference. It doesn't make too much difference whether using stepwise or exhaustive selection methods, if rational regression checking has been performed.

4.2 WEIGHTING FUNCTION PARAMETERS AND MODEL SELECTION

Development of weighting functions was described in section 2.1, Step 4. The weighting function development is different for each approach (ANNA, BROI, NROI) but weighting functions were developed for each approach in order to evaluate the performance of each approach. For the ANNA scheme, the number of neighbors required by test site need to be determined. For the BROI and NROI scheme, the weighting function parameters for selecting the neighbors need to be determined. Thus, the development of different weighting function parameters identified the number of neighbors for ANNA scheme and the weighting function parameters for BROI and NROI schemes.

The model with best performance must be chosen. The O/E ratio, defined in Section 2.1. Step5, was used to check the model credibility and, thus, choose the best performance model.

As described in general Step 1 (Section 2.1. Step 1), first the biodiversity data from 129 reference sites in Maryland were used to construct a Bray-Curtis matrix and NMDS was performed. Regression functions using physical parameters from the same reference sites were obtained in Section 4.1 according to the coordinate values (obtained from NMDS). The physical parameters from the same 129 reference sites then served as calibration test site data to predict the probability of presence for each taxon. As mentioned in Section 3.1, the physical parameters from the remaining 29 reference sites were used as validation test site data. The taxon observed data (O) were added, depending on the predicted taxon expected (E) as defined in Section 2.1. Step5 and the O/E ratio was then calculated. A histogram of O/E ratios from both data sets was drawn afterward. The mean and standard deviation of O/E ratios distribution from both data sets was attained. If the predictive model is well built, the taxon prediction (E) from the model at any reference site will match the taxon observed (O) on-site and thus the O/E score ideally is one. Therefore, the standard deviation of O/E score distribution from the calibration data set (also validation data set) is an indication of the range around mean 1 and is an estimate of the model's precision. To identify a well performing model, the goal is to determine which model is closest to mean 1, and has the lowest standard deviation. A scatter plot between taxon expected (E) and observed (O) was drawn for both data sets and the linear regression line between them was developed. The ideal regression line slope should be 1, which

means the model prediction (E) at the reference sites matched the observed (O). The intercept of the regression line ideally is zero. The R-square (r^2) is also calculated. R-square, sometimes referred to as the proportion of explained variation, is a useful summary for any linear model. Of the above 5 indicators (mean, standard deviation, slope, intercept, and r^2), RIVPACS models in the literature mostly emphasized the standard deviation (Moglen, 2005; Moss et al., 1999; Wright et al., 2000). In the ANNA scheme example by Linke et al. (2005), they ranked the slope, intercept and r^2 and calculated the geometric mean of the ranks of the three parameters. They explained that the geometric mean is preferred over arithmetic mean because it gives more weight to low ranks. The lowest rank score was then the criterion for determining the overall best model. Van Sickle et al. (2006), on the other hand, used root mean squared error (RMSE) as a single, overall measure of model performance. The bias of a predictive model is given by $B = [Mean(O/E) - 1.0]$, where Mean (O/E) denotes the average O/E over a set of reference sites. And the RMSE is given by:

$$RMSE(O/E) = \sqrt{\frac{1}{n} \sum \left(\frac{O}{E} - 1 \right)^2} = \sqrt{B^2 + SD^2} \quad (33)$$

where the sum is across a collection of n reference sites and SD is the standard deviation of O/E for those sites.

Either the geometric mean of slope, intercept and r^2 or the single measure RMSE to check the model creditability is reasonable and rational. Because previous RIVPAC-style models have most often used only standard deviation or RMSE for

performance checking, in this study RMSE was used for measuring the model performance in order to compare the models with RIVPACS-style models.

Different Region-style models with different weighting functions were described in Section 2.3. Each weighting function parameters has its range that depends on the definitions and limitations. In the preliminary study, different ranges and interval spans were tried for iterations in each scheme. The final ranges and spans were chosen based on the initial research and the available literature. It is described in the following.

4.2.1 For the ANNA scheme

An empirical iteration approach was used to find the appropriate number of nearest neighbors for the ANNA predictive scheme in Maryland. A total of 129 reference sites were used as calibration data. The numbers from two through 120 were iterated as neighbor reference sites for the ANNA scheme. As described in Section 2.1, RMSE was calculated and served as the model performance measure.

4.2.2 For the BROI and NROI schemes

Unlike the ANNA scheme, both BROI and NROI schemes derived from hydrology are more complicated in selecting reference neighbors. It considers both the physical parameter distance between sites and the number of reference neighbor included (Sections 2.3.2, 2.3.3). Just as with the ANNA scheme, the empirical iteration approach was used to determine the weightings of parameters. In both BROI and NROI scheme, all the parameters that define the desired ranges of influence, θ_L , θ_U , TP , TPP , were determined by the percentile of the entire distance

distribution from calibration sites. For example, for the family level with only GISHydro2000 predictors, the lower threshold, θ_L , was tested as 4%, 6%, 8%, and 10% of the entire distance data, which will provide the physical similarities between all the calibration reference sites. With the regression functions obtained from Stepwise AIC selection method, the percentile was 0.045, 0.052, 0.063, 0.070 respectively. The rest of the thresholds are obtained in the same manner with different percentage ranges. Also the minimum number of reference sites included, NST , and the weighting function parameter, n , were also approached empirically. Table 14 lists the ranges and values that have been tried for each Region of Influence method.

Table 14. Weighting function parameter testing ranges in different region-style methods

Scheme	Selecting parameters/Weighting parameters	Testing range
ANNA	Number of neighbors	2, 3, ..., 120
BROI	θ_L	5%, 10%, 15%
	θ_U	25%, 30%, 50%
	TP	50%, 75%, 85%
	NST	8, 9, 10, ..., 15
	n	2, 2.5, 3, 3.5
NROI	θ_L	4%, 6%, ... 10%
	θ_U	30%, 40%, ..., 90%
	NST	4, 5, ..., 8
	n	2, 2.5, ..., 4

ANNA: Assessment by Nearest Neighbour Analysis

BROI: Burn's Region of Influence

NROI: New Datum Region of Influence

The listed ranges might seem to be arbitrary. In the preliminary search, these ranges can better detect the model's performance. More iterations were tried with NROI scheme than BROI scheme, because fewer parameters were required for NROI scheme. Running the iterations required time, NROI with fewer parameters allows us to do a more thorough investigation.

CHAPTER 5: REGION-STYLE MODEL RESULTS

RMSE was used to choose the best performance model. The best performance model based on the calibration, however, does not necessarily perform the best with the validation data. Van Sickle et al. (2006) used RMSE to measure model performance both in calibrating and validating data. They demonstrated that RMSE from calibration data decreases as model order increases up until a specific model order was reached (Van Sickle et al., 2006, Figure 3). RMSE from validation data did not have specific trends throughout the different model orders. I consulted the author about their model selection process. After they determined the model order that was not overfitting by cross-validation, they selected models that had the best RMSE with validation data and then, from those models, chose the model with the best RMSE from calibration data. Generally, to build a model, its best performance is first attempted with calibration data. Validation data then are used to test a model's performance. I agree with Van Sickle's point of view that validation data are important. It tells us how a model really performs outside of calibration data. But, validation data are usually from a small number of test sites, in our case only 29 sites. As modeling addresses more complex topics and topics for which data are not routinely collected, the problem of limited data for validation will occur more frequently. It is not convincing to use validation results as the main model performance criteria and use that to select the best model. So the following is an alternative process for selecting the best model:

1. Find and rank the best 10 models that have the smallest RMSE during calibration phase
2. From these 10 models, choose the best model with the smallest RMSE using validation results

Step 1 emphasizes the calibration data results. Because there are more calibration data than validation data, the step can better represent the model performance. Step 2 selects model with lowest RMSE, thus emphasizing the importance of validation data. The model selected by these two steps will hopefully predict well for the entire region. This best model selection process is not perfect, but is a more representative approach than any other. So the following in Sections 5.1 and 5.2 is the best performance model for difference taxon levels, using different region selection schemes. Only the best performance from the parameter selection methods (Stepwise AIC, Exhaustive AIC, or Exhaustive BIC) in each predictive scheme (ANNA, BROI, NROI) was presented in the following.

Note also, that of the 29 validation sites, the data for only two sites departs far from the predicted results from the O vs. E scatter plot in the case of $p_c > 0.5$ (the third column on the right of Figure 9). The univariate data for 2 sites compared to the means and standard deviations for all calibration sites and validation sties are summarized in Table 15.

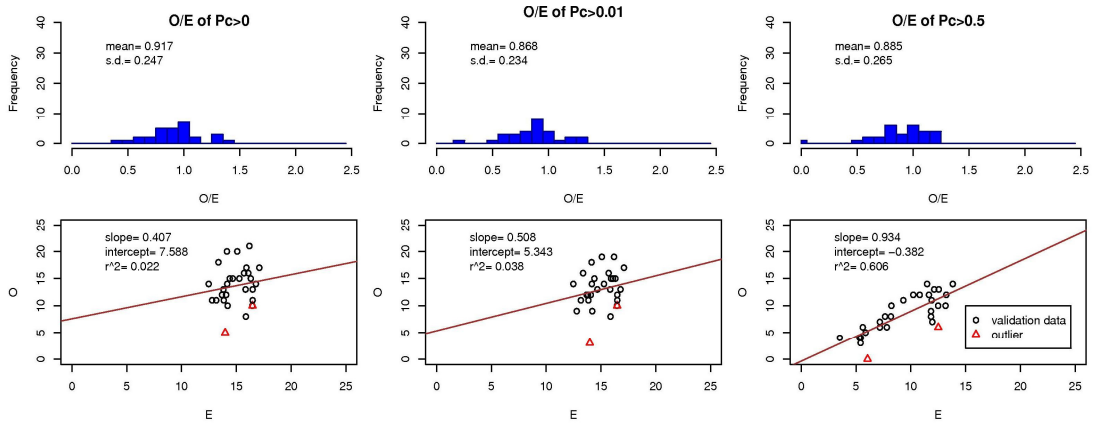


Figure 9. The validation data outliers in the O, E plots. O/E histograms are on the top and O vs.E scatter plots are on the bottom. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are from the left to right.

Table 15. O, E values comparison between outliers and calibration, validation data.

O, E, O/E	Calibration data		Validation data		Outlier 1 (CH-S-217-201-95)	Outlier 2 (PRUN-106-R-2001)	
	Mean	S.D.	Mean	S.D.	Values	Values	
$p_c > 0$	O	14.36	3.95	13.69	3.67	5	10
	E	14.65	1.42	14.99	1.33	13.98	16.47
	O/E	0.98	0.27	0.95	0.22	0.36	0.61
$p_c > 0.01$	O	13.65	3.74	12.97	3.49	3	10
	E	14.65	1.42	14.99	1.33	13.98	16.47
	O/E	0.93	0.25	0.90	0.20	0.22	0.61
$p_c > 0.5$	O	8.42	3.40	8.45	3.66	0	6
	E	8.92	2.77	9.46	3.05	6.05	12.50
	O/E	0.94	0.25	0.93	0.19	0.00	0.48

Table 15 shows that these two site have observed values apart from the means of calibration and validation data, thus they are apart from the predicted values from the model. These two sites were taken out of later results and graphs to have fair validation predictions (Table 16 to Table 19 and Figure 10 to Figure 21). After all the rules described, the region-style model results can be delivered.

5.1 MODELS WITH ONLY GISHYDRO2000 PREDICTORS

5.1.1 Family Level

Results from the region-style model at the family level and developed using only GISHydro2000 data are summarized in Table 16. All three predictive schemes (ANNA, BROI, NROI) are presented. This table shows only the results from capture probability, $p_c = 0.5$, with which most of the literature is concerned. The O/E histograms and E vs. O scatter plots of $p_c = 0$, $p_c = 0.01$, and $p_c = 0.5$ using calibration (129 sites) and validation (27 sites) data sets for each predictive scheme are shown in Figures 10, 11 and 12.

In Table 16, the three different predictive schemes are in the first column, along with corresponding parameter selection methods in the second column. Number of neighbors or parameter values for the weighting functions (Section 2.3) are listed in the third column. The two different data sets used appear in the fourth column. Since O/E is unitless, its statistics are unitless, as well. The means and standard deviation from O/E scores are listed in the fifth and sixth columns next to their corresponding model case and calibration or validation dataset. The slope and intercept of the regression function from E. vs. O scatter plot are listed in seventh and eighth columns. The R^2 of the regression is listed under the ninth column and the RMSE as defined in equation 33, for each case is listed under the tenth column. This format is applied to Tables 16, 17, 18, and 19. From the calibration results, the NROI scheme has the smallest RMSE at 0.194, which is followed by BROI and ANNA, at 0.219 and 0.255 respectively. Also, NROI has the highest R^2 at 0.733, with BROI and

ANNA following at 0.706 and 0.608 respectively. However the validation results show that the ANNA scheme has the smallest RMSE (0.204), followed by BROI and NROI at 0.232, and 0.237 respectively. ANNA also has the highest R^2 at 0.701 which is higher than R^2 from ANNA calibration data. Despite the difference, the RMSE from all the models are close in value. The ANNA scheme model has a mean close to 1.0, small standard deviation and intercept, and high R^2 . The ANNA predictive scheme is the best model for the family level model using only GISHydro2000 predictors.

Table 16. The O/E statistics at $P_c=0.5$ from all the Region-style models for the case in family level with GISHydro2000 predictors

Region Selection scheme	Best Selection Parameter method	Weighting Function parameters	Capture Probability $P_c = 0.5$	mean	s.d.	slope	intercept	R^2	RMSE
ANNA	Exhaustive BIC	Neighbor no. = 21	Calibration	0.944	0.249	0.959	-0.127	0.608	0.255
			Validation	0.932	0.192	0.925	0.087	0.701	0.204
BROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 30\%$, $TP = 50\%$, $NST = 8$, $n = 2.5$.	Calibration	1.004	0.219	1.021	-0.159	0.706	0.219
			Validation	0.875	0.195	0.769	1.066	0.521	0.232
NROI	Exhaustive AIC	$\theta_L = 10\%$, $\theta_U = 40\%$, $NST = 5$, $n = 2.5$.	Calibration	1.043	0.189	0.967	0.668	0.733	0.194
			Validation	0.877	0.202	0.744	1.317	0.49	0.237

Validation data are without the two outlier sites.
Mean and s.d. are from O/E values histogram.
Slope, intercept, and R^2 are from E vs. O scatter plots.
RMSE is defined in equation 33.

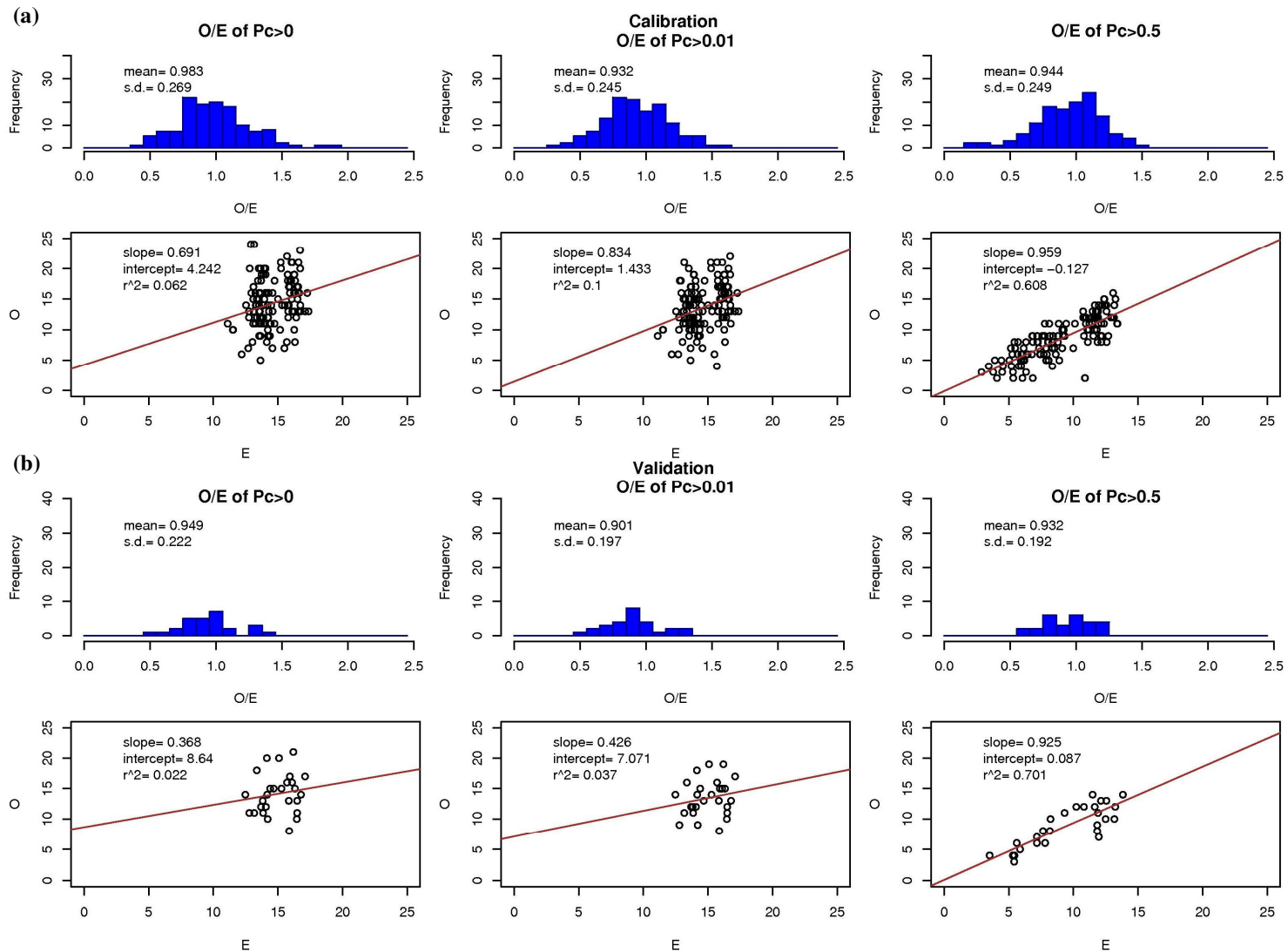


Figure 10. The O/E histograms and E vs. O scatter plots from best ANNA model at the family level with GISHydro2000 predictors. Exhaustive BIC method was used. Neighbor number is 21. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

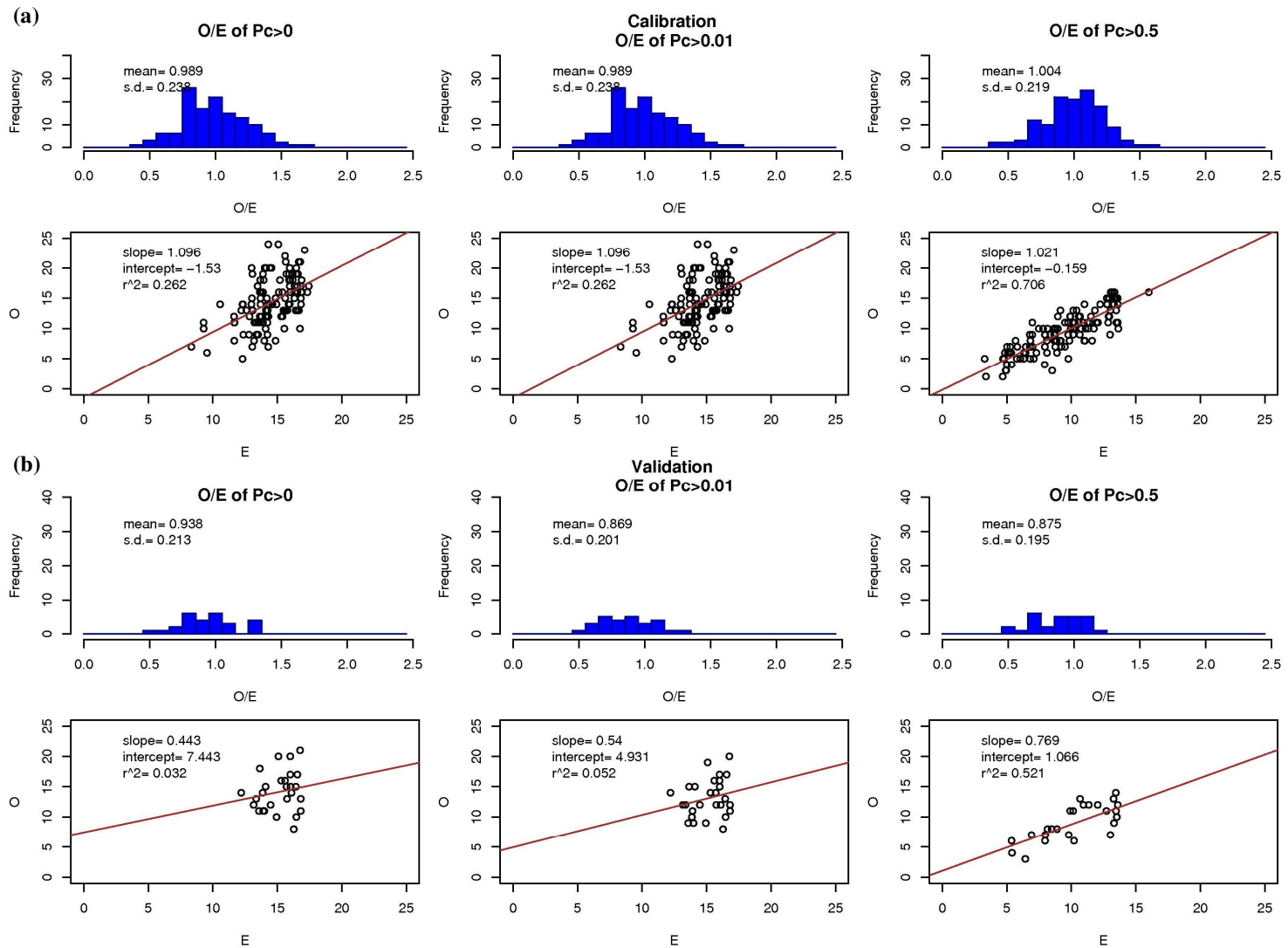


Figure 11. The O/E histograms and E vs. O scatter plots from best BROI model at the family level with GISHydro2000 predictors. Stepwise AIC method was used. $n = 2.5$, $NST = 8$, $\Theta_L = 10\%$, $\Theta_U = 30\%$, $TP = 50\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

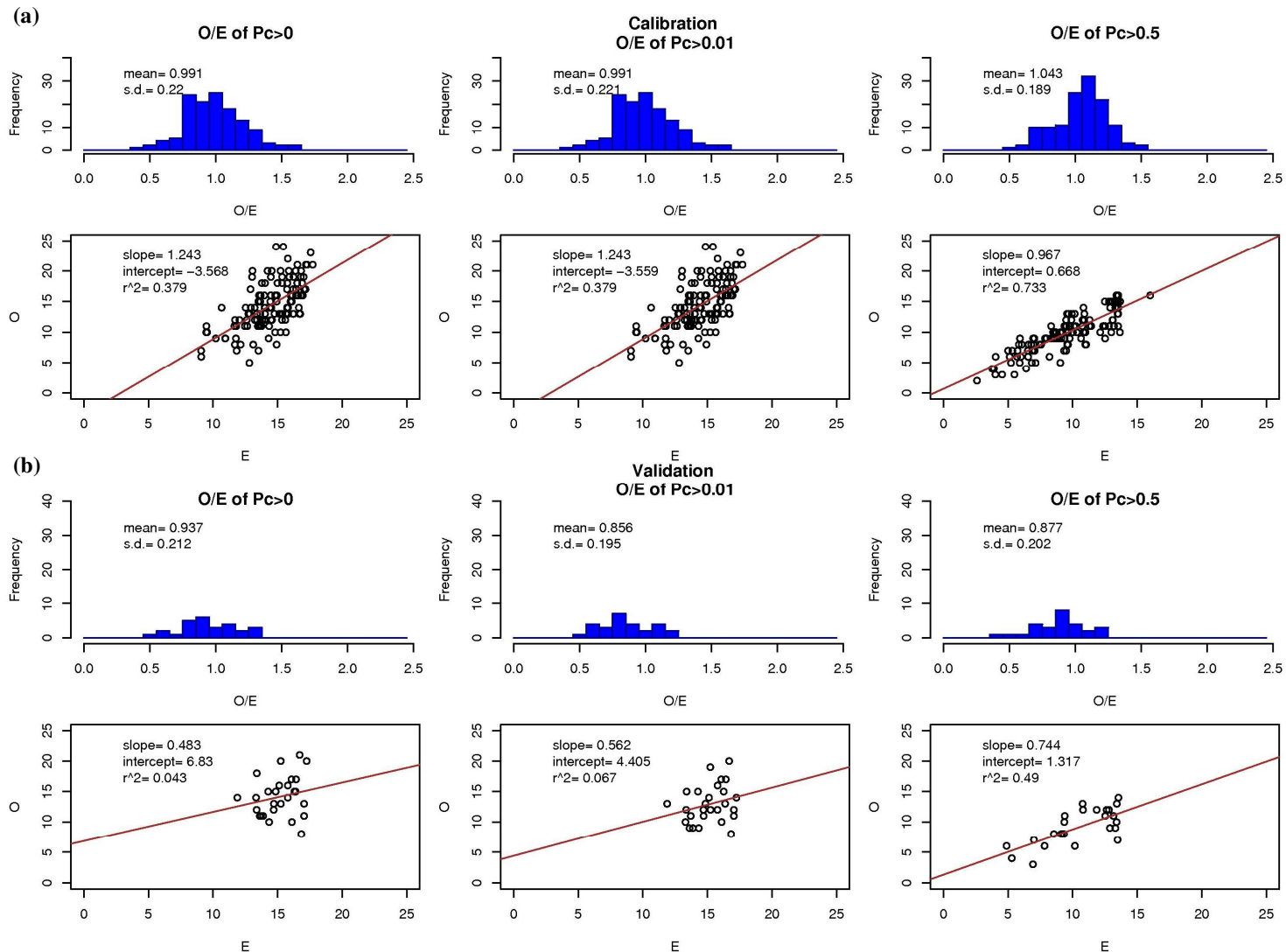


Figure 12. The O/E histograms and E vs. O scatter plots from best NROI model at the family level with GISHydro2000 predictors. Exhaustive AIC method was used. $n=2.5$, NST = 5, $\Theta_L = 10\%$, $\Theta_U = 40\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

5.1.2 Genus Level

Results from the region-style model at the genus level and developed using only GISHydro2000 data are summarized in Table 17. All three predictive schemes (ANNA, BROI, NROI) are presented. This table shows only the results from $p_c = 0.5$. For each model, both calibration and validation results are listed for more selected model performance information. The O/E histograms and E vs. O scatter plots of $p_c = 0$, $p_c = 0.01$, and $p_c = 0.5$ using calibration (129 sites) and validation (27 sites) data sets for each predictive scheme are shown in Figures 13, 14, and 15.

From the calibration results in Table 17, the NROI scheme has the smallest RMSE at 0.274, followed by BROI and ANNA at 0.281 and 0.324, respectively. Then NROI has the highest R^2 at 0.817, with BROI and ANNA at 0.816 and 0.745, respectively. However the validation results show that BROI has the smallest RMSE at 0.288, followed by ANNA and then NROI at 0.296 and 0.302 respectively. BROI also has the highest R^2 at 0.589. For the genus level with only GISHydro2000 predictors, the model using BROI performs the best.

Table 17. The O/E statistics at $P_c=0.5$ from all the Region-style models for the case in genus level with GISHydro2000 predictors

Region Selection scheme	Best Selection Parameter method	Weighting Function parameters	Capture Probability $P_c = 0.5$	mean	s.d.	slope	intercept	R^2	RMSE
ANNA	Exhaustive AIC	Neighbor no. = 16	Calibration	0.88	0.301	0.94	-0.36	0.745	0.324
			Validation	0.904	0.28	0.709	1.263	0.531	0.296
BROI	Exhaustive AIC	$\theta_L = 5\%$, $\theta_U = 30\%$, $TP = 75\%$, $NST = 11$, $n = 2$.	Calibration	1.001	0.281	0.999	-0.001	0.816	0.281
			Validation	0.912	0.274	0.699	1.258	0.589	0.288
NROI	Exhaustive AIC	$\theta_L = 6\%$, $\theta_U = 40\%$, $NST = 4$, $n = 4$.	Calibration	1.081	0.262	1.011	0.479	0.817	0.274
			Validation	0.836	0.253	0.668	1.263	0.445	0.302

Validation data are without the two outlier sites.
Mean and s.d. are from O/E values histogram.
Slope, intercept, and R^2 are from E vs. O scatter plots.
RMSE is defined in equation 33.

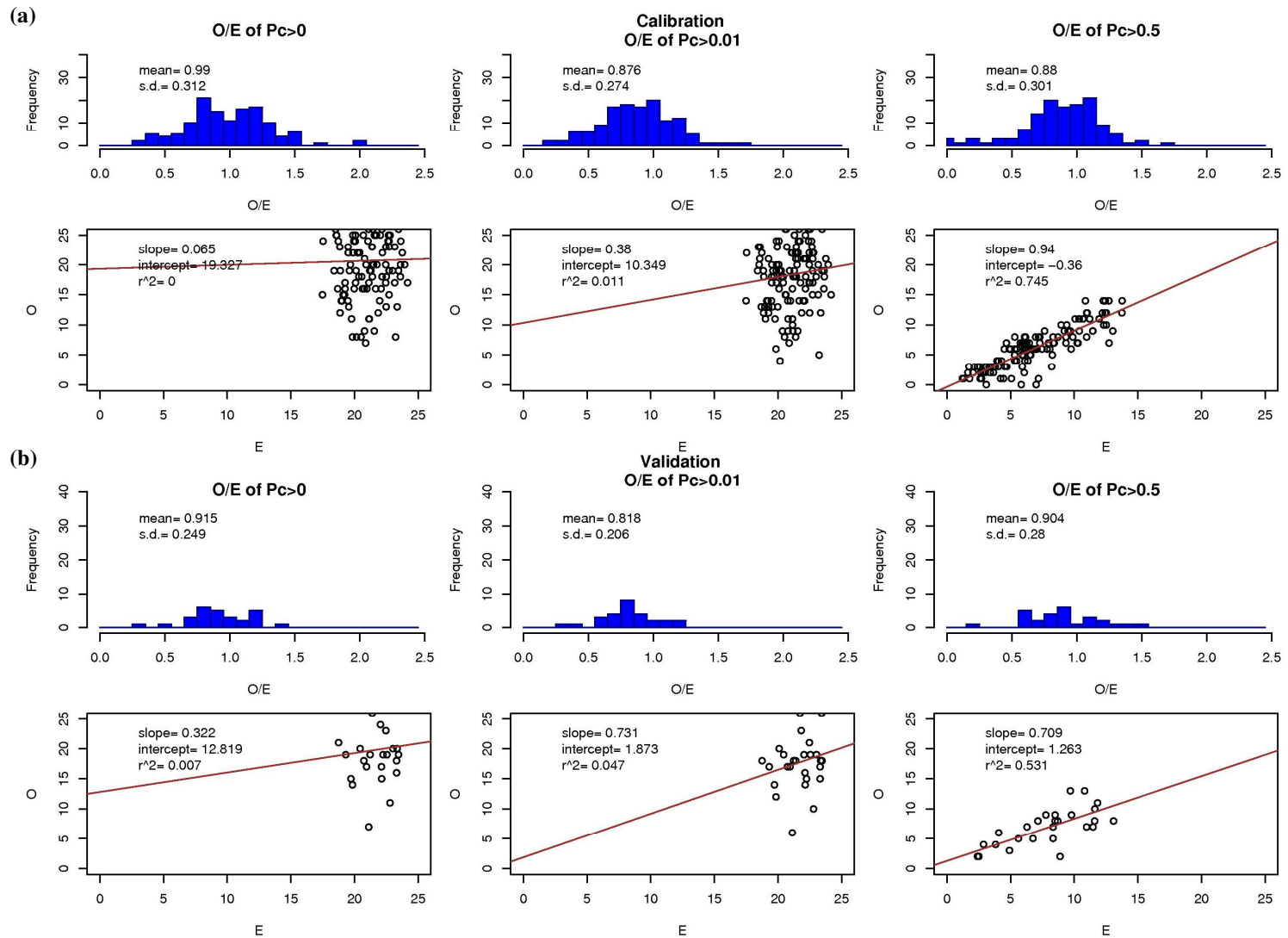


Figure 13. The O/E histograms and E vs. O scatter plots from best ANNA model at the genus level with GISHydro2000 predictors. Exhaustive AIC method was used. Neighbor number = 16. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

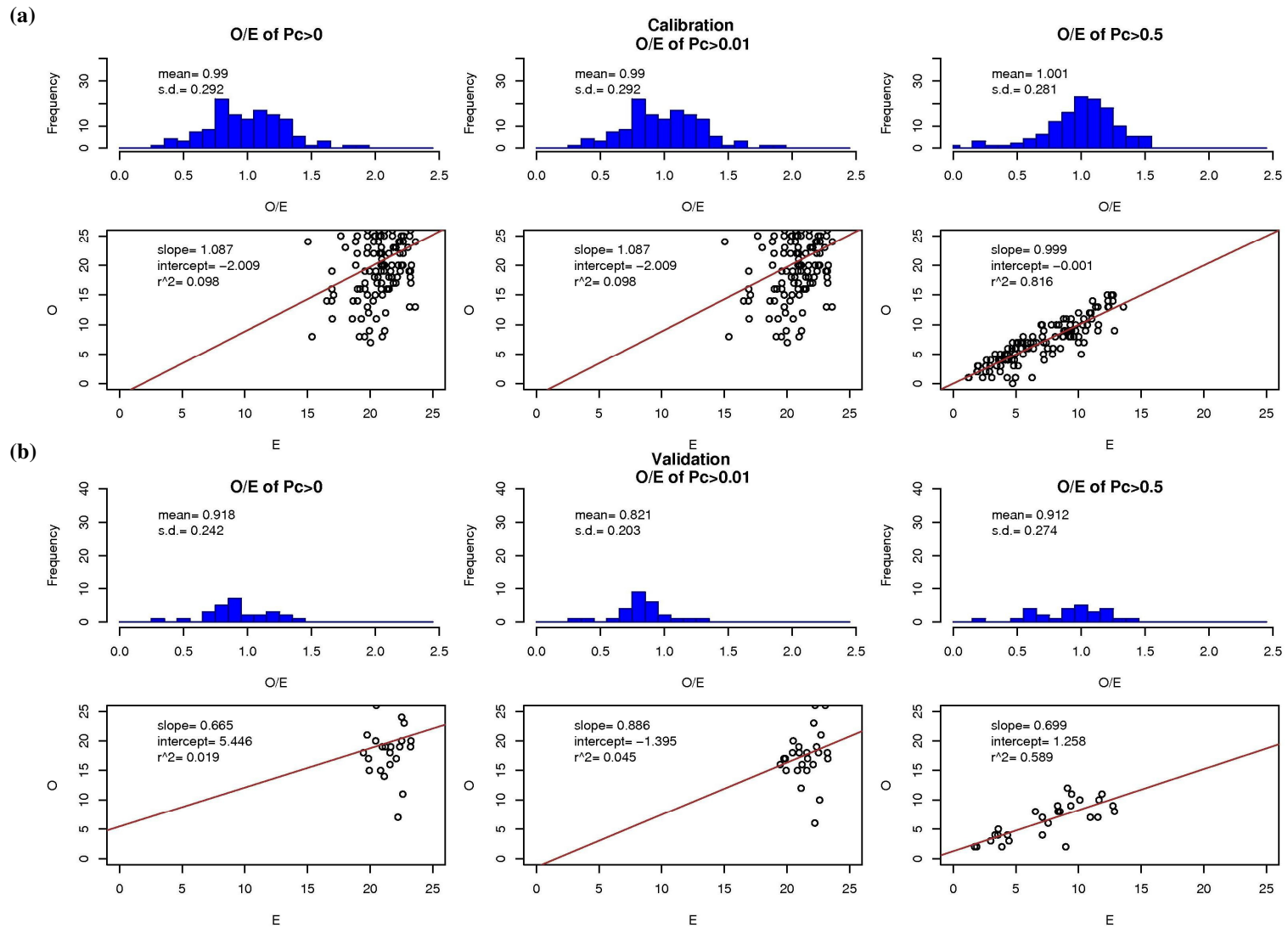


Figure 14. The O/E histograms and E vs. O scatter plots from best BROI model at the genus level with GISHydro2000 predictors. Exhaustive AIC method was used. $n = 2$, $NST = 11$, $\Theta_L = 5\%$, $\Theta_U = 30\%$, $TP = 75\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

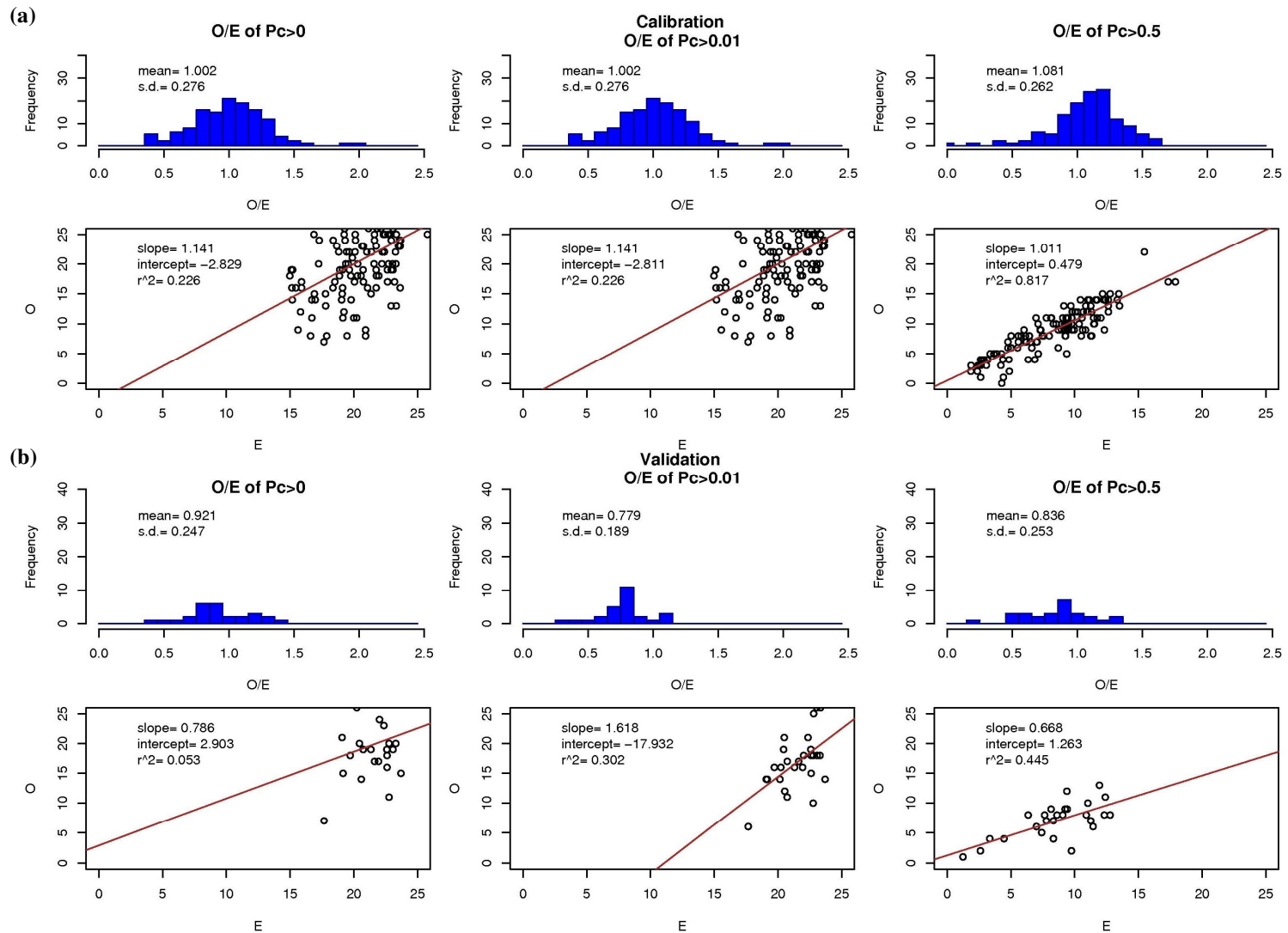


Figure 15. The O/E histograms and E vs. O scatter plots from best NROI model at the genus level with GISHydro2000 predictors. Exhaustive AIC method was used. $n = 4$, $\Theta_L = 6\%$, $\Theta_U = 40\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

5.2 MODELS WITH ALL PREDICTORS

5.2.1 Family Level

Results from the region-style model at the family level and developed using both GISHydro2000 and MBSS data are summarized in Table 18. All three predictive schemes (ANNA, BROI, NROI) are presented. The table shows only the results from $p_c = 0.5$. For each model, both calibration and validation results are listed for more model performance information. The O/E histograms and E vs. O scatter plots using calibration (124 sites) and validation (27 sites) data sets for each predictive scheme are shown in Figures 16, 17, and 18.

From the calibration results in Table 18, the NROI scheme has the smallest RMSE at 0.179, followed by BROI and ANNA at 0.202 and 0.238, respectively. NROI has the highest R^2 at 0.768, with BROI and ANNA having R^2 values at 0.708 and 0.675, respectively. The validation results show that ANNA has the smallest RMSE at 0.201, followed by NROI and BROI at 0.232 and 0.236, respectively. All the RMSE from difference predictive schemes and datasets have close values, except NROI has the lower RMSE for the calibration data set. The model using the ANNA scheme has the highest R^2 at 0.699 for the validation data set. It is hard to decide which of the Region-style model is the best. In terms of prediction, they all perform equally.

Table 18. The O/E statistics at $P_c=0.5$ from all the Region-style models for the case in the family level with all predictors

Region Selection scheme	Best Selection Parameter method	Weighting Function parameters	Capture Probability $P_c = 0.5$	mean	s.d.	slope	intercept	R^2	RMSE
ANNA	Exhaustive BIC	Neighbor no. = 34	Calibration	0.969	0.236	0.977	-0.069	0.675	0.238
			Validation	0.964	0.198	0.896	0.628	0.699	0.201
BROI	Exhaustive AIC/ Stepwise AIC	$\theta_L = 5\%$, $\theta_U = 25\%$, $TP = 50\%$, $NST = 8$, $n = 3.5$.	Calibration	1.007	0.202	0.989	0.135	0.708	0.202
			Validation	0.875	0.2	0.655	2.117	0.477	0.236
NROI	Exhaustive AIC/ Stepwise AIC	$\theta_L = 6\%$, $\theta_U = 40\%$, $NST = 6$, $n = 3$.	Calibration	1.028	0.177	1.009	0.149	0.768	0.179
			Validation	0.872	0.193	0.745	1.191	0.597	0.232

Validation data are without the two outlier sites.
Mean and s.d. are from O/E values histogram.
Slope, intercept, and R^2 are from E vs. O scatter plots.
RMSE is defined in equation 33.

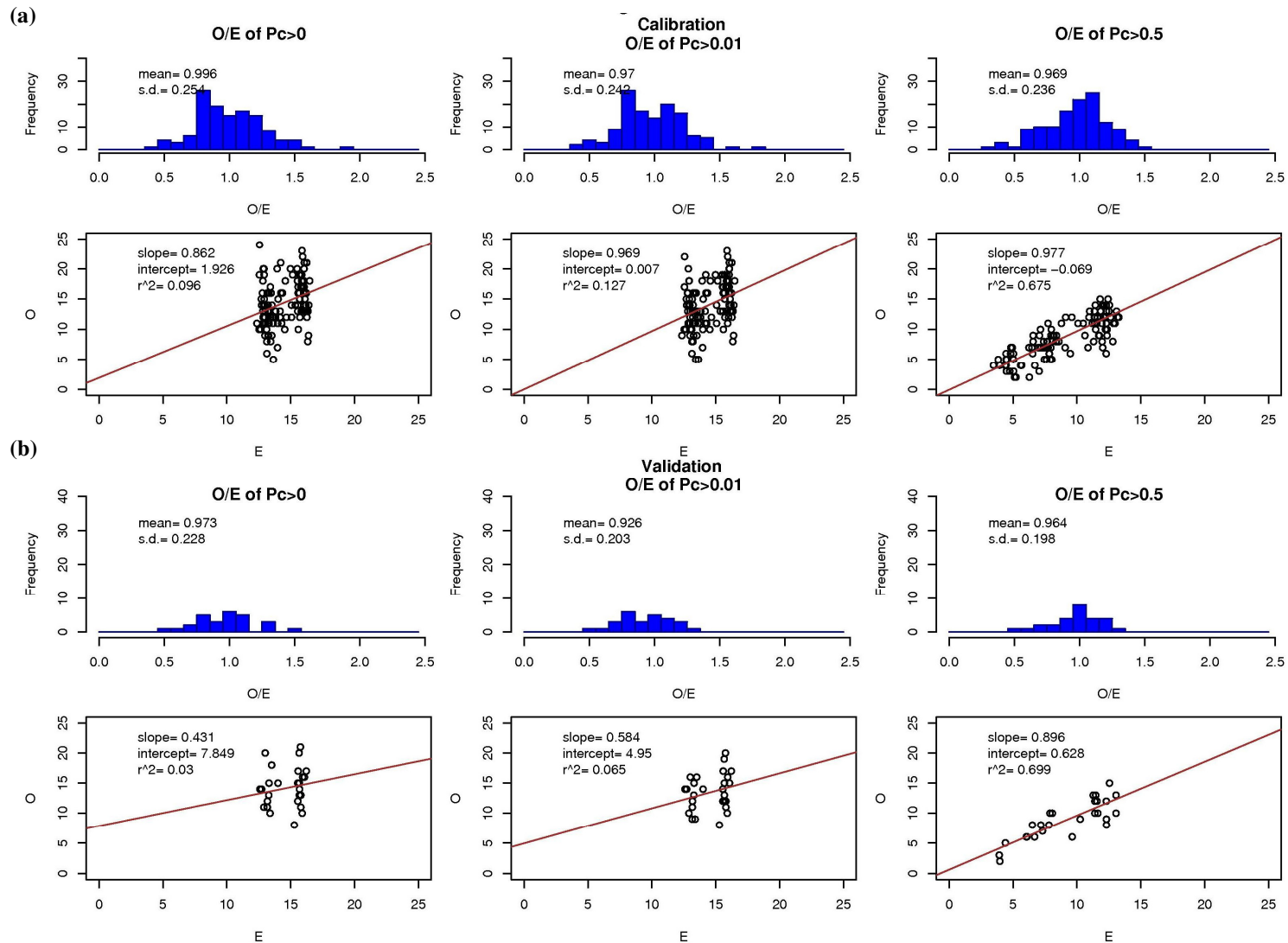


Figure 16. The O/E histograms and E vs. O scatter plots from best ANNA model at the family level with All predictors. Exhaustive BIC method was used. Neighbor number = 34. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

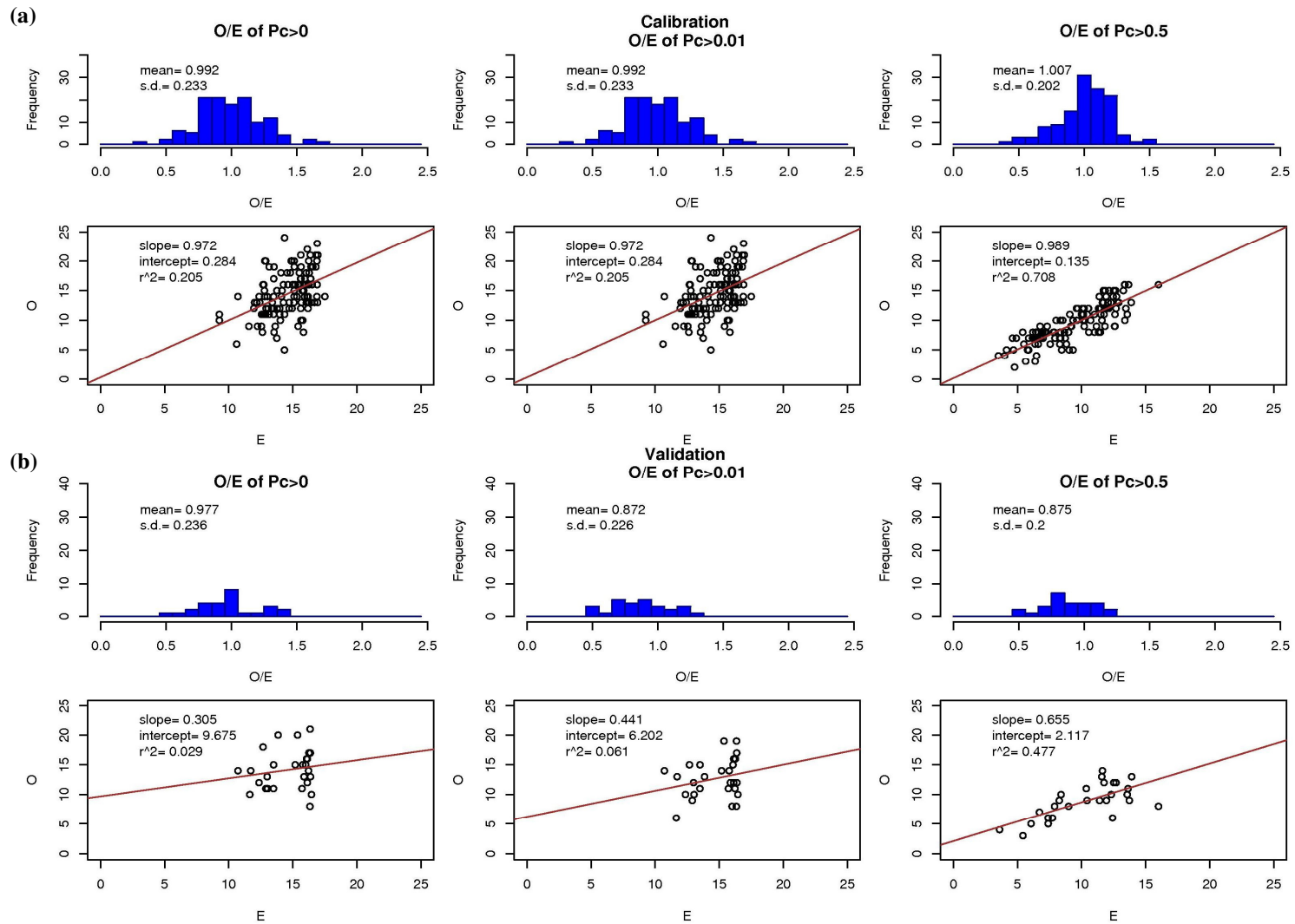


Figure 17. The O/E histograms and E vs. O scatter plots from best BROI model at the family level with All predictors. Exhaustive AIC method was used. $n = 3.5$, NST = 8, $\Theta_L = 5\%$, $\Theta_U = 25\%$, TP = 50%. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

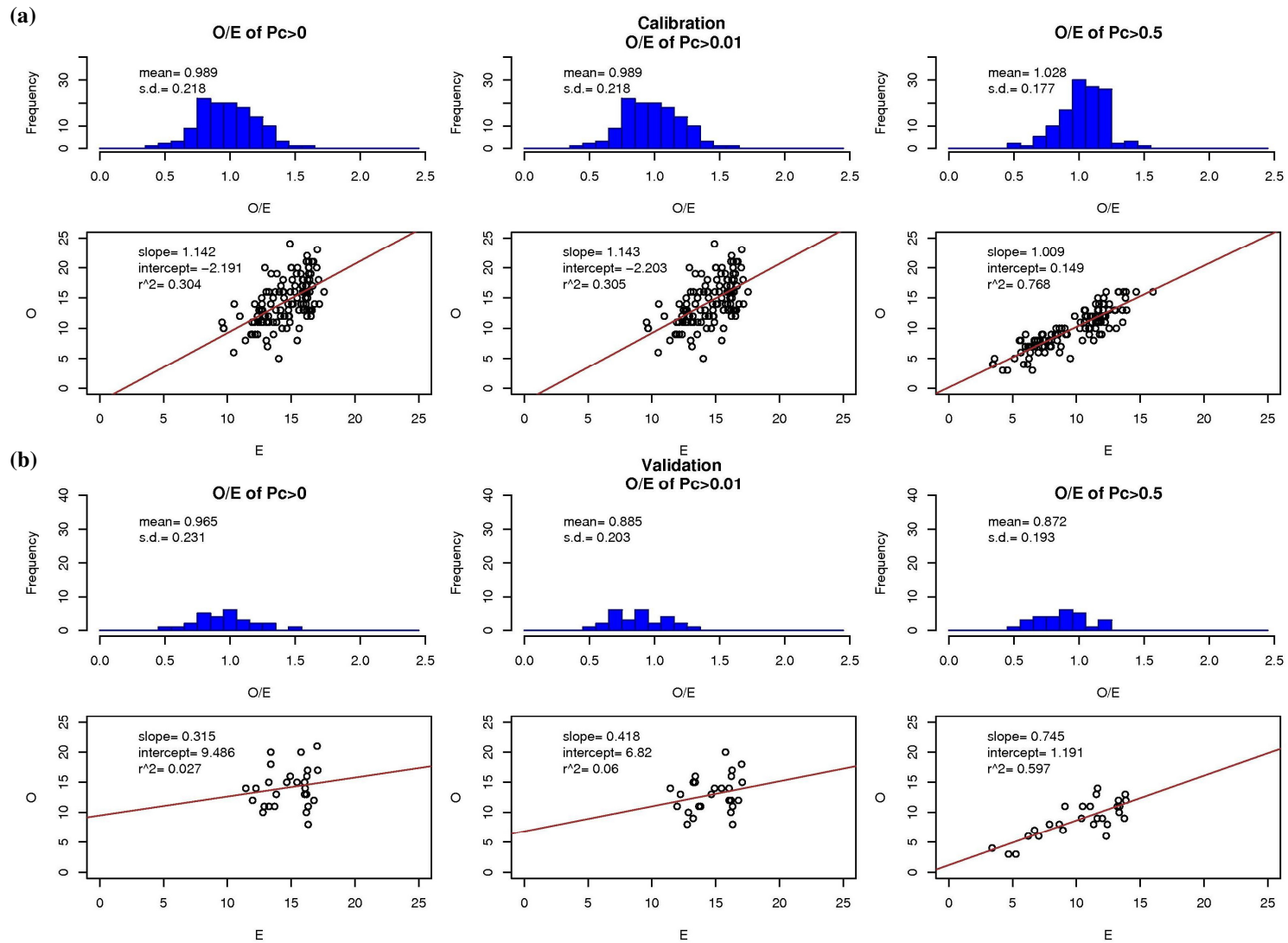


Figure 18. The O/E histograms and E vs.O scatter plots from best NROI model at the family level with All predictors. Exhaustive AIC method was used. $n=3$, NST = 6, $\Theta_L = 6\%$, $\Theta_U = 40\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

5.2.2 Genus Level

Results from the region-style model at the genus level and developed using both GISHydro2000 and MBSS data are summarized in Table 16. All three predictive schemes (ANNA, BROI, NROI) are presented. This table shows only the results from $p_c = 0.5$. For each model, both calibration and validation results are listed for more model performance information. The O/E histograms and E vs. O scatter plots using calibration (124 sites) and validation (27 sites) data sets for each predictive scheme are shown in Figures 19, 20, and 21.

From the calibration results, the models from both NROI and BROI schemes have the smallest RMSE at 0.253, and ANNA has it as 0.276. NROI has the highest R^2 at 0.765, with BROI and ANNA having R^2 value at 0.754 and 0.745, respectively. The validation results show NROI with the smallest RMSE at 0.288 and with ANNA and BROI at 0.290 and 0.293 respectively. NROI has the highest R^2 as 0.676. For the genus level with all predictors, the NROI scheme results in the best model for both calibration results and validation results.

Table 19. The O/E statistics at $P_c=0.5$ from all the Region-style models for the case in genus level with all predictors

Region Selection scheme	Best Selection Parameter method	Weighting Function parameters	Capture Probability $P_c = 0.5$	mean	s.d.	slope	intercept	R^2	RMSE
ANNA	Stepwise AIC	Neighbor no. = 25	Calibration	0.923	0.265	0.995	-0.411	0.745	0.276
			Validation	0.959	0.287	0.977	-0.185	0.63	0.290
BROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 25\%$, $TP = 50\%$, $NST = 13$, $n = 2.5$.	Calibration	1.05	0.248	1.066	-0.072	0.754	0.253
			Validation	0.943	0.287	0.943	-0.036	0.616	0.293
NROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 40\%$, $NST = 7$, $n = 4$.	Calibration	1.065	0.245	1.07	-0.024	0.765	0.253
			Validation	0.872	0.258	1.02	-1.058	0.676	0.288

Validation data are without the two outlier sites.
Mean and s.d. are from O/E values histogram.
Slope, intercept, and R^2 are from E vs. O scatter plots.
RMSE is defined in equation 33.

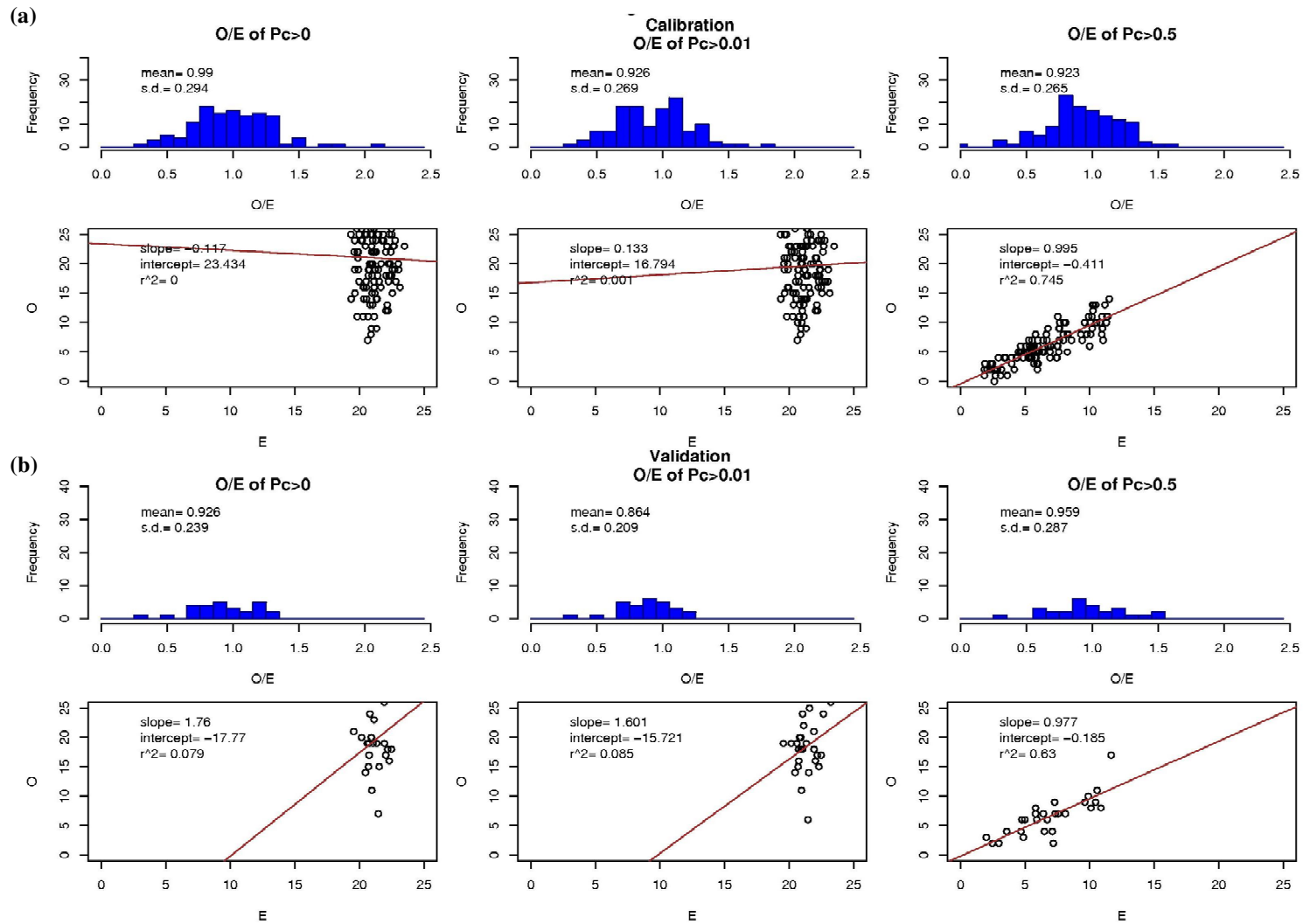


Figure 19. The O/E histograms and E vs. O scatter plots from best ANNA model at the genus level with All predictors. Stepwise AIC method was used. Neighbor number = 25. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

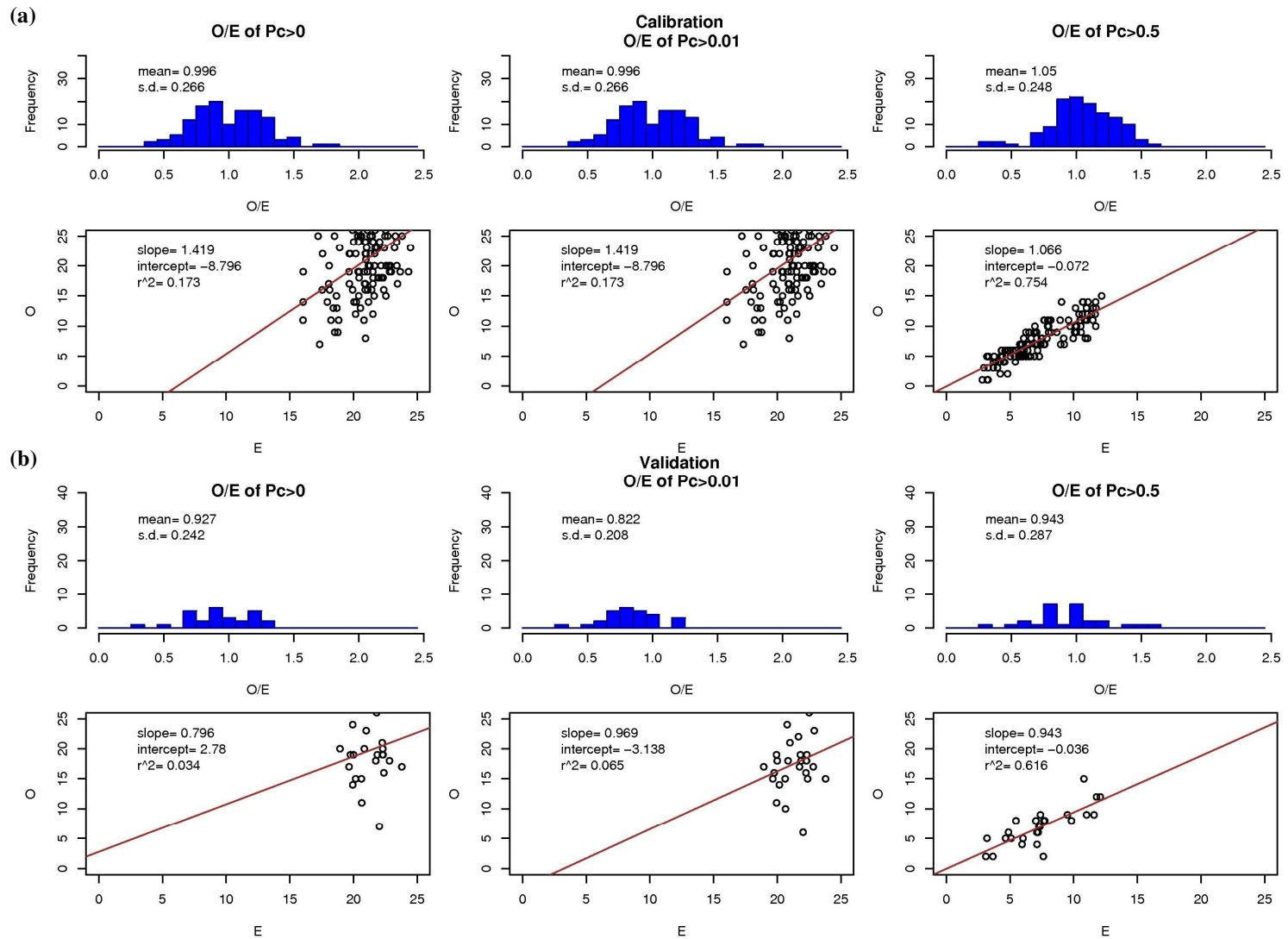


Figure 20. The O/E histograms and E vs. O scatter plots from best BROI model at the genus level with All predictors. Stepwise AIC method was used. $n = 2.5$, NST = 13, $\Theta_L = 10\%$, $\Theta = 25\%$, TP = 50%. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

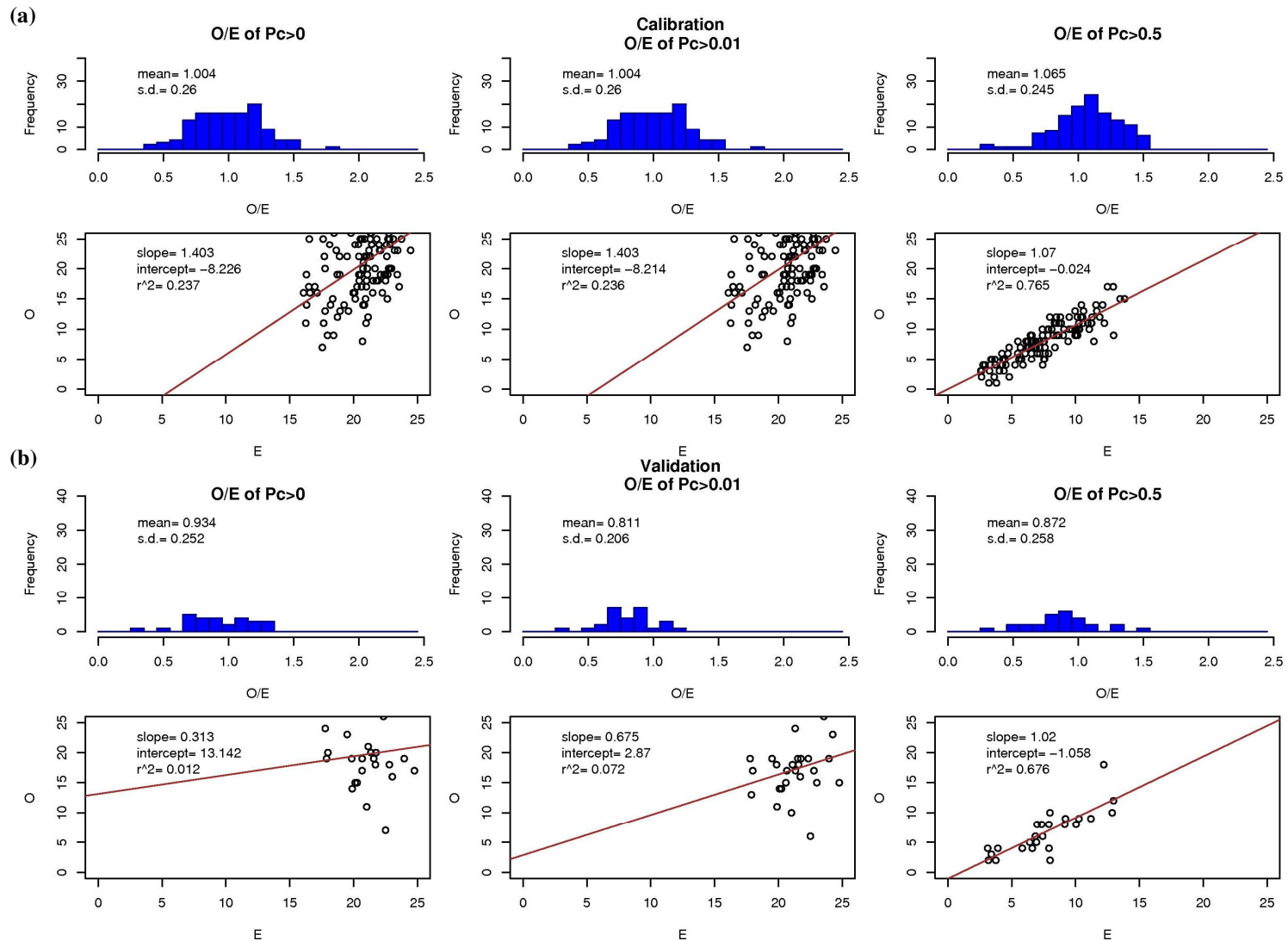


Figure 21. The O/E histograms and E vs. O scatter plots from best NROI model at the genus level with All predictors. Stepwise AIC method was used. $n = 4$, $NST = 7$, $\Theta_L = 10\%$, $\Theta_U = 40\%$. (a) Calibration data, (b) Validation data. The cases of capture probability $P_c > 0$, $P_c > 0.01$, and $P_c > 0.5$ are shown from the left to right.

CHAPTER 6: DISCUSSION AND CONCLUSIONS FROM THE MODEL RESULTS

6.1 COMPARING THE REGION-STYLES

When comparing RMSE values, predictions for the family taxon level from Region-style models are better than those for the genus level. The RMSE value difference between these two taxon levels varies from 0.036 to 0.092. Family level prediction is better than genus level prediction by 13.8% to 31.1%. In the Region-style models, ANNA clearly defines the neighbor number for different taxon levels with different sets of predictors. BROI and NROI on the other hand define the region of influence using percentage thresholds from the entire set of data-collection sites. Each site thus has a different number of reference neighbors, depending on the similarity of reference sites.

6.1.1 Family Level

Tables 16 and 18 in Chapter Five show the family level results with the different sets of predictors. It can be seen that adding MBSS predictors to GISHydro2000 predictors decreases the RMSE value in all three Region-style models. In the calibration data, the RMSE values decrease by 0.015 or 0.017. In the validation data, the decrements are smaller than in the calibration data, only by 0.003 to 0.005. Even though the magnitude of improvement is surprisingly small, this supports the notion that adding predictors, especially on-site physical feature predictors, helps models predict better. When looking at Table 13 in Chapter Four for

the parameter selection for the family level, it is found that the presence of BOULDLT2 and COBBLE is critical to improving Region-style models. MAXDEPTH and DATE_SPR are minor factors. BOULDLT2 and COBBLE represent substrates in the stream bed. With substrate information, Region-style models predict the presence of macroinvertebrates better. This further verifies the importance of stream bed substrates to the macroinvertebrate species.

Tables 20 and 21 summarize the number of reference sites from different selection schemes for the family level. ANNA used 21 reference sites when using GISHydro2000 predictors and 34 reference sites when using all predictors. The reference neighbors for test sites using BROI and NROI schemes are automatically selected according to the physical similarity (Sections 2.3.2, 2.3.3). The number of neighbors for each test thus varies. BROI and NROI generally use fewer reference sites than ANNA to predict taxon appearance no matter which predictor set is used. In the case of the GISHydro2000 predictors, BROI used 14 reference sites (median value) for calibration data and 18 reference sites for validation data. In all the predictors' cases, it used 10 reference sites for calibration and 11 reference sites for validation. NROI uses 14 and 18 reference sites, respectively, for calibration and validation when using GISHydro2000 predictors. It used 12 and 13 reference sites, respectively, for calibration and validation when using all of the predictors. The reference site results are compared with prediction results in Tables 16 and 18 in Chapter Five. For the calibration data, NROI had better prediction precision using either just the GISHydro2000 predictors or using all the predictors. Both RMSE values were under 0.2. The means are a little above 1.000. The standard deviations

are 0.189 and 0.177 respectively. For validation data, ANNA had the best prediction precision using either GISHydro2000 or using all the predictors. The RMSE values from ANNA were both around 0.2. The means were less than 1. The standard deviations were both around 0.190. ANNA performed best with validation data but worst with calibration data, while NROI performed best with calibration data but worse with validation data. There is no absolute relation between the number of reference sites and the prediction precision for the family level macroinvertebrate prediction.

Table 20. Region-Style Model: number of reference sites for the family level using only GISHydro2000 predictors

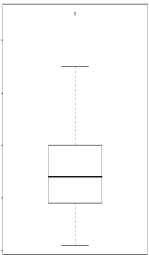
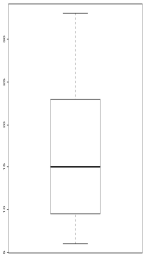
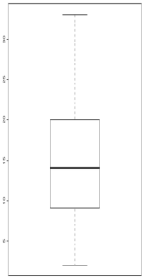
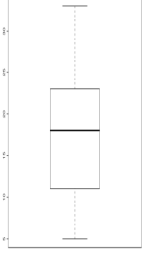
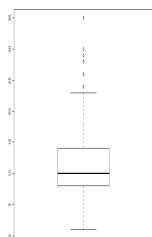
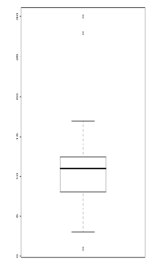
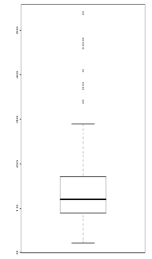
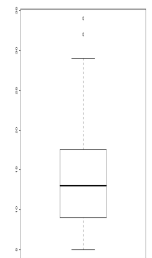
Region Selecting scheme	Best Selecting Parameter method	Weighting Function parameters	Number of Reference Sites					
			mean	max	min	median	Box Plot	
ANNA	Stepwise AIC	Neighbor no. = 21	Calibration				21	
			Validation					
BROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 30\%$, $TP = 50\%$, $NST = 8$, $n = 2.5$.	Calibration	15.26	45	1	14	
			Validation	16.81	33	6	15	
NROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 40\%$, $NST = 5$, $n = 2.5$.	Calibration	15.28	33	2	14	
			Validation	17.56	33	5	18	

Table 21. Region-Style Model: number of neighbors' selection for the family level using both GISHydro2000 and MBSS predictors

Region Selecting scheme	Best Selecting Parameter method	Weighting Function parameters	Number of Reference Sites					
			mean	max	min	median	Box Plot	
ANNA	Exhaustive BIC	Neighbor no. = 34	Calibration					
			Validation				34	
BROI	Exhaustive AIC/ Stepwise AIC	$\theta_L = 5\%$, $\theta_U = 25\%$, $TP = 50\%$, $NST = 8$, $n = 3.5$.	Calibration	11.48	35	1	10	
			Validation	11.44	34	1	11	
			Calibration	14.48	54	2	12	
			Validation	14.37	34	5	13	

6.1.2 Genus Level

Tables 17 and 19 in Chapter Five shows genus level results with different predictor sets. Combining MBSS predictors with GISHydro2000 predictors decreases the RMSE values in all three Region-style models over using just the GISHydro2000 predictors. In the calibration data, the RMSE values decrease by 0.048 to 0.021, whereas in the validation data, the decrements are smaller, only by 0.006 or 0.014. This result indicates that adding predictors, especially on-site physical feature predictors, helps models predict better, just as was found using family level data. Table 13 in Chapter Four also presents parameter selection results for the genus level. BOULDLT2, COBBLE, and DATE_SPR are critical factors in improving Region-style models. AVGWID, MAXDEPTH, and JDAY are minor factors. Contrary to the family level prediction, the predictor selection for genus level uses almost every on-site predictor provided except BEDROCK. This means that on-site predictors are required to achieve better prediction at the genus level.

Tables 22 and 23 summarize the data that represent the number of reference sites from different selection schemes at the genus level. ANNA uses 16 reference sites when using GISHydro2000 predictors and 25 reference sites when using all predictors. When using GISHydro2000 predictors, BROI uses more reference sites than ANNA, while NROI uses fewer reference sites than ANNA. Both BROI and NROI use fewer reference sites than ANNA when using all predictors. Comparing these results with Tables 16 and 18 in Chapter Five, when using GISHydro2000 predictors, BROI performs the best for validation data and second best for the calibration data. When using all the predictors, NROI performs about the same for

both calibration and validation data. In these cases, NROI predicts better than the other two schemes and with fewer reference sites. This indicates that a good prediction scheme can work well even when using fewer reference sites.

It is clear that NROI is the best choice for modeling the genus level when using all predictors. However, it would be hard to conclude from the number of used reference sites which prediction scheme is the best in general because the results vary with both level (genus vs. family) and data sets (calibration vs. validation).

The family level predictive models has better prediction precision than genus level model. The genus level model, however, catch the trend fairly well (Section 9.1). The possible reason may be caused by the sampling process that MBSS adapts. Only 100 individuals were subsampled from the each monitoring sample. The genus level composition cannot be described completely for limited individual counts, which cause the lower expected probabilities. According to the consultation with Dr. William Lamp, at stream side it is easiler to identify macroinvertebrate to family level than genus level. But to discriminate or interprete the stream condiction, genus may give more information for restoration. The prediction ability is different from discrimination.

Table 22. Region-Style Model: number of reference sites at the genus level with GISHydro2000 predictors


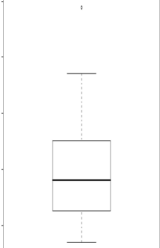
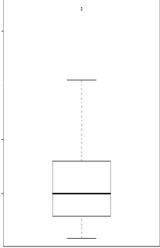
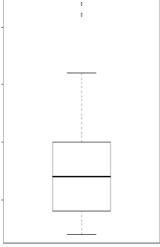
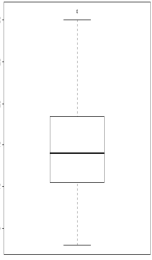
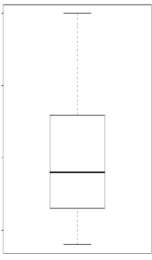
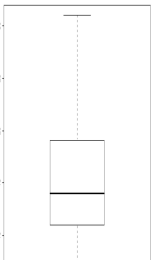
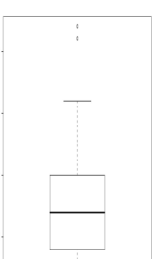
Region Selecting scheme	Best Selecting Parameter method	Weighting Function parameters	Number of Reference Sites					
			mean	max	min	median	Box Plot	
ANNA	Exhaustive AIC	Neighbor no. = 16	Calibration				16	
			Validation					
BROI	Exhaustive AIC	$\theta_L = 5\%$, $\theta_U = 30\%$, $TP = 75\%$, $NST = 11$, $n = 2$.	Calibration	18.06	51	1	17	
			Validation	20.33	49	7	18	
NROI	Exhaustive AIC	$\theta_L = 6\%$, $\theta_U = 40\%$, $NST = 4$, $n = 4$.	Calibration	11.70	44	2	10	
			Validation	16.93	44	4	14	

Table 23. Region-Style Model: number of neighbors' selection at the genus level with all predictors

Region Selecting scheme	Best Selecting Parameter method	Weighting Function parameters	Number of Reference Sites					
			mean	max	min	median	Box Plot	
ANNA	Stepwise AIC	Neighbor no. = 25	Calibration				25	
			Validation					
BROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 25\%$, $TP = 50\%$, $NST = 13$, $n = 2.5$.	Calibration	15.05	31	3	14	
			Validation	15.15	25	9	14	
NROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 40\%$, $NST = 7$, $n = 4$.	Calibration	15.53	31	7	14	
			Validation	15.52	27	7	15	

6.1.3 Individual validation reference site comparison

With the help of GISHydro2000, the neighbor selection from the Region-style models can be shown on the map. The locations of 27 validation reference sites were loaded into Geographic Information System (GIS) to run the Region-style models in GISHydro2000. The details of this application will be explained in Chapter Seven. The selection results maps from these 27 validation reference sites were also obtained to observed if there is any pattern and shown in Appendix D. The red points in the graphs are those individual validation reference sites that were used as test sites. The colored irregular polygons are the delineated watershed when using test sites as watershed outlets. The blue points are calibration reference sites. The points highlighted in yellow are the selected sites used to predict benthic macroinvertebrates. These sites were selected automatically based on their physical features similarity to the test sites. From these graphs, some of the selected reference sites are found at close distances to test sites and some of the selected reference sites are far from the test site locations. No particular pattern was found in this observation. More detail in comparing the neighbor selections with different Region-style models will be described in Chapter Eight using restoration sites as examples.

In Sections 6.1.1 and 6.1.2, only median of reference site numbers was examined to determine the relationship with prediction. Here, the numbers of reference neighbors selected by individual validation sties are listed in Table 24 (family level) and Table 25 (genus level), as well as the observed values (O) and prediction probabilities (E). The prediction from each model was examined to determine if the prediction was close to the observed reference condition and if the

number of reference sites used affect the prediction and observed. At site ZEKI-106-R-2001 in Table 24, the ANNA scheme used 21 sites to predict the probable number of taxa as 5.61, the BROI scheme used 8 sites to predict the probable taxa as 6.91, and the NROI used only 7 sites to predict the probable number of taxa present as 7.0, which meets the observed value exactly. At site SAVA-225-S-2001 in Table 25, the ANNA scheme used 16 sites to predict the taxa present as 7.67, the BROI used 49 sites to predict 7.77, and the NROI used 44 sites to predict 8.33, but the observed value on-site was just 7. From the relationship between number of reference sites used and prediction precision, using more sites did not improve the prediction. The number of reference sites used is not equal to the reference site quality and thus not equal to the prediction precision.

In these two tables, the prediction values from the BROI and NROI are slightly higher than the ANNA's value for each site. This was also observed from the results in Tables 16, 17, 18, and 19 in Chapter Five. The mean of O/E from BROI and NROI were slightly higher than one, while ANNA's mean was slightly lower than one. As defined in Section 2.1. Step5, the observed value (O) is selected based on the taxon prediction (E). If the probability prediction is higher, it affects our selection of observed data. The higher taxon probability prediction from BROI and NROI not only raised the expected taxon appearances (E) at each site, more observed taxa were selected into observed data (O) as well. When adding the observed values under each O column, it shows in both tables that more observed taxon data were selected due to higher probability prediction using BROI and NROI schemes. The predictive models using BROI and NROI schemes expect more taxon appearance in stream reference

condition.

Table 24. Region-Style number of neighbors' selection for validation sites in family level with GISHydro2000 predictors (Pc = 0.5)

Station ID	Family								
	ANNA			BROI			NROI		
	Total Number of Reference Sites	E	O	Total Number of Reference Sites	E	O	Total Number of Reference Sites	E	O
AL-A-318-126-95	21	11.90	11	17	13.46	11	20	12.52	11
AL-A-480-205-96	21	10.81	12	18	12.04	12	5	12.63	12
BA-P-025-102-96	21	7.81	6	10	10.20	6	11	10.17	6
BA-P-077-315-96	21	9.31	11	11	9.91	11	13	9.35	10
BA-P-206-108-95	21	8.18	8	15	8.88	8	14	9.32	8
BRIG-307-R-2000	21	8.23	10	8	10.14	11	12	9.38	11
CA-S-119-210-97	21	3.51	4	8	5.36	6	8	4.88	6
CASS-111-R-2000	21	11.96	7	33	13.03	7	33	13.48	7
CH-S-020-322-95	21	5.31	4	7	6.43	3	7	6.92	3
CN-N-024-113-96	21	5.41	3	15	5.40	4	24	5.29	4
GA-A-002-312-96	21	13.23	12	33	12.72	11	32	13.16	11
GA-A-008-213-96	21	13.17	10	19	13.52	10	19	13.40	10
GA-A-121-210-96	21	12.49	10	25	13.52	10	23	12.88	9
GA-A-159-202-96	21	11.65	12	18	13.60	12	18	12.81	12
GA-A-185-309-95	21	7.17	7	15	8.15	8	7	8.54	8
LOGU-106-R-2002	21	7.64	8	15	8.47	8	18	9.10	8
MO-P-064-328-97	21	7.17	6	9	9.79	7	12	9.32	8
SAVA-116-R-2002	21	13.83	14	23	13.27	13	23	13.42	13
SAVA-206-R-2002	21	12.55	13	23	11.35	12	22	11.86	12
SAVA-225-S-2001	21	11.84	9	30	13.30	9	31	13.31	9
SAVA-225-S-2002	21	11.84	8	30	13.30	9	31	13.31	9
SAVA-312-R-2002	21	12.13	13	25	13.45	14	23	13.55	14
SM-S-039-127-95	21	5.43	4	6	7.97	7	9	9.15	8
TA-N-001-210-97	21	5.87	5	6	7.96	6	11	7.82	6
YOUG-203-C-2000	21	10.22	12	15	10.94	12	29	10.80	12
YOUG-432-S-2001	21	11.48	14	12	10.67	13	12	10.77	13
ZEKI-106-R-2001	21	5.61	6	8	6.91	7	7	7.00	7
Total			239			247			247

*All these 27 reference sites can be looked up from MBSS data set with the listed Station ID in "Site name". <http://mddnr.chesapeakebay.net/mbss/search.cfm>

Table 25. Region-Style number of neighbors' selection for validation sites in genus level with GISHydro2000 predictors (Pc = 0.5)

Genus

Station ID*	ANNA			BROI			NROI		
	Total Number of Reference Sites	E	O	Total Number of Reference Sites	E	O	Total Number of Reference Sites	E	O
AL-A-318-126-95	16	7.39	6	34	7.57	6	20	9.06	8
AL-A-480-205-96	16	11.52	8	20	10.94	7	32	10.91	8
BA-P-025-102-96	16	6.81	8	11	6.55	8	13	6.35	8
BA-P-077-315-96	16	5.12	6	11	8.36	8	14	8.15	9
BA-P-206-108-95	16	5.36	3	29	3.88	2	7	11.44	6
BRIG-307-R-2000	16	7.21	7	13	9.38	9	14	9.25	9
CA-S-119-210-97	16	2.60	4	9	3.56	4	6	7.00	6
CASS-111-R-2000	16	8.28	2	22	8.98	2	9	9.75	2
CH-S-020-322-95	16	1.80	1	28	1.70	2	42	1.23	1
CN-N-024-113-96	16	2.32	3	19	1.87	2	20	2.60	2
GA-A-002-312-96	16	10.26	11	37	9.45	11	7	11.94	13
GA-A-008-213-96	16	12.18	9	15	12.87	8	18	12.78	8
GA-A-121-210-96	16	11.99	8	15	12.74	9	20	12.32	8
GA-A-159-202-96	16	11.16	7	18	11.54	7	21	11.26	7
GA-A-185-309-95	16	4.97	3	32	4.45	3	5	7.78	7
LOGU-106-R-2002	16	6.38	7	12	8.51	8	16	8.61	8
MO-P-064-328-97	16	3.58	4	7	4.32	4	8	4.46	4
SAVA-116-R-2002	16	12.59	11	14	11.87	11	13	12.40	11
SAVA-206-R-2002	16	10.13	10	18	10.10	10	20	9.41	9
SAVA-225-S-2001	16	7.67	7	49	7.11	7	44	8.33	7
SAVA-225-S-2002	16	7.67	4	49	7.11	4	44	8.33	4
SAVA-312-R-2002	16	10.92	10	19	11.63	10	21	11.03	10
SM-S-039-127-95	16	2.25	1	13	2.98	3	4	8.33	4
TA-N-001-210-97	16	3.76	4	10	3.33	4	14	3.34	4
YOUG-203-C-2000	16	9.28	9	12	8.33	9	12	7.64	8
YOUG-432-S-2001	16	9.95	14	18	9.11	12	8	9.39	12
ZEKI-106-R-2001	16	2.43	4	15	3.59	5	5	7.41	5
Total			171			175			188

*All these 27 reference sites can be looked up from MBSS data set with the listed Station ID in "Site name". <http://mddnr.chesapeakebay.net/mbss/search.cfm>

6.1.4 Residual Analyses

To understand the relationship between O and E from the Region-style models, two residual analyses were performed to observe the potential patterns. Here only the case with capture probability $P_c = 0.5$ is discussed and compared. First, the residual is defined as:

$$\text{Residual} = E_{at P_c=0.5} - O_{at P_c=0.5} \quad (34)$$

The Residual vs. E from difference model and cases are drawn in Figures 22, 23, 24, and 25.

Residual Analysis 1.

The number above the line of residual value = 0 was counted, as well as the number of reference sites below the line of residual value = 0. Those above zero shows the expected taxon probabilities at these reference site were overestimated, while the number below zero shows the expected taxon probabilities at these reference sites were underestimated. A well performing predictive model should have equally same number of reference site above and below residual value = 0. The Analysis results from the best models of difference predictive schemes are summarized under Residual Analysis 1 in Tables 26, 27, 28 and 29.

From the tables, they show that ANNA scheme in different cases has more number of reference sites above the residual = 0 than below the residual = 0. So, the predicted taxon probabilities from ANNA scheme are higher than observed occurrence probabilities in many calibration and validation reference sits. The BROI and NROI schemes interestingly have less number of reference sites below the

residual = 0 than above residual =0 in calibration data, but they have higher number of reference sites above the residual = 0 than below residual = 0 in validation data. The predicted taxon probabilities from BROI and NROI schemes are lower than observed occurrence probabilities in many calibration reference sites, but higher in validation sites. The scatter plots in Figures 22, 23, 24, and 25 show these slightly imbalance around the line at residual = 0.

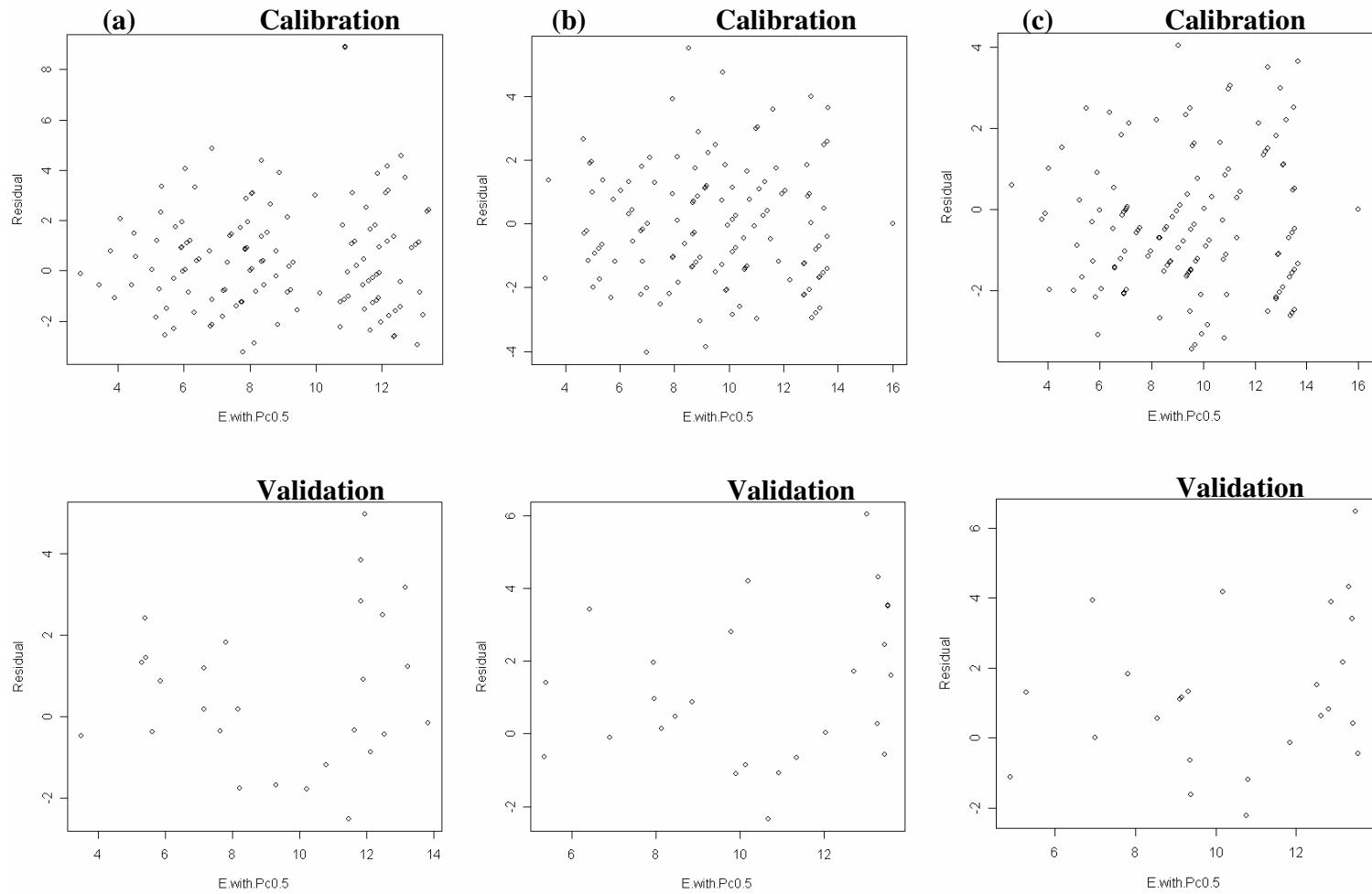


Figure 22. E vs. Residual at Pc=0.5 for (a) ANNA, (b) BROI, and (c) NROI schemes at the family level with only GISHydro2000 predictors. The calibration cases are on the top, and the validation cases are on the bottom.

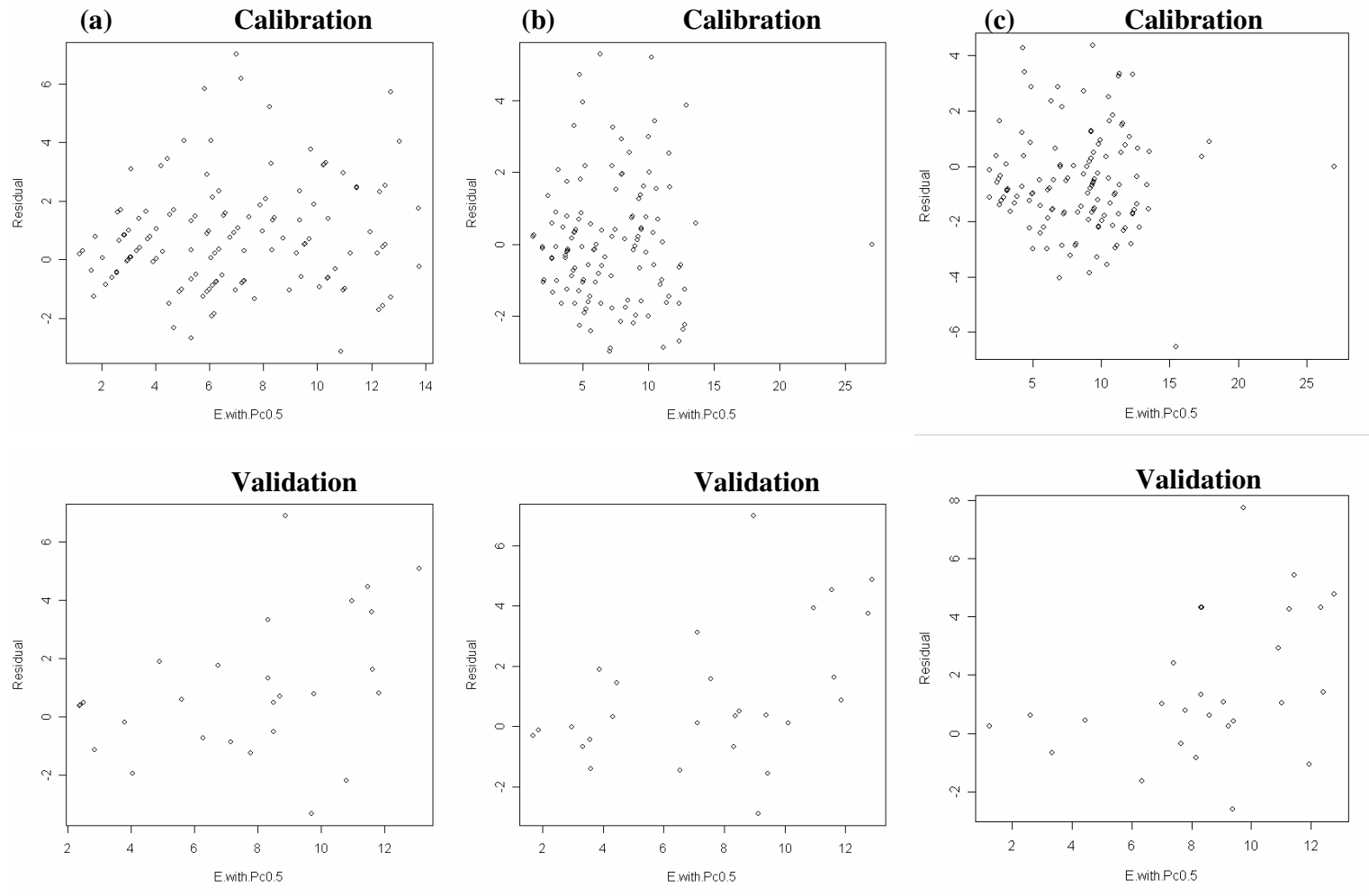


Figure 23. E vs. Residual at Pc=0.5 for (a) ANNA, (b) BROI, and (c) NROI schemes at the genus level with only GISHydro2000 predictors. The calibration cases are on the top, and the validation cases are on the bottom.

Table 26. Residual Analyses for Region-Style Models at the family level with only GISHydro2000 predictors

Region Selecting scheme	Best Selecting Parameter method	Weighting Function parameters		Residual Analysis 1		Residual Analysis 2				
				Number of Reference Sites		Number of Reference Sites		Residual Means		P-value
				Above	Below	Left	Right	Left	Right	
ANNA	Stepwise AIC	Neighbor no. = 21	Calibration Validation	71 15	58 12	71 13	58 14	0.42 0.36	0.59 0.86	0.651 0.48
BROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 30\%$, $TP = 50\%$, $NST = 8$, $n = 2.5$.	Calibration	59	70	64	65	0.02	-0.10	0.72
			Validation	17	10	12	15	0.21	1.63	0.08
NROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 40\%$, $NST = 5$, $n = 2.5$.	Calibration	46	83	67	72	-0.41	-0.31	0.71
			Validation	19	8	13	14	1.02	1.71	0.41

Table 27. Residual Analyses for Region-Style Models at the genus level with only GISHydro2000 predictors

Region Selecting scheme	Best Selecting Parameter method	Weighting Function parameters		Residual Analysis 1		Residual Analysis 2				
				Number of Reference Sites		Number of Reference Sites		Residual Means		P-value
				Above	Below	Left	Right	Left	Right	
ANNA	Exhaustive AIC	Neighbor no. = 16	Calibration Validation	85 18	44 9	70 11	59 16	0.53 0.07	1.08 1.62	0.10 0.06
BROI	Exhaustive AIC	$\theta_L = 5\%$, $\theta_U = 30\%$, $TP = 75\%$, $NST = 11$, $n = 2$.	Calibration	57	72	67	62	-0.07	0.09	0.60
			Validation	17	10	12	15	0.21	1.63	0.09
NROI	Exhaustive AIC	$\theta_L = 6\%$, $\theta_U = 40\%$, $NST = 4$, $n = 4$.	Calibration	44	85	58	71	-0.65	-0.51	0.66
			Validation	21	6	13	14	-3.15	0.62	0.18

Residual Analysis 2.

And then the residual analysis 2 examined the left and right balance in the scatter plots. Every scatter plot in Figure 22, 23, 24, and 25 was divided by a vertical line at the average of the reference site $\sum E_{at Pc=0.5}$ in the data set. The points on the left are residual values with $E_{at Pc=0.5} < avg(\sum E_{at Pc=0.5})$, and the points on the right are residual values with $E_{at Pc=0.5} \geq avg(\sum E_{at Pc=0.5})$. By separating the scatter points in to half, the number of reference sites at each side was counted and the Student t-test was conducted to examine the mean difference of left and right side. The hypothesis is defined as:

H₀: The true difference in residual means between left and right groups is equal to zero.

H₁: The true difference in residual means between left and right groups is not equal to zero.

This analysis results are summarized in the same Tables 26, 27, 28, and 29 under Residual Analysis 2. The number of reference sites from left and right sides are listed on the seventh and eighth columns. The residual means of left and right side groups are listed in ninth and tenth columns. The p-values from the Student t-test are listed in the 11th column.

The difference of number of reference sites between left and right side groups is small, especially not as large as that between above and below residual = 0. No particular pattern was observed from left or right side of residual points. The means from two sides are slightly different in most predictive scheme cases, but no p-values

of the Student t-test for the two sides shows the means from two sides are significantly different. Thus, no pattenen or difference was found between the two side groups by the Residual Analysis 2.

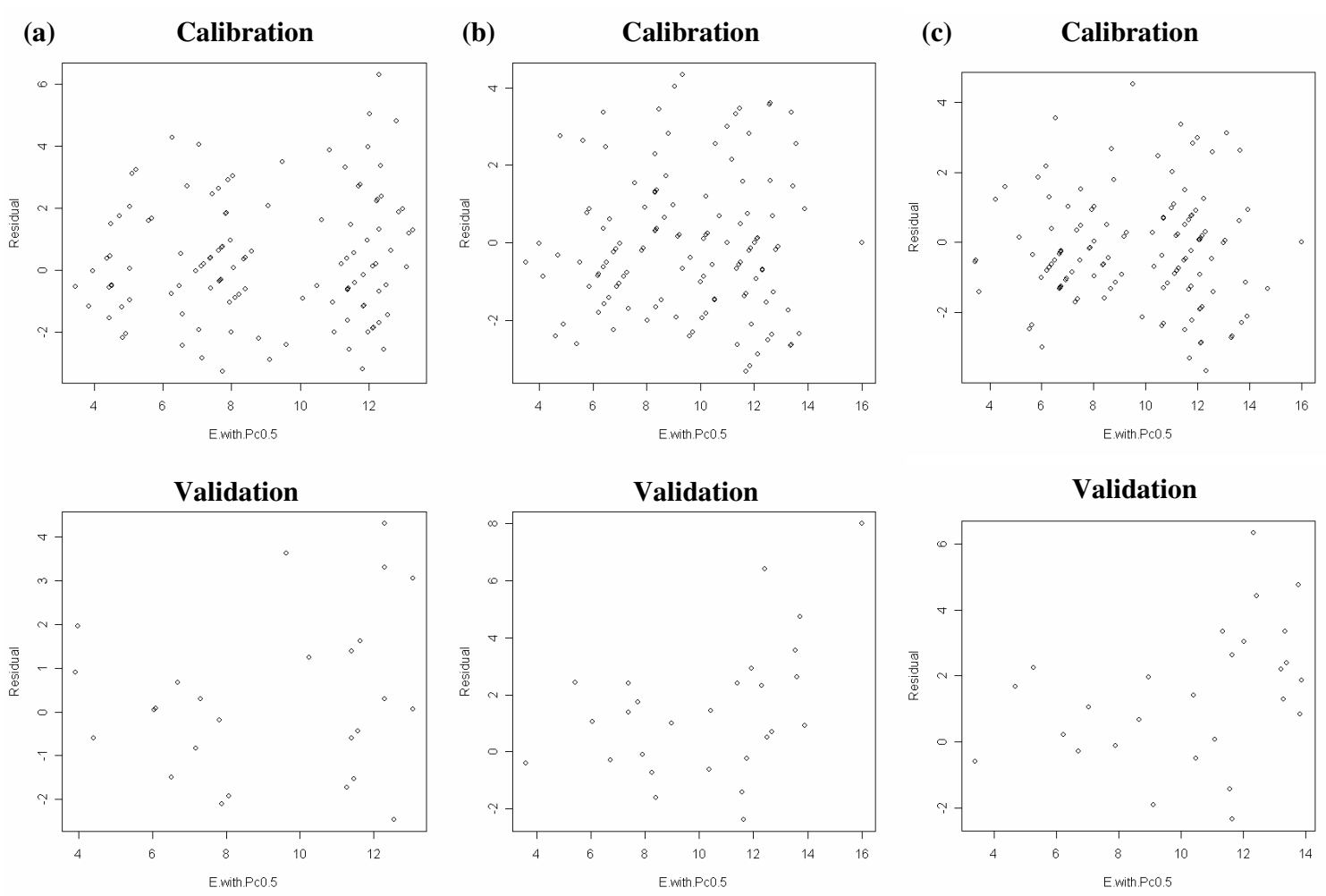


Figure 24. E vs. Residual at Pc=0.5 for (a) ANNA, (b) BROI, and (c) NROI schemes at the family level with all predictors. The calibration cases are on the top, and the validation cases are on the bottom.

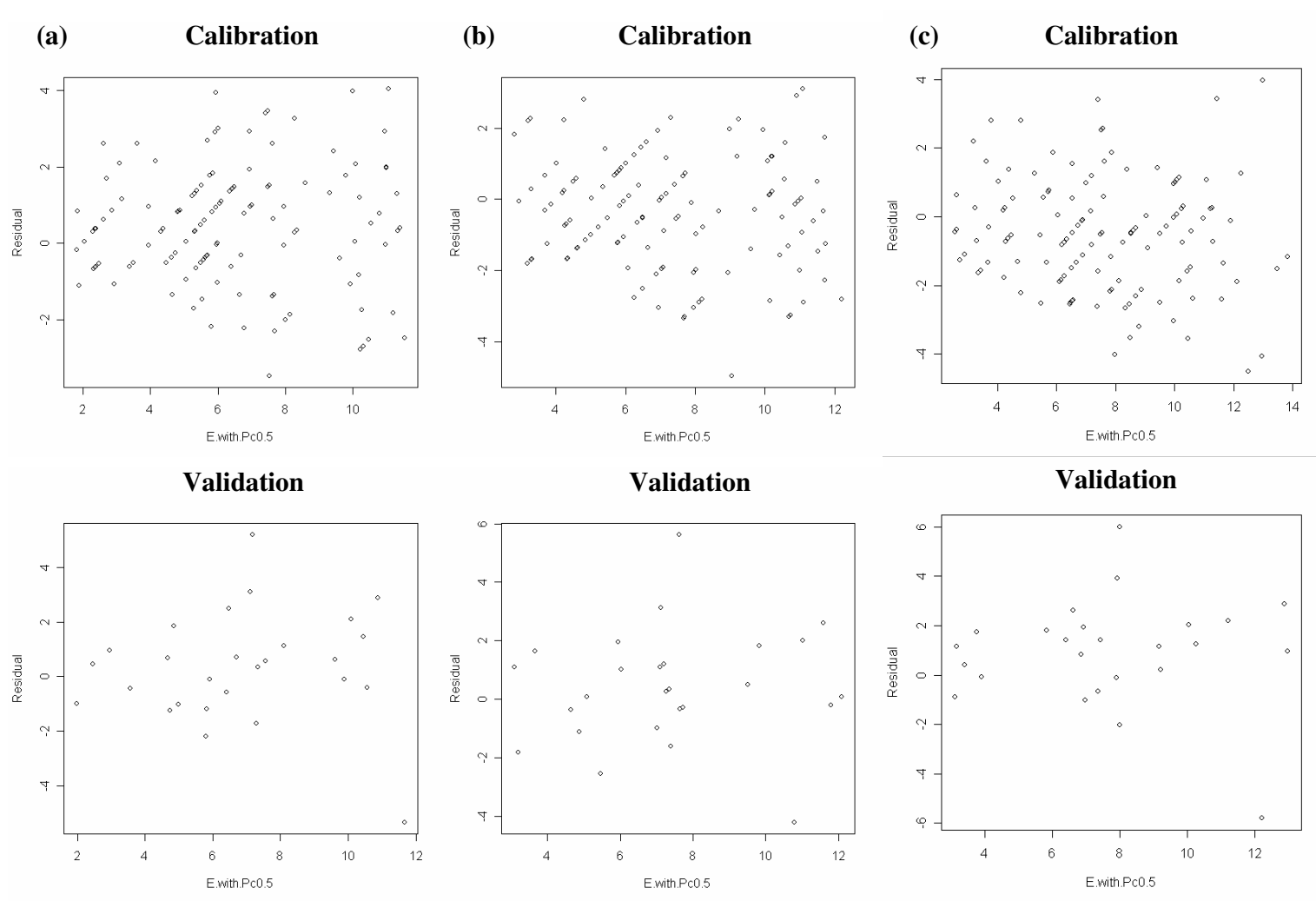


Figure 25. E vs. Residual at Pc=0.5 for (a) ANNA, (b) BROI, and (c) NROI schemes at the genus level with all predictors. The calibration cases are on the top, and the validation cases are on the bottom.

Table 28. Residual Analyses for Region-Style Models at the family level with all predictors

Region Selecting scheme	Best Selecting Parameter method	Weighting Function parameters		Residual Analysis 1		Residual Analysis 2				
				Number of Reference Sites		Number of Reference Sites		Residual Means		P-value
				Above	Below	Left	Right	Left	Right	
ANNA	Stepwise AIC	Neighbor no. = 34	Calibration	64	60	65	59	0.25	0.31	0.88
			Validation	16	11	12	15	-0.26	0.83	0.10
BROI	Stepwise AIC	$\theta_L = 5\%$, $\theta_U = 25\%$, $TP = 50\%$, $NST = 8$, $n = 3.5$.	Calibration	55	69	59	65	0.13	-0.18	0.32
			Validation	18	9	11	6	0.62	1.99	0.10
NROI	Stepwise AIC	$\theta_L = 6\%$, $\theta_U = 40\%$, $NST = 6$, $n = 3$.	Calibration	52	72	58	66	-0.16	-0.31	0.59
			Validation	20	7	10	17	0.49	1.98	0.04

Table 29. Residual Analyses for Region-Style Models at the genus level with only all predictors

Region Selecting scheme	Best Selecting Parameter method	Weighting Function parameters		Residual Analysis 1		Residual Analysis 2				
				Number of Reference Sites		Number of Reference Sites		Residual Means		P-value
				Above	Below	Left	Right	Left	Right	
ANNA	Stepwise AIC	Neighbor no. = 25	Calibration	75	49	70	54	0.48	0.39	0.76
			Validation	15	12	14	13	-2.45	0.86	0.33
BROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 25\%$, $TP = 50\%$, $NST = 13$, $n = 2.5$.	Calibration	51	73	69	55	-0.25	-6.0	0.24
			Validation	17	10	16	11	0.39	0.54	0.86
NROI	Stepwise AIC	$\theta_L = 10\%$, $\theta_U = 40\%$, $NST = 7$, $n = 4$.	Calibration	44	80	65	59	-0.31	-0.72	0.17
			Validation	19	8	13	14	0.82	0.97	0.86

6.2 COMPARING THE REGION-STYLES AND RIVPACS-STYLE MODEL

RESULTS

One of the goals of this study is to compare the model performance between Region-style models and RIVPACS-style model. Moglen and Paul (2005) built the RIVPACS-style for Maryland using only GISHydro2000 predictors. The model results using both GISHydro2000 and MBSS predictors are from Paul and his symposium presentation (Paul, unpublished 2006). The previous RIVPACS-style models applied to Maryland results have shown that predictions are better at the family level than at the genus level. This result is similar to that of Region-style models. The RIVPACS-style model using predictors from both GISHydro2000 and MBSS data makes better predictions than when it uses GISHydro2000 predictors alone. This is similar in Region-style models as well. A more detailed comparison of family and genus level models is described below.

6.2.1 Family Level

Moglen and Paul (2005) used the predictors from GISHydro2000 to build the RIVPACS-style model for Maryland. The statistics summary of the O/E scores from this model are shown in Table 30 and Figure 26 shows the O/E histogram and E vs. O scatter plot.

Table 30. RIVPACS-style family level result with GISHydro2000 predictors (adapted from Moglen and Paul, 2005)

Family		Mean	SD	Best	Null	%Explained
9 groups	Pc>0.5	1	0.25	0.18	0.32	50

Predictors: Area, Elevation, Hypsometry, Longitude, Percent soil type D

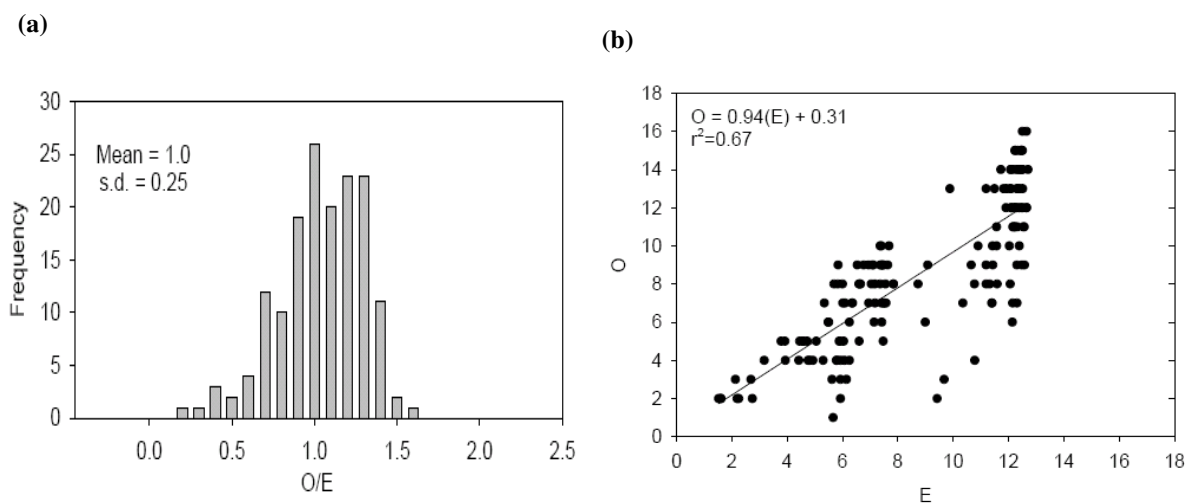


Figure 26. RIVPACS-style family level results using GISHydro2000 predictors, 9 groups, $P_c > 0.5$. (a) O/E scores frequency histogram, (b) E vs. O scatter plot. (Adapted from Moglen and Paul, 2005)

These are the O/E score results based on calibration data (the same calibration data set as Region-style models). The results from validation data are not available in the source literature. The mean of O/E score from RIVPACS-style model has a value as 1, and the standard deviation is 0.25. When these results are compared with the same case results shown in Table 16 in Chapter Five, Region-style models performed better than RIVPACS-style in the calibration phase. The mean value of ANNA's O/E score is less than one (0.944), while BROI and NROI values are slightly larger than one at 1.004 and 1.043, respectively. The standard deviations using Region-style modeling in Table 16 are either equally small (0.249) or smaller values (0.219, 0.189) than that using RIVPACS-style modeling. The number under the "Null" column in Table 30 is the standard deviation value for the predictive result if not using any multivariate analysis method or weighting in selecting reference sites to predict taxon appearance (Van Sickle, 2005). It can serve as the control case in comparing the

predictive models. The number under the “Best” column is the best standard deviation value that the RIVPACS-style model can achieve in Maryland. The NROI predictive scheme in Region-style models has a standard deviation of 0.189, which performs almost as well as the RIVPACS-style’s best case. The RIVPACS-style’s best case is the model with as many predictors as possible in its discriminant functions, which may imply that there is an overfitting effect in the model. When building Region-style models in our case, the predictors in multiple regression functions have been screened with the rational multiple regression checking and applied strict selection criteria. The overfitting problem in the models is thus alleviated. Also according to the RMSE definition (eq. 33), both BROI and NROI have smaller RMSE values than the RIVPACS-style does. Region-style models using only GISHydro2000 parameters perform better than RIVPACS-style model.

Compare the graphs in Figure 26 with the Region-style models’ graphs in Figures 10, 11, and 12 in Chapter Five. The graph sets in the upper right of Figures 10, 11, and 12 are the O/E frequency histograms and E vs. O scatter plots for $P_c > 0.5$ with calibration data of the Region-style models. All of the O/E frequency histograms concentrate around the mean value of one fairly well no matter which style model they come from. However, the E vs. O scatter plot from the RIVPACS-style model shows a more dispersed and separated pattern, while the plots from Region-style models aggregated along the regression line. The R-squares in BROI and NROI, 0.706 and 0.733, respectively, are higher than the 0.67, in the RIVPACS-style model. The BROI and NROI models using only GISHydro2000 predictors perform better than RIVPACS-style model.

Paul (unpublished, 2006) used the predictors from both GISHydro2000 and MBSS physical data to build the RIVPACS-style model for Maryland. These results are shown in Table 31 and as an O/E histogram and E vs. O scatter plot in Figure 27.

Table 31. RIVPACS-style family level results with All predictors (adapted from Paul, 2006 symposium presentation)

Family		Mean	SD	Best	Null	%Explained
9 groups	Pc>0.5	1	0.23	0.17	0.3	54

Predictors: Average Width, Cobble, Easting, Hypsometry, Northing, 2-year Precipitation, Percent soil type D.

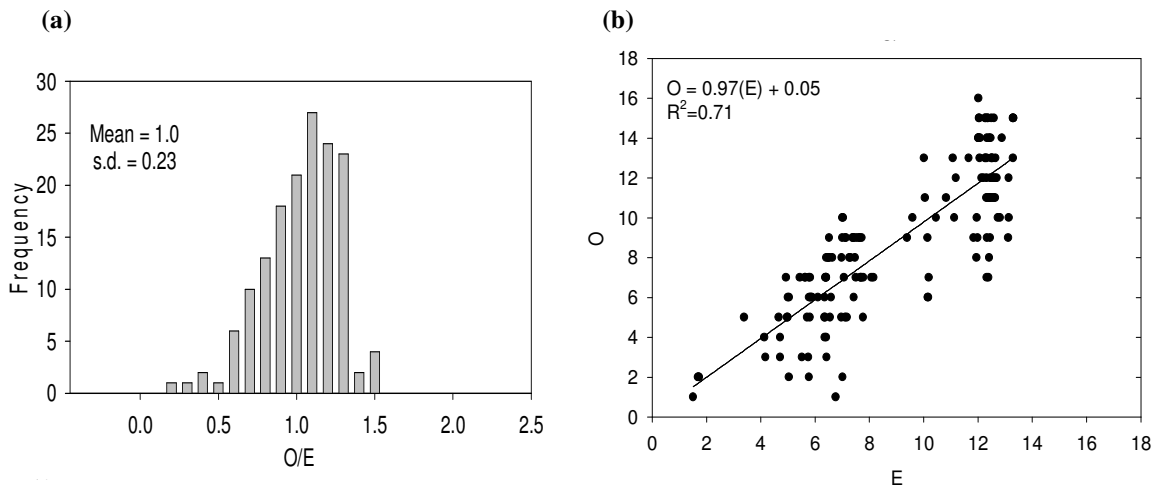


Figure 27. RIVPACS-style family level results using all predictors, 9 groups, Pc>0.5. (a) O/E scores frequency histogram, (b) E vs. O scatter plot. (Adapted from Paul, 2006 symposium presentation)

These are the O/E ratio results based on calibration data (the same calibration data set as Region-style models). The results from validation data are not available in the source literature. Comparing them with the calibration results in Table 18 in Chapter Five shows that Region-style models have better performance for calibration data. For the O/E mean, ANNA has the smaller value at 0.969, while BROI and

NROI have values a little larger than one at 1.007 and 1.028 respectively. The standard deviations in Region-style models have either equal or smaller values than the standard deviation in RIVPAC-style model in Table 31. NROI has a standard deviation of 0.177, which means NROI performs almost as well as the RIVPACS-style's best case. And both BROI and NROI have smaller RMSE values than the RIVPACS-style results. Region-style models using all parameters perform better than RIVPACS-style model.

Compare the graphs in Figure 27 with Region-style models' results in Figures 16, 17, and 18. The graph sets in the upper right in Figures 16, 17, and 18 are the O/E frequency histograms and E vs. O scatter plots for $P_c > 0.5$ with calibration data for Region-style models. The O/E frequency histograms of all the models are slightly skewed to the left. The E vs. O scatter plot from the RIVPACS-style model with all predictors still shows a more dispersed and separated pattern, while the plots from the Region-style models are more closely congregated. The R-square of BROI is as good as that of the RIVPACS-style model. NROI has a higher R-square value, 0.768, than the RIVPACS-style model. Region-style models have more continuous shape than RIVPACS-style model. Only NROI scheme has higher R-square than the RIVPACS-style model.

6.2.2 Genus Level

The genus level results using GISHydro2000 predictors are listed in Table 32 and Figure 32 has the O/E histogram and O vs. E scatter plot from Moglen and Paul (2005).

Table 32. RIVPACS-style genus level result with GISHydro2000 predictors (adapted from Moglen and Paul, 2005)

Genus		Mean	SD	Best	Null	%Explained
7 groups	Pc>0.5	1	0.31	0.27	0.46	79

Predictors: Area, Elevation, Hypsometry, Longitude, 2-year precipitation, and Relief.

These are the O/E score results based on calibration data. Comparing them with the results in Table 17 shows that Region-style models have better performance with calibration data. For the O/E mean, ANNA has the smallest value at 0.88, while BROI and NROI have values slightly larger than one at 1.001 and 1.081 respectively. The standard deviations of Region-style models, however, are either equal to or smaller value than that of the RIVPACS-style model. NROI has a standard deviation of 0.262, which indicates NROI performs even better than the RIVPACS-style's best case 0.27. And both BROI and NROI have smaller RMSE values than the RIVPACS-style does. Region-style model with NROI and BROI scheme perform better than RIVPACS-style model in genus level with GISHydro2000.

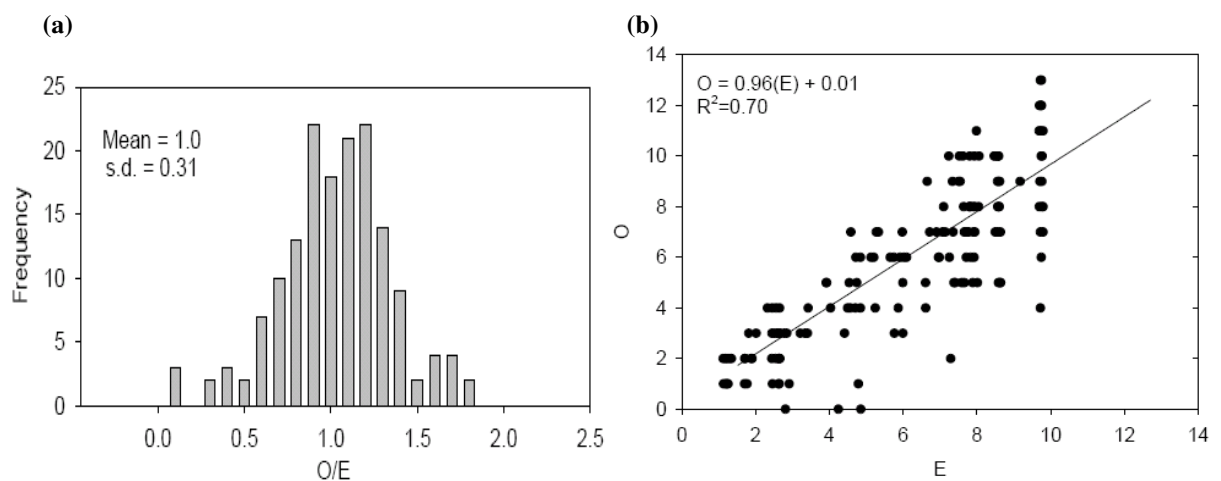


Figure 28. RIVPACS-style genus level results using GISHydro2000 predictors, 7 groups, Pc>0.5. (a) O/E scores frequency histogram, (b) E vs. O scatter plot. (Adapted from Moglen and Paul, 2005)

A comparison of the graphs in Figure 28 with the Region-style models' results in Figures 13, 14, and 15 in Chapter Five was done. The graph in the upper right in each of Figures 13, 14, and 15 are the O/E frequency histograms and E vs. O scatter plots for $P_c > 0.5$ and calibration data for the Region-style models. The O/E score histograms concentrate mostly around mean value one except for NROI. The O/E ratio histogram of NROI has a higher concentration value than mean value 1. The E vs. O scatter plot from the RIVPACS-style model has a more dispersed plot, while the plots from the Region-style models gather along the regression line. The R-squares of Region-style models, 0.745, 0.816, and 0.817, are all higher than the 0.70, of the RIVPACS-style model. Region style models perform better than RIVPACS-style model in Maryland with GISHydro2000 predictors.

The results with both GISHydro2000 and MBSS physical data in genus level are shown in Table 33. Figure 29 is the O/E histogram and O vs. E scatter plot from Paul (unpublished, 2006).

Table 33. RIVPACS-style genus level results with all predictors (adapted from Paul, 2006 symposium presentation)

Genus		Mean	SD	Best	Null	%Explained
7 group	$P_c > 0.5$	1	0.29	0.23	0.45	73

Predictors: Boulder, Cobble, Catchment Slope, Easting, Elevation, 2-year Precipitation.

These are the O/E score results based on calibration data. Comparing them with the results in Table 19 shows that Region-style models perform better with calibration data. For the O/E mean, ANNA has the smallest value at 0.923, while BROI and NROI have values a little larger than one at 1.05 and 1.065 respectively.

The standard deviations of Region-style models, however, all have smaller values than the RIVPACS-style model. NROI has a standard deviation of 0.245, which is nearly as good as the RIVPACS-style's best case. And all the Region-style models built with all predictors have smaller RMSE values than the RIVPACS-style. The Region-style models with all predictors performed better than RIVPACS-style model in Maryland.

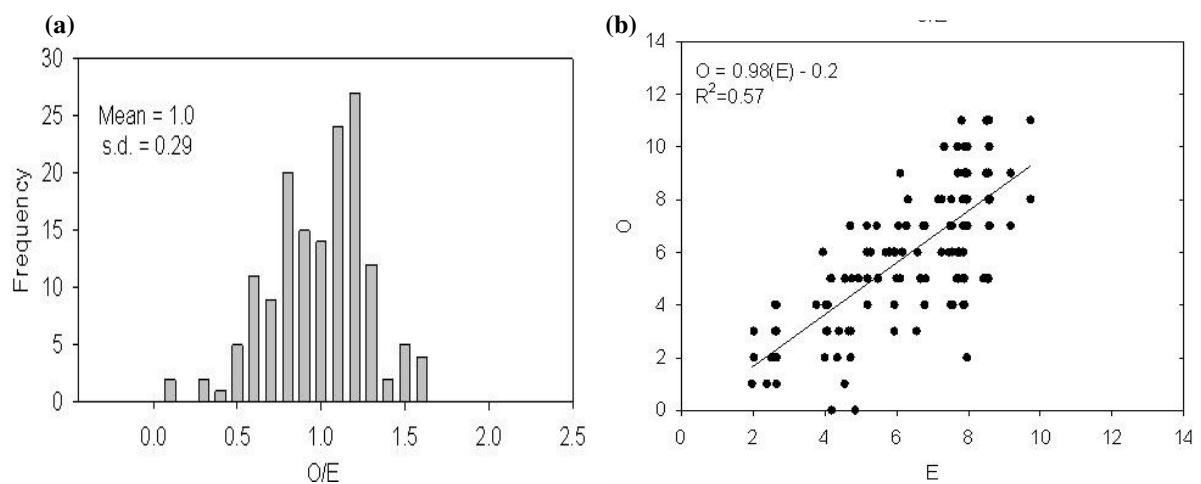


Figure 29. RIVPACS-style genus level results using all predictors, 7 groups, $P_c > 0.5$. (a) O/E scores frequency histogram, (b) E vs. O scatter plot (Adapted from Paul, 2006 symposium presentation)

A comparison of the graphs in Figure 29 with Region-style models' results in Figures 19, 20, and 21 was done. The graph in the upper right in each of Figures 19, 20, and 21 are the O/E frequency histograms and E vs. O scatter plots for $P_c > 0.5$ with calibration data for the Region-style models. The O/E frequency histograms concentrate mostly around mean value one except for ANNA. The ANNA O/E histogram concentrates on the lower side of mean value 1. The E vs. O scatter plot from the RIVPACS-style model with all predictors still shows a more dispersed plot,

while the plots from the Region-style models all show more aggregation along the regression line. The R-square values of the Region-style models, 0.745, 0.754, and 0.765, are all much better than the 0.57, of the RIVPACS-style model. Region style models with all predictors perform better than RIVPACS-style model in Maryland.

6.3 CONCLUSIONS

The Region-style models use similar reference sites to predict the occurrence of macroinvertebrate taxa. Different region selection schemes utilize different numbers of reference sites. The prediction results, however, show no preference to models with more reference sites or models with fewer reference sites. Only in the Genus level with all predictors does NROI make the best predictions with fewer reference sites. This suggests that a good region selection scheme can provide good predictions regardless of the number of reference sites. In order to have precise prediction, different predictive styles were tried to find sufficient reference sites for improving the prediction. The sufficient reference sites however do not rely on the number of reference sites but rather rely on the quality of reference sites. Due to the variability of reference sites within the state, fixed number of reference sites or adding more reference sites does not necessarily increase the quality of reference sites, thus additional reference sites do not necessarily achieve better prediction.

Region-style predictive schemes were trying to add the flexibility in choosing reference data. Generally Region-style models perform better than the earlier RIVPACS-style model at both family and genus taxon levels for Maryland. When using GISHydro2000 predictors, the performances of Region-style models are much

better than the RIVPACS-style model. This is very encouraging for later GISHydro2000 assessment applications in stream restoration sites. The more precise prediction models can thus be used for the restoration effort assessment.

CHAPTER 7: APPLYING THE PREDICTIVE MODEL TO STREAM RESTORATION DATA WITH GIS TOOL

One of the objectives of this research is to use the macroinvertebrate predictive model as a systematic standard measurement for assessing stream restoration efforts. The macroinvertebrate predictive models have been built. These models can now be used to evaluate the stream restoration projects in Maryland.

The Montgomery County Department of Environmental Protection is dedicated to protecting and restoring watersheds within Montgomery County, Maryland. It continues tracking restoration projects with pre- and post- restoration monitoring, which is very rare, and provides the monitoring data for examining the stream restoration effort.

7.1 MONTGOMERY COUNTY AND STREAM RESTORATION PROJECTS

BACKGROUND

7.1.1 Montgomery County

Montgomery County, is located in Central Maryland and is adjacent to Virginia and the nation's capital, Washington DC (Figure 30). It includes 497 square miles of land area. The topography is rolling with small hills. Elevations range from 52 feet above sea level near the District Line to 850 feet in the northern portion of the County near Damascus. One-third of its land is still preserved for agricultural use. Montgomery County is located at the epicenter of the Mid-Atlantic's thriving federal and advanced technology marketplace.

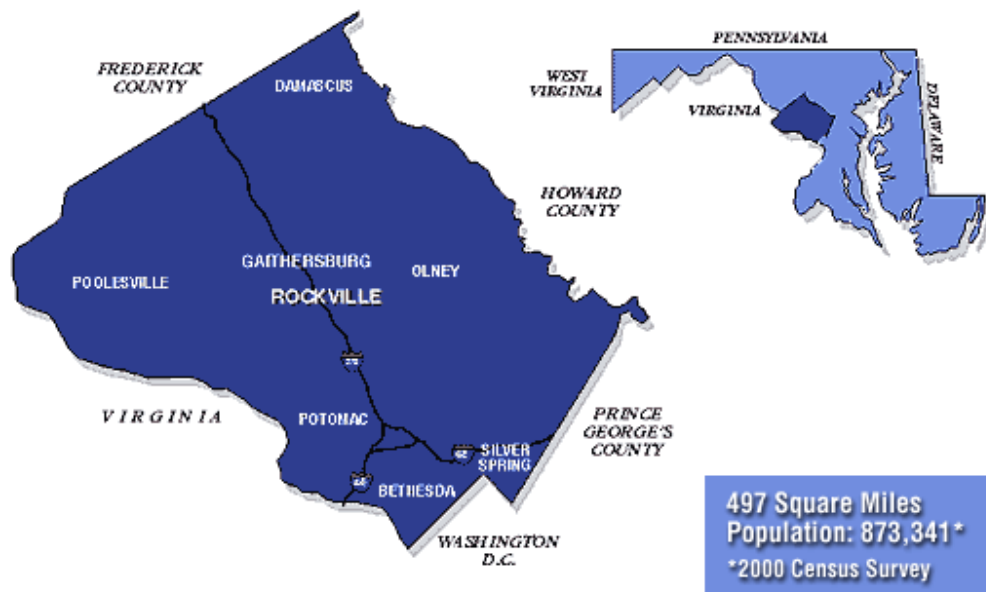


Figure 30. Location of Montgomery County in Maryland. (From Montgomery County Government website)

The programs that are run by the Montgomery County Department of Environmental Protection (DEP) involve air quality, energy, noise control, grass, forest and trees, groundwater, stormwater control, streams and watersheds, and more. The DEP has a Watershed Management Division especially for watershed conservation in the county. The stream restoration monitoring data used in this study were obtained from the Watershed Management Division.

7.1.2 Stream Restoration Projects by the Montgomery County DEP

The Montgomery County DEP has four programs that are dedicated to protecting and restoring the streams in Montgomery County: the Countywide Stream Protection Strategy, Biological Monitoring, Streams and Watersheds, and Watershed Restoration. After years of study and practice, Best Management Practices (BMP) within entire watersheds are recognized as more important than specific stream site restorations. Montgomery County DEP does put more emphasis on entire watershed

management. The watersheds within Montgomery County are illustrated in Figure 31. Streams in Montgomery County suffer from being located in a highly urbanized area which experienced rapid development over the last 60 years. Some subwatershed imperviousness is as high as 32 %. As a result, some of the most common problems are:

- Excessive sediment in streams
- Unstable stream banks
- Spills or illegal discharges of pollutants
- Severely eroding stream channels
- Lack of forested stream buffer areas
- Uncontrolled runoff from high stormwater flows

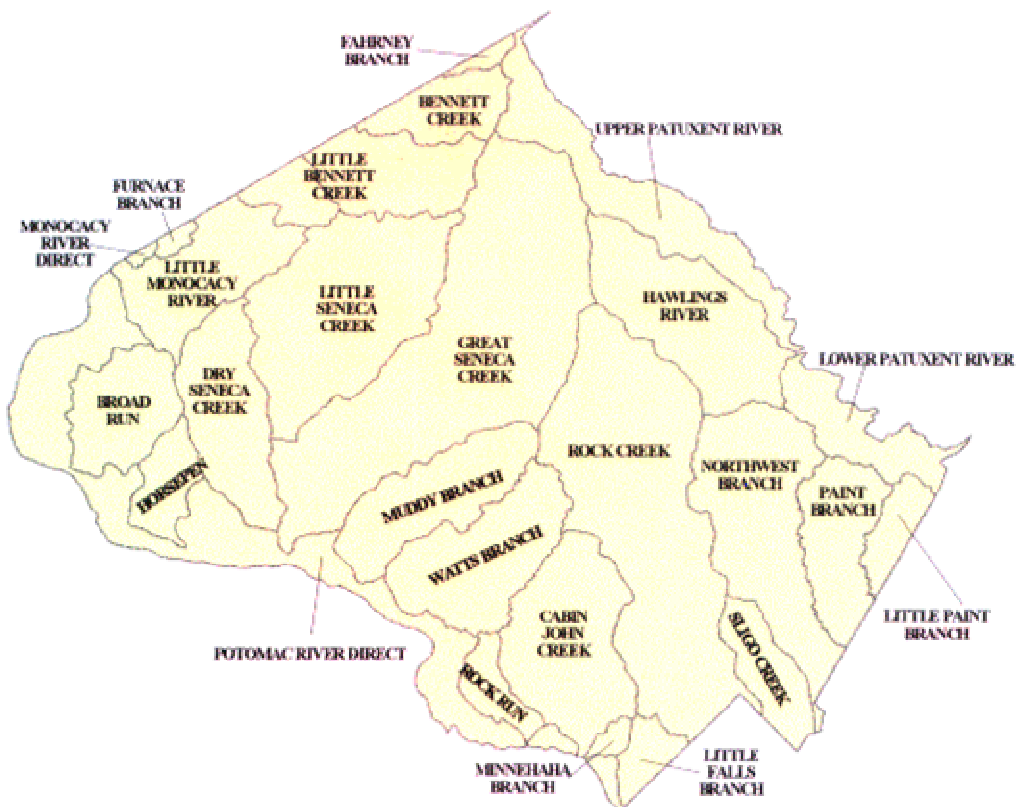


Figure 31. The Watersheds within Montgomery County. (From Montgomery County Government website)

The Montgomery County DEP has conducted an assessment of conditions in its 1,500 miles of streams through an intensive effort which evaluated data from over 200 monitoring stations. In their watersheds, Montgomery County DEP has taken various stream restoration actions to alleviate problems. Stream restoration actions were applied according to different stream restoration project situations. The various actions are summarized below.

Three types of stormwater management were used in the designs:

- Stormwater Ponds – Stormwater ponds provide water quantity and quality control by storing stormwater. They are designed to reduce downstream flooding and erosion and allow sediments to settle.
- Stormwater Wetlands – Stormwater wetlands are designed to use the water quality improvement functions similar to natural wetlands to treat and contain runoff.
- Sand Filters – A sand filter is used to strain particulate matter from stormwater runoff through a horizontal bed of sand. It is designed mainly to provide water quality control.

Instream stream restoration structures include:

- Rock and log vanes – They direct water away from unstable stream banks, and form down stream scour pools.
- Rock cross vanes – They function as grade control and slow the erosive process of stream down-cutting.
- Rootwad revetments – They are instream structures that use lowermost portion of a tree trunk to help stabilize stream banks and create scour holes and overhead cover for fish.
- Boulder rock – It is installed at the toe of steam bank slopes to stabilize the area of stream channel that is subject to great erosive or shear stress.
- Rock packing – a technique used to stabilize trees along stream banks that involves packing rocks underneath and around the tree roots.

They also enhance the riparian habitats and stream buffers:

- Create shallow vernal pools or wetlands – to treat water quality of

stormwater discharges.

- Rehabilitate the stream buffers – Planting native trees and shrubs along the streams

The valuable thing in these restoration projects is that Montgomery County DEP performed biomonitoring before and after stream restoration construction. These data were obtained from the Watershed Management Division in Montgomery County DEP for this research. The biomonitoring data were obtained from Keith Van Ness and the project completion dates were verified with Daniel Harper. Table 30 shows the different projects in each watershed along with the Station ID and stream monitoring site locations in latitude and longitude. The following goals for the projects are also included in Table 34.

- Improve benthic population or density
- Improve fish population or density
- Restore habitat
- Reduce erosion
- Reduce thermal impacts
- Create or improve wetlands
- Create amphibian habitat
- Remove fish blockage
- Riparian reforestation
- Improve storm water management quantity control
- Improve water quality

As shown in Table 34, all of the projects have the goal to improve benthic population or density along with their other goals in water quantity and quality controls.

Table 34. The summary table of stream restoration projects by the Montgomery County DEP. (Adapted from Montgomery County DEP data sheet) (! : First post-monitoring done in the same year as construction)

Restoration Project	Station ID	LAT_N	LONG_W	Restoration Completed (yr)	Pre-construction monitoring	Post-Biomonitoring Time						Restoration Goals												
						1st year post-monitoring	2nd year post-monitoring	3rd year post-monitoring	4th year post-monitoring	5th year post-monitoring	6th year post-monitoring	Improve Benthic pop/density	Improve Fish pop/density	Restore Habitat	Reduce Erosion	Reduce Thermal Impacts	Create or Improve Wetlands	Create Amphibian Habitat	Remove Fish Blockage	Riparian Reforestation	Improve SWM Quantity Control	Improve Water Quality		
ROCK CREEK																								
Sycamore Creek																								
Upper	LRSB101A	39.09129715	-77.09407043	2004	3/25/2003	4/25/2005		4/5/2007					X	X	X	X		X	X		X			
Middle	LRSB101B	(no lat/long)		2004	4/4/2002	4/25/2005		4/5/2007					X	X	X	X		X	X		X			
Lower	LRSB101C	39.08611298	-77.10669708	2004	4/4/2002	4/25/2005		4/5/2007					X	X	X	X		X	X		X			
Josephs Branch Mainstem																								
Above Beach Drive	LRJB204	39.04035187	-77.08709717	Jan-05	3/24/2003	4/26/2005		4/18/2007					X	X	X	X		X	X		X			
Between Beach and Connecticut	LRJB203B	39.04614258	-77.0814743	Jan-05	3/28/2002	4/26/2005		4/5/2007					X	X	X	X		X	X		X			
Below Connecticut	LRJB203A	39.04839325	-77.07665253	Jan-05	3/28/2002	4/26/2005		4/5/2007					X	X	X	X		X	X		X			
Turkey Branch																								
Lower Turkey Branch																								
Just Above Veirs	LRTB203B	39.06033707	-77.08860016	2008	4/19/2002, 3/28/2006								X	X	X	X					X	X		
Between Veirs and Connecticut	LRTB203A	39.06846619	-77.08174896	2008	4/17/2002, 3/28/2006								X	X	X	X						X		
Above Elizabeth Street	LRTB101	39.07221603	-77.08058929	2008	4/17/2002, 4/21/2006								X	X	X	X						X		
Middle Turkey Branch (Between Connec. and Ga.)	LRTB202	39.07252121	-77.07501221	2008	4/14/2002, 4/26/2006								X	X	X	X						X		
Coquelin Run																								
Between Jones Mill Rd and Falls	LRCR101B	38.99544144	-77.06613159		3/22/2002								X	X	X	X						X		
Between Falls and Dunlop Trib	LRCR101A	38.99420166	-77.07450104		2/21/2002								X	X	X	X						X		
Stream Valley Drive																								
	URLF201A	(no lat/long)		2004	4/13/2004								X	X	X	X		X	X		X			
Little Falls																								
Mainstem below Mass. Ave (LFI)	LFLF301B	38.95342636	-77.11096954	2001		4/26/2002		4/12/2005					X	X	X	X						X		
Mainstem above Dalecarlia (LFII)	LFLF301C	38.94844055	-77.11264801	2001	4/26/2002																			

Table 34-continued. The summary table of stream restoration project by Montgomery County DEP. (Adapted from Montgomery County DEP data sheet) (! : First post-monitoring done in the same year as construction)

Restoration Project	Station ID	LAT_N	LONG_W	Restoration Completed (yr)	Pre-construction monitoring	Post-Biomonitoring Time						Restoration Goals														
						1st year post-monitoring	2nd year post-monitoring	3rd year post-monitoring	4th year post-monitoring	5th year post-monitoring	6th year post-monitoring	Improve Benthic pop/density	Improve Fish pop/density	Restore Habitat	Reduce Erosion	Reduce Thermal Impacts	Create or Improve Wetlands	Create Amphibian Habitat	Remove Fish Blockage	Riparian Reforestation	Improve SWM Quantity Control	Improve Water Quality				
Sligo Creek																										
SLIGO II																										
between Wheaton Br and Flora Lane	SCFT101	39.01294327	-77.03125763	1994		4/13/2004																				
Northwest Branch																										
<i>Northwest branch South of Randolph Rd.</i>																										
within restoration	NWNW408A	39.06155396	-77.02495303	2002	4/8/2002	4/29/2003		4/6/2005		4/11/2007		X	X	X	X									X		
within restoration	NWNW408B	39.06049728	-77.0194931	2002	4/8/2002	4/29/2003		4/6/2005		4/11/2007		X	X	X	X									X		
within restoration	NWNW407A	39.06773376	-77.02894592	2002	4/8/2002	4/29/2003		4/6/2005		4/18/2007		X	X	X	X									X		
above restoration	NWNW407B	39.06748199	-77.02893045	2002	4/8/2002	4/29/2003		4/6/2005		4/18/2007		X	X	X	X									X		
above restoration	NWNW407C	39.06552124	-77.02893045	2002	4/8/2002	4/29/2003		4/6/2005		4/18/2007		X	X	X	X									X		
above restoration	NWNW407D	39.06330872	-77.02638245	2002	4/8/2002	4/29/2003		4/6/2005		4/11/2007		X	X	X	X									X		
Anacostia II																										
Sherwood Forest	NWSF201	(no lat/long)		2005	4/13/2004	3/28/2006						X	X	X	X											
Bachelors Run East	NWBB104	(no lat/long)		2005	4/16/2004							X	X	X	X									X		
Bryants Nursary Run I,II	NWNW205	(no lat/long)			4/14/2004							X	X	X	X											
Upper Northwest Branch	NWNW304	39.11750031	-77.02236176		5/7/1997, 4/5/2001, 4/14/2004							X	X	X	X		X	X								
Sandy Spring I,II,III																	X						X			
Woodlawn, Sandyspring West	NWNW301	39.12732315	-77.02091217		4/14/2001, 3/29/2002, 4/14/2004							X	X	X	X											
Bachelors Run III													X	X	X								X			
Bachelors Run II	NWBF302A	(no lat/long)			4/13/2004							X	X	X	X											
Other Northwest Branch																										
Lockridge (ANACOSTIA I)	NWLR102	39.03256989	-77.00975037	2002	3/28/2001	! 4/19/2002		3/30/2004		4/8/2007		X			X		X	X								
Lamberton	NWLT101	39.04808044	-77.01570892	2002	5/3/2001	! 4/19/2002		3/30/2004		4/18/2007		X	X	X	X									X		
Northwood	NWND201	(no lat/long)		2004						4/18/2007		X	X	X	X									X		
Dumont Oaks	NWDO102	(no lat/long)		2005	3/31/2005		4/11/2007					X			X	X	X	X					X			

Table 34-continued. The summary table of stream restoration project by Montgomery County DEP. (Adapt from Montgomery County DEP data sheet) (! : First post-monitoring done in the same year as construction)

Restoration Project	Station ID	LAT_N	LONG_W	Restoration Completed (yr)	Pre-construction monitoring	Post-Biomonitoring Time						Restoration Goals												
						1st year post-monitoring	2nd year post-monitoring	3rd year post-monitoring	4th year post-monitoring	5th year post-monitoring	6th year post-monitoring	Improve Benthic pop/density	Improve Fish pop/density	Restore Habitat	Reduce Erosion	Reduce Thermal Impacts	Create or Improve Wetlands	Create Amphibian Habitat	Remove Fish Blockage	Riparian Reforestation	Improve SWM Quantity Control	Improve Water Quality		
Glen Allen Trib																								
Glen Allen Trib	NWGT101	39.058967	-77.041992		4/12/2005								x		x								x	
Pine Lake Trib	NWPT101	39.057639	77.035361		4/12/2005								x		x								x	
Paint Branch																								
Peachwood Park																								
Piping Rock Drive	PBGH108	39.09552765	-76.99691772	2001	3/24/2000	4/22/2002			4/20/2004	4/29/2005														
Fairland to 29				2001									x	x	x	x							x	
within restoration	PBPP305D	39.07136536	-76.97570801		4/26/2001	4/23/2002				4/6/2005														
within restoration	PBPP305E	39.07090378	-76.9755249		4/26/2001	4/23/2002	4/30/2003			4/6/2005														
within restoration	PBPP305F	39.06991959	-76.97512054		4/26/2001	4/23/2002	4/30/2003			4/6/2005														
above restoration	PBPP305A	39.08232498	-76.97692108		4/26/2001	4/23/2002	4/30/2003			4/6/2005														
above restoration	PBPP305B	39.07906342	-76.97797394		4/26/2001	4/23/2002	4/30/2003			4/6/2005														
above restoration	PBPP305C	39.07853317	-76.97728729		4/3/2001	4/29/2002	4/14/2003		4/16/2004	4/6/2005														
Gum Springs Farm SWM				2006																			x	x
Middle Gum Springs 700ft DS of Fireside Dr	PBGS102B	39.09543228	-76.97930908		4/5/1999, 4/13/2004								x	x	x	x								
Middle Gum Springs 452ft S of Bart Dr, W of Ansted Rd	PBGS111	39.08925629	-76.97554779		4/5/1999, 3/27/2000, 4/2/2001, 4/22/2002, 3/26/2003, 4/13/2004, 4/11/2005								x	x	x	x							x	
Hawlings River																								
7A Brighton Dam Road	HWHW314D	39.20306	-77.04583	2005	5/13/2003	4/16/2004							x	x	x	x						x	x	
Cabin John																								
Booze Creek	CJBC203	38.985356	-76.848033		4/12/2005								x	x	x	x							x	

Most of the stream monitoring sites have pre-construction biomonitoring data, and most of them have at least one post-restoration monitoring data set. Some of the projects have monitoring data up to 6 years after restoration construction. By its very nature, stream restoration takes time. The monitoring data can thus be used to examine if the stream ecosystem has recovered after restoration actions and if the restoration practices have achieved the goal of improving benthic populations. Project completion dates and dates of pre- and post-restoration biomonitoring are also shown in Table 34. The post-monitoring time is listed in order roughly by the year after restoration. Only a few projects were monitored annually after construction. Some of the projects were monitored every other year. Some of the first post-monitoring data were done in the same year as construction, which are noted by an exclamation point. These data may have been collected too soon to show any ecosystem recovery.

Many of the stream monitoring sites have latitude and longitude coordinates. These stream sites can be located on the map with the help of a Geographic Information System (GIS) tool. Figure 32 shows maps from ArcView with the scope including all of Montgomery County and is enlarged to the area with most of the stream restoration monitoring sites. The blue points are the stream monitoring locations. The geographic coordinates in decimal degrees were converted to meters to match the Maryland State Plane, NAD, 1983.

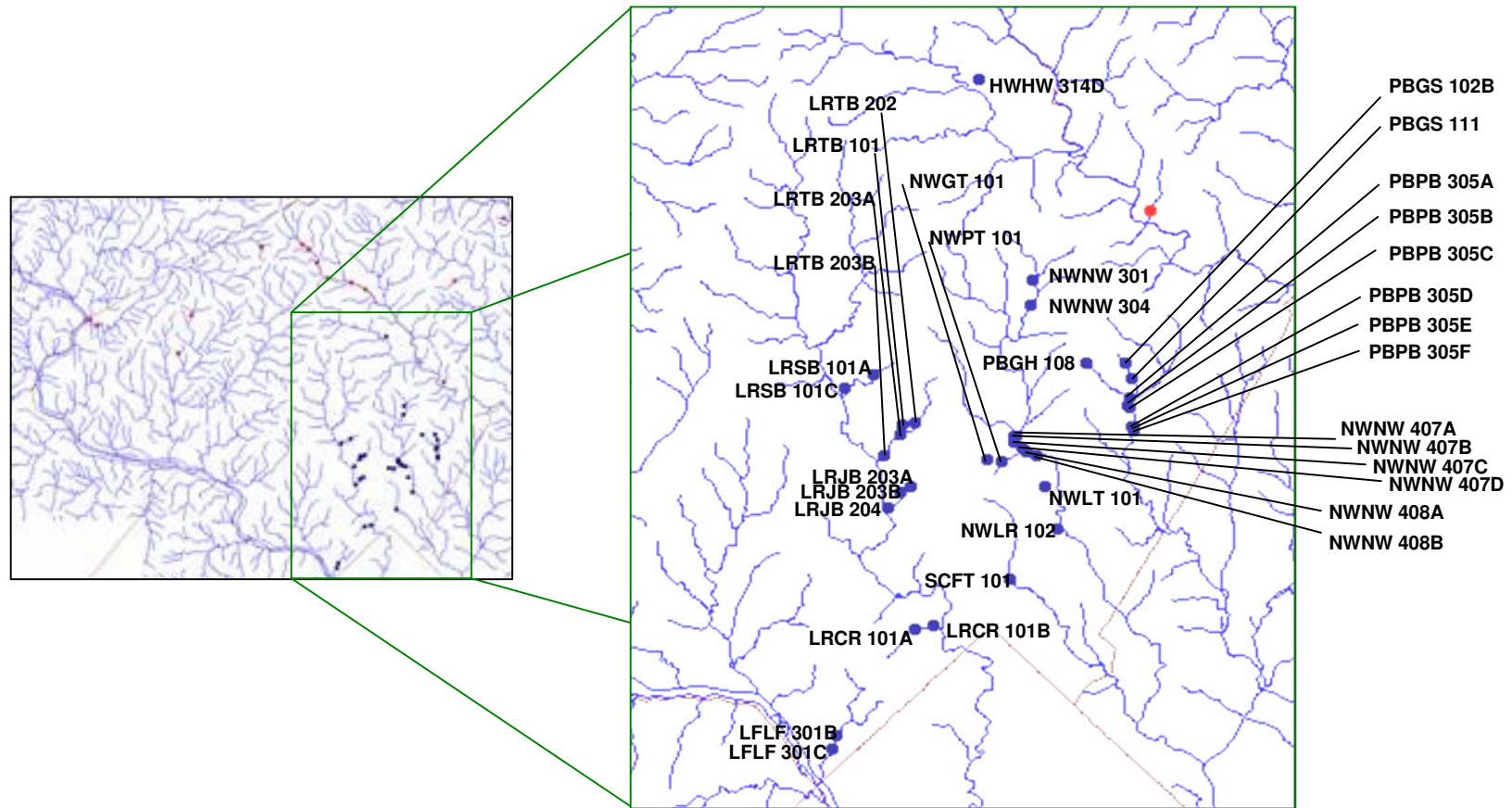


Figure 32. The stream restoration sites in Montgomery County

7.2 APPLICATION METHODS

The stream restoration biomonitoring data were obtained with their locations on the map in GIS. As mentioned in Chapter Six, the macroinvertebrate predictive models with only map-based predictors were built in GISHydro2000. The GIS can provide the physical predictors at any stream site without the need to measure on-site, thus checking and comparing the observed macroinvertebrate with the expected taxa from the model is possible.

7.2.1 GISHydro2000 Applications

GISHydro2000 was an ArcView GIS-based application software for conducting hydrologic analyses in the state of Maryland. It is sponsored by the Maryland State Highway Administration, developed and maintained by Dr. Glenn E. Moglen, an Associate Professor in the Department of Civil and Environmental Engineering at the University of Maryland, College Park. GISHydro2000 integrates an extensive database of terrain, land use, and soil data with tools for assembling and evaluating hydrologic models in Maryland. It is a valuable data source because watershed features can be generated from the GIS data layers and used as the predictors for macroinvertebrate predictive models.

Moglen and Paul (2005) built the RIVPACS-style model of Maryland into GISHydro2000. From the discussion in Chapter Six, Region-style models of Maryland with GISHydro2000 predictors have better prediction results than the RIVPACS-style model does. The Region-style models in Maryland were built into GISHydro2000 to assess Montgomery County stream restoration projects.

Building Region-Style Models in GISHydro2000

The MBSS reference sites biomonitoring data were represented as present/absent matrix at both family and genus taxon levels as described in Section 2.1. Step1. These two matrices (family and genus level) served as the reference for predicting the taxon appearance at any desired stream site. Table 2 in Chapter Three lists the potential predictors that can be generated from GISHydro2000. In order to apply the predictive models, our goal at the beginning was to obtain these potential predictors with a mouse click on any watershed outlet. These potential predictors were then substituted into the multiple regression functions that were derived in Section 4.1.3, and the physical feature distances between desired stream sites and reference stream sites was calculated and applied to the macroinvertebrate predictive models.

Because the model is targeted to be build in GISHydro2000, the map-based candidate predictors need to be generated from GISHydro2000 for the models. Some potential parameters were provided by built-in GISHydro2000 parameters. GISHydro2000 provides AREA, WSLOPE, CSLOPE, RELIEF, LIME, P2, EASTING, NORTHING, and soil type percentages values at any point. For the rest of the potential predictors, HUC 14 and HUC 8 were looked up from a 14-digit hydrologic unit code; PERIM and LENGTH were calculated from the selected watershed; SIN was calculated along the stream channel; P100 was obtained from a 100-Year, 24 hour storm precipitation map; HIGHELEV and HYPISO were calculated from a selected watershed area and its elevation. As for LAT and LONG, since the GISHydro2000 is set in the Maryland State Plane, NAD 1983, the selected locations

were projected back to geographic coordinates in decimal degrees. All of these calculation and look-up processes were programmed in Avenue, which allows the values to be immediately available at any selected point. Much of this programming was done earlier by Dr. Moglen for the RIVPACS-style model of Maryland. These Avenue codes were adapted for the application in Region-style models. This step is part of the new extension to GISHydro2000. Note that although the GIS data are stored in metric units, the calculated statistics within GISHydro2000 are stored in English units and the predictors for the Region-style models are also in English units as noted in Table 2. The parameters generated from GIS data need to agree with the predictors in the models. Once all the potential parameters were collected from the GIS data layers at a selected stream site, this predictor list was exported as a text file.

In Table 13 of Chapter Four, the three axis predictor lists from Stepwise AIC, Exhaustive AIC, and Exhaustive BIC parameter selection methods are summarized. Tables 9 and 11 in Chapter Four provide the multiple regression functions with only GISHydro2000 predictors from different parameter selection methods. Combined with different region selection schemes, ANNA, BROI, and NROI, a total of nine cases (3 region selection schemes \times 3 parameter selection methods) were programmed in FORTRAN 90 and the executable files (*.exe) were generated. These files use the predictor list from GISHydro2000 data as input and create text file output for the nine cases. The FORTRAN scripts take the parameters from an input file and insert them into multiple regression functions. The two present/absent matrices serve as references and the different region selection schemes were used to determine the weighting for the taxon prediction (in either family or genus level). The outputs from

these files are a list of expected taxon present probabilities and a list of MBSS reference sites that were included in the region of influence for the desired stream site.

A master script was programmed in Avenue within ArcView to execute all the steps. After selecting the desired stream site, the built-in basin statistics from GISHydro2000 as predictors was obtained (e.g., AREA, WSLOPE, CSLOPE, etc.) and those that are not built-in parameters were calculated (e.g., HUC8, HUC14, P100, etc.). Based on user inputs, taxon level (Family or Genus), the region selection scheme (ANNA, BROI, or NROI), and parameter selection scheme (Stepwise AIC, Exhaustive AIC, or Exhaustive BIC) were applied and the corresponding executable file was utilized to complete the Region-style macroinvertebrate prediction. The output from the models were then displayed on the ArcView screen.

As mentioned earlier, STATSGO soil data cover the entire region of GISHydro2000, while Ragan soil data cover only the state of Maryland. The earlier RIVPACS-style model of Maryland used soil type percentage data from Ragan soil cover. So, the results of this research used Ragan soil data as well in order to compare with the earlier model results in the same data base. To use Ragan soil data as predictors, this soil data source from “Ragan” was selected at the primary hydrologic analysis. On the other hand, no matter which soil data source is selected, GISHydro2000 automatically calculates the soil type percentage in STATSGO soil data. So, it is easier for the GISHydro2000 user if the model is built using STATSGO soil data. Region-style models using STATSGO soil data was also developed for user convenience. The Avenue scripts for the Region-style models for GISHydro2000 are

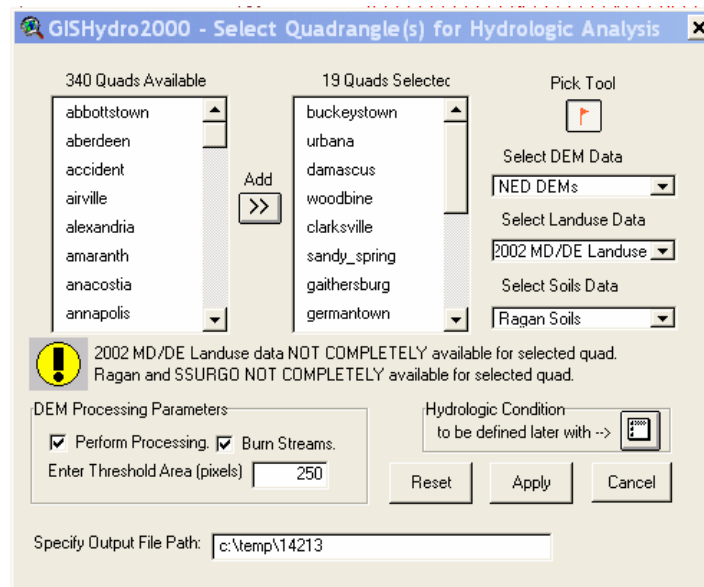
shown in Appendix C-1. The FORTRAN scripts for Region-style models are attached in Appendix C-2.

Manual for using Region-Style Models in GISHydro2000

-- using stream restoration sites in Montgomery County as an example

Please refer to GISHydro2000+ User's Manual (Moglen 2005) for details on how to use GISHydro2000. Here only the necessary steps are provided in order to use Region-style macroinvertebrate predictive models.

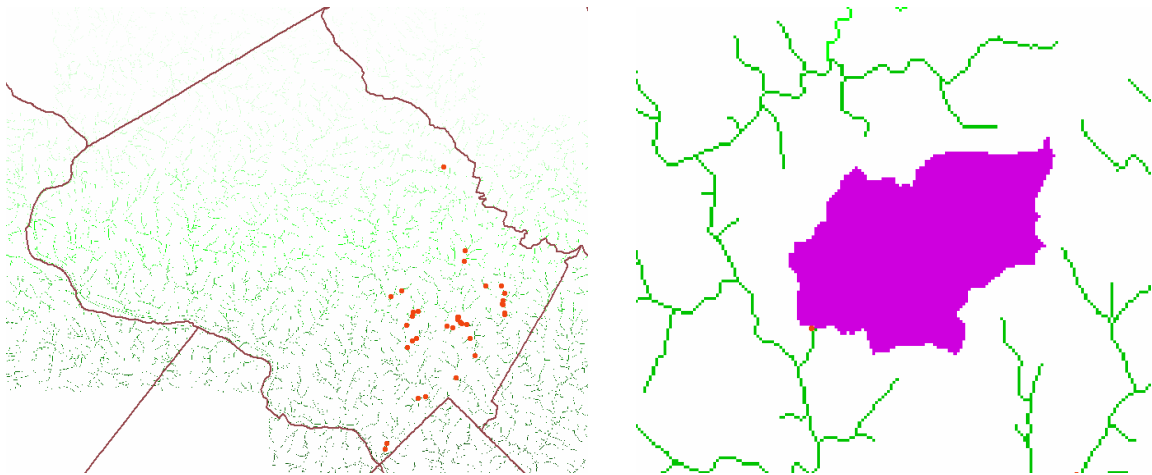
1. Select the quadrangles of interest and decide the Digital Elevation Model (DEM), landuse, and soil data sources.




In GISHydro2000, once you select the desired quadrangles, which includes all of Montgomery County in our case, “DEM”, “landuse”, and “soil data” are processed in the selected area. The default selections for all these data choices are applied in this example. They are NED DEMs, 2002 MD/DE landuse, and Ragan Soil. As

mentioned earlier, GISHydro2000 processes STATSGO soil data automatically, and also the selected soil data, Ragan Soil, is applied here. Note that if you would like to use the Region-style model built with Ragan Soil data, you need to select Ragan as the soil data source. Depending on the number of quadrangles selected, this primary hydrologic process can take a while. A new View named as “Area of Interest” will be generated in ArcView.

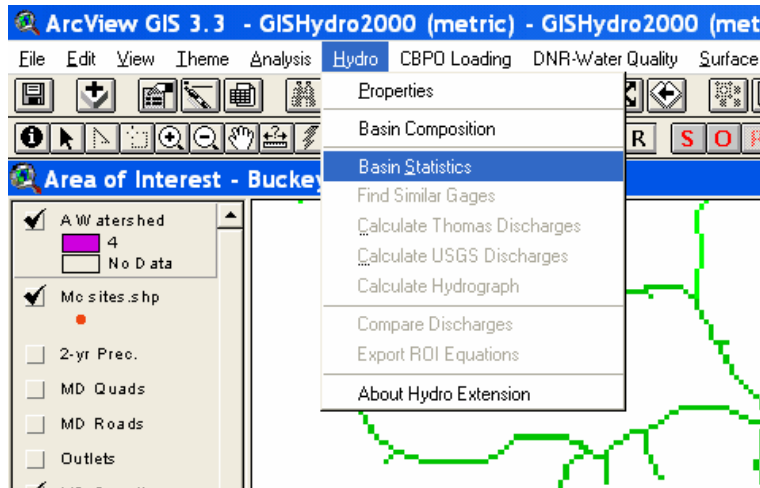
2. Locate the target stream site and delineate the watershed



Once all the map data layers are ready for the area of interest, load the interested stream sites layer (“stream restoration sites of Montgomery County”) onto the map. If leaving only the “Stream Links” and the “stream restoration sites of Montgomery County” layers on the View window, the interested sites on “stream restoration sites of Montgomery County” layer falls approximately onto the stream channels on the “Stream Links” layer as above left figure. Then use the watershed tool, “”, on the tool bar to click the interested stream site. It will treat the stream site as a watershed outlet to delineate its watershed area. ArcView then creates the “A

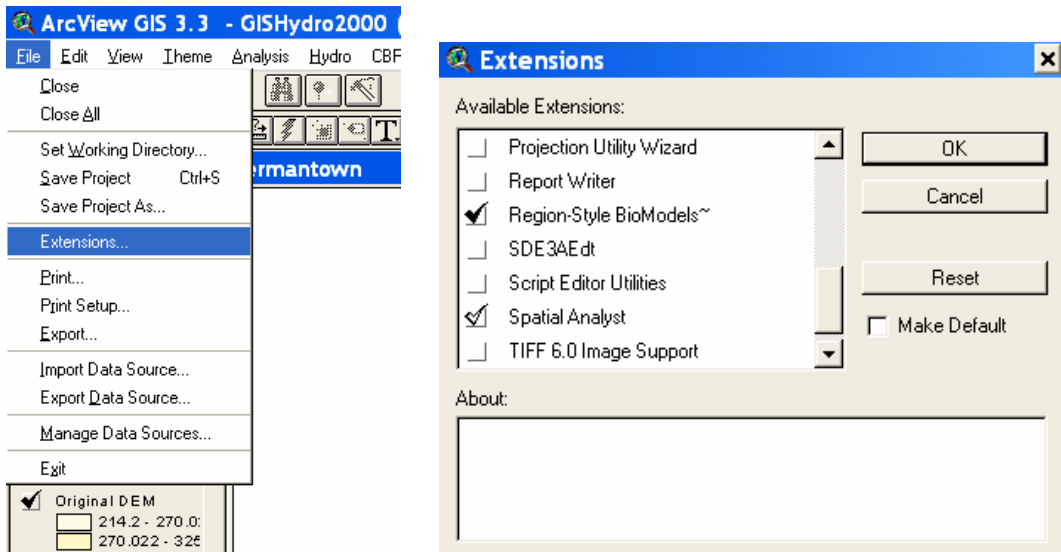
Watershed” theme for the delineated watershed as above right figure.


3. Calculate the GISHydro2000 built-in basin statistics




Click on the “Hydro” on the menu and choose “Basin Statistics” from the choices as above. GISHydro2000 then calculates all the basic basin statistics for the delineated watershed, including outlet location, drainage area, channel slope, land slope, and more. Some of these statistics are used as the predictors for the developed Region-style models. (A pop-up window will allow you to save these basin statistics for future reference.)

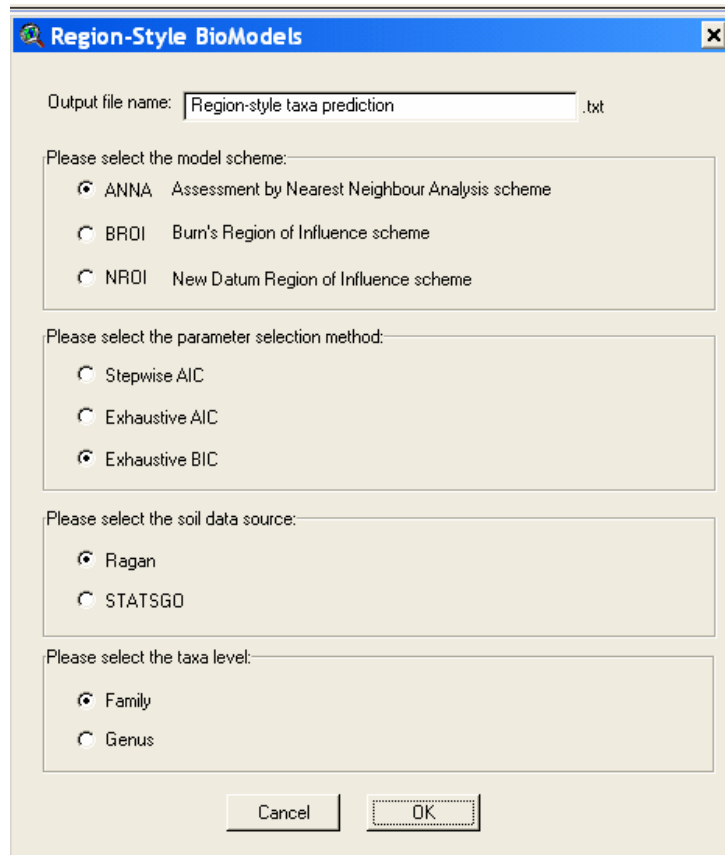
4. Run the Region-style models



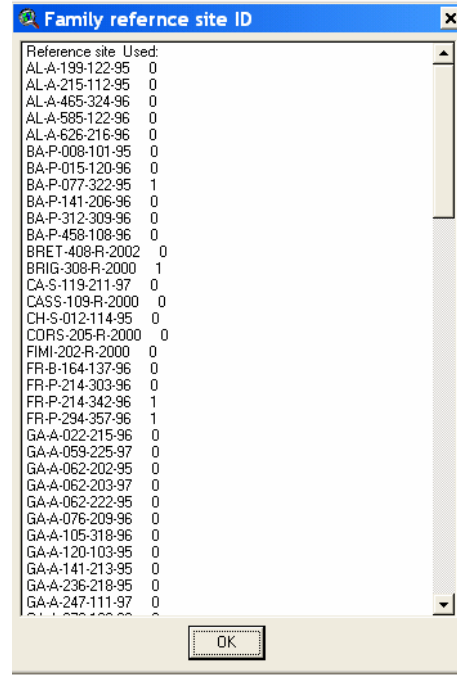
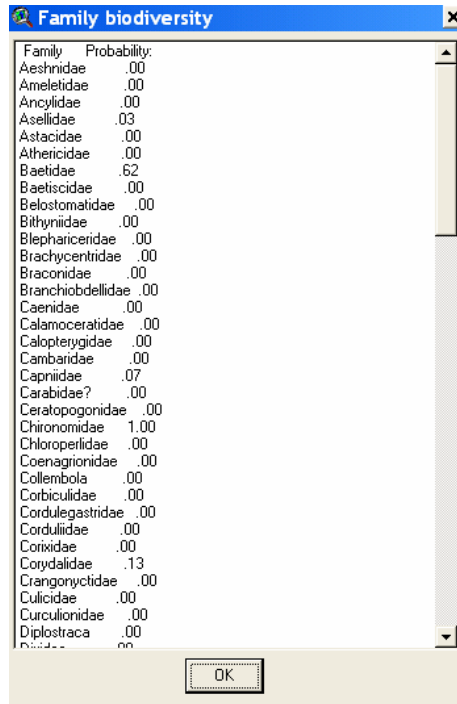
Click on the “File” on the menu and choose the “Extension” from the choice. A extension list window will show up. Scroll down the menu and check the “Region-style Biomodels~”. One “dragonfly” icon “” will show on the tool bar.

Run the Region-style model by clicking on the . A dialog windows will pop-up and ask you to :

- Input the name of the output file.
- Choose the region selection scheme (ANNA, BROI, NROI).
- Choose the parameter selection method (Stepwise AIC, Exhaustive AIC, Exhaustive BIC).
- Select the soil data sources (Ragan, STATSGO).
- Choose the desired prediction taxon level (Family, Genus).

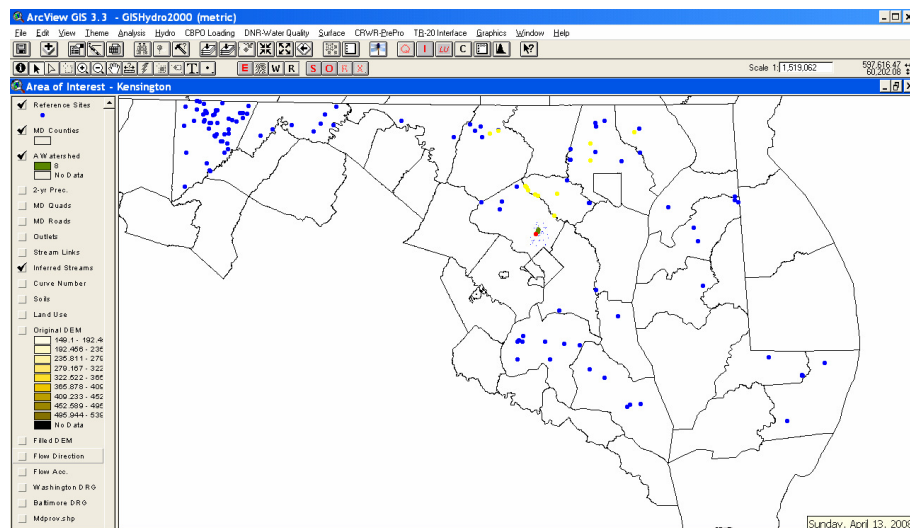


This extension script obtains and calculates all the parameters needed from GIS layers. The built-in FORTRAN executable file chooses the region of influence from MBSS reference sites and calculates the predicted taxon appearance probabilities. The taxon prediction results will be shown in the window as well as the used reference sites.



All the results are saved in a text file (*.txt) under the name of the project that was assigned in the “Region-Style Biomodels” dialogue box.

5. Interpret and utilize the model output



One “Reference Sites” theme will also be added to the “Area of Interest” view. Setting the “Reference Sites” to visible by checking this theme and zooming

7.2.2 Application of Stream Restoration Biomonitoring Data

The pre- and post-stream restoration raw biomonitoring data from Montgomery County need to be processed before the comparison can be done with the predictive model results.

Identifying the observed taxon names in the family level

In our macroinvertebrate predictive model, the models were built for both the family and genus levels. To check the stream restoration site results, the observed macroinvertebrate data need to be in family and genus levels as well. The Montgomery County DEP has identified their macroinvertebrate samples down to the genus level, whenever possible. However, some of the macroinvertebrates may be fragile or not complete, so they can be identified only to family or even order level. There are 124 taxa at the family level and 231 taxa at the genus level in the MBSS data (Section 3.1). The raw observed data were then assigned one by one to the genus name list. From there the family level name of the observed taxon data were looked up depending on the genus level name in the master taxon list table (Southerland et al., 2005, Appendix Table 2). The Integrated Taxonomic Information System (ITIS) (<http://www.itis.gov/>) also provided updated taxon information in our assignation process. The observed taxon data for the family level were then assigned one by one to our family name list. Thus the observed taxon lists in family and genus levels were ready for each stream monitoring restoration site.

Assigning the O according to E

The definition of Observed (O) and Expected (E) were described in Section 2.1.Step5. The observed (O) at each site is counted depending on the taxon occurrence probability (p) predicted from the models and the capture probability (p_c). As in the literature, the capture probability was set as $p_c = 0.5$. If the taxon occurrence probability (p) was larger than capture probability (p_c), then the probability p was included into taxon expected (E) and the taxon observed (presence =1, absence = 0) was included in the calculation of observed (O). All three Region-style models were applied to calculate the occurrence probability (p). The taxon occurrence probability values from different models can be different. The O and E selections thus vary with different models even with same observed data. O and E were summed for different prediction models and the results are shown in the next chapter.

CHAPTER 8: STREAM RESTORATION SITE APPLICATION

RESULTS

Maryland's geology can be described as five distinct landform regions (Smith et al., 2003). These regions are called physiographic provinces. Maryland's highest elevations are found in the Appalachian Plateau. The lowest are in the Coastal Plain. In between from west to east lies the variable landscape of the Ridge and Valley, the narrow mountainous strip called the Blue Ridge, and the broad Piedmont Plateau that transitions to the Coastal Plain surrounding the Chesapeake Bay. Montgomery County is mostly located in the Piedmont region. Part of western Montgomery County belongs to Blue ridge. The stream restoration sites in this study are located in the eastern Montgomery County, the Piedmont region. Region-style models use the reference sites with similar physical features to predict the macroinvertebrate. These selections are shown on the map in Section 8.1 and noted if the physiographic locations of selected reference sites are different than that of the test site. The multivariate and multimetric analysis applying to stream restoration monitoring data are described in Sections 8.2 and 8.3.

8.1 MACROINVERTEBRATE PREDICTIVE RESULTS FROM DIFFERENT

REGION-STYLE MODELS

With GISHydro2000, the different reference sites picked from different Region-style schemes can be examined visually. A Montgomery County stream restoration site, LFLF 301B in Table 34, is shown in this section as an example.

LFLF 301B is the restoration site that falls on the mainstem south of Massachusetts Avenue, in the Little Falls watershed. The Little Falls watershed is one of the County's most urban stream systems, with part of its drainage from the northwest portion of the District of Columbia. This watershed contains some of the oldest developed areas of the County. There are several historic elements in the watershed such as Battery Bailey, the route of the old trolley line, the Washington Aqueduct, and the C&O Canal, all of which have played a role in shaping the watershed's landscape (<http://www.montgomerycountymd.gov/>). In the following graphs, the red point is the restoration site location (LFLF 301B), and the other color points are reference sites in MBSS data that were used to build the models. The points highlighted in yellow are the reference sites used to predict macroinvertebrates in the Region-style models.

8.1.1 The ANNA Scheme

The Region-style model with ANNA scheme requires 21 reference sites in the family level and 16 reference sites in the genus level to predict the presence or absence of taxa. The reference site selection results with ANNA are shown in Figure 33. The family level modeling included the following predictors – HUC14, HUC8, RSB, SIN, P100, HYPISO, LONG, X, and Y. The genus level modeling included the following predictors – WSLOPE, CSLOPE, RELIEF, LIME, HUC14, HUC8, PERIM, RSC, SIN, P100, HIGHELEV, HYPISO, LAT, X, and Y as predictors. The family and genus level models used different parameters and regression functions to estimate the similarity between sites, so the ways that the reference sites were used for prediction were different. The models for both levels used primarily Piedmont

sites for reference. In the family level Figure 33(a), the 21 sites highlighted in yellow are the closest reference sites that have similar physical features to the target stream restoration site. These reference sites are located mostly in the Piedmont, except for two in the Coastal Plain. These reference sites have similar hydrologic unit codes, soil type B percentage, channel sinuosity, 100-year precipitation, hypsometric area ratio, and site location. The parameter averages within different physiography are listed in Table 35 along with the value at LFLF 301B site.

In the genus level Figure 33(b), the 16 sites highlighted in yellow are the closest reference sites with similar physical features to the target stream restoration site. Three of these sites are in the Coastal Plain; one of them is in the Ridge and Valley, while the rest are in the Piedmont. These sites are similar in watershed and channel slope, basin relief, percent of limestone, hydrologic unit codes, watershed perimeter length, soil type C percentage, sinuosity, 100-year precipitation, percent watershed above 200 feet, hypsometric area ratio, and site locations. The parameter averages within different physiography are listed in Table 36 along with the value at LFLF 301B site. Even though the ANNA model at the family level picked more reference sites than the model at the genus level, the reference sites they picked are quite different. The only Coastal Plain reference site picked by the model at the family level wasn't picked by the model at the genus level. The ANNA model at the genus level on the other hand picked another three reference sites in the Coastal Plain. This difference is due to different multiple regression functions used by the two models. This implies that family and genus level models treat the closest neighbor differently. One reference site may have physical features close to the target site at the family

level, but not necessarily at the genus level.

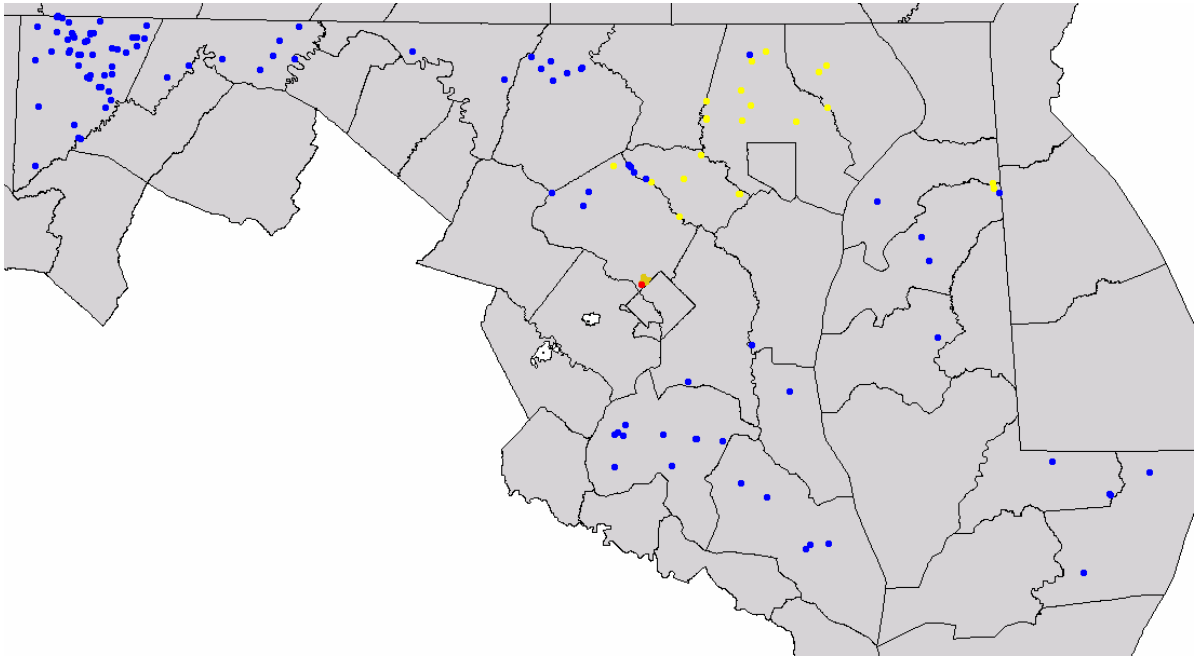
Table 35. The parameter averages of reference sites chosen by ANNA for LFLF 301B stream restoration site in the family level

Parameters	LFLF 301B value	Coastal Sites Average (2)	Piedmont Sites Average (19)	All Sites Average (21)
HUC14	40226	N/A	N/A	N/A
HUC8	2070008	N/A	N/A	N/A
RSB	91.71	80.1	73.93	74.51
SIN	1.33	1.31	1.40	1.39
P100	8.44	8.66	8.51	8.53
HIGHELEV	0	0	0	0
HYPISO	0.56	0.64	0.52	0.54
X	390381	505174.5	420709.7	428754
Y	142842	174940.5	194151.7	192322.1

Table 36. The parameter averages of reference sites chosen by ANNA for LFLF 301B stream restoration site in the genus level

Parameters	LFLF 301B value	Coastal Sites Average (2)	Ridge and Valley Site Average(1)	Piedmont Sites Average (19)	All Sites Average (21)
WSLOPE	0.052	0.066	0.123	0.077	0.076
CSLOPE	59.3	210.97	450.02	99.96	119.13
RELIEF	143.1	38.23	155.20	131.38	121.35
LIME	0	0	0	0	0
HUC14	40226	N/A	N/A	N/A	N/A
HUC8	2070008	N/A	N/A	N/A	N/A
PERIM	12.71	1.17	5	9.04	8.38
RSC	0.98	7	57.1	18.09	16.7
SIN	1.33	1.30	1.19	1.35	1.36
P100	8.44	8.67	6.34	8.40	8.43
HIGHELEV	0	0	0	0	0
HYPISO	0.56	0.61	0.19	0.53	0.54
LAT	38.9534	38.791	39.6337	39.29196	39.27633
X	390381	435165	315721	401882.7	408022
Y	142842	124961.6	218813	180438.9	178734.9

(a) Family Level



(b) Genus Level

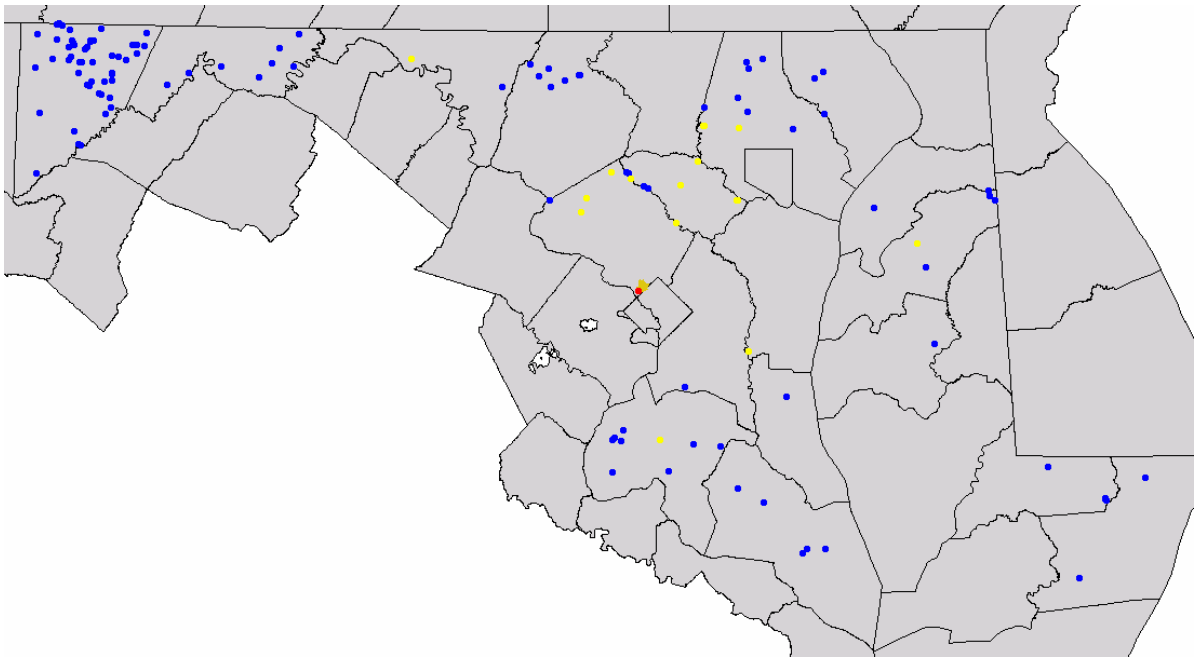


Figure 33. The reference sites used by ANNA for Site LFLF 301B in Montgomery County

8.1.2 The BROI Scheme

The thresholds and weighting functions for the BROI scheme were defined in Section 2.3.2 and the final weighting function parameters chosen for Maryland were listed in Tables 16 and 17. The Region-style model with BROI used eight reference sites to predict family level taxa and twenty-eight reference sites to predict genus level taxa for LFLF 301B site. The reference sites selected by the models are listed in Table 37. Interestingly, the reference sites chosen by these two model have only five sites in common (BA-P-077-322-95, HO-P-068-220-95, HO-P-068-231-96, PATL-124-R-2000, RKGR-101-R-2002). The family level model used parameters – LIME, HUC14, HUC8, LENGTH, RSB, RSD, SIN, P100, HYPISO, LONG, X, and Y as predictors. The genus level model used parameters – WSLOPE, CSLOPE, RELIEF, LIME, HUC14, HUC8, PERIM, RSC, SIN, P100, HIGHELEV, HYPISO, LAT, X, and Y as predictors. The parameter averages within different physiography are listed in Table 38 and 39 along with the value at LBLB 301B site. The family and genus level models used different parameters to measure the similarity, so different reference sites for prediction were used.

Table 37. The reference sites chosen by BROI for LFLF 301B stream restoration site

Model	Reference Sites
Family Level	BA-P-077-322-95, HA-P-001-205-96, HO-P-068-220-95, HO-P-068-231-96, LOCH-404-R-2002, PATL-124-R-2000, QA-N-111-312-95, RKGR-101-R-2002
Genus Level	BA-P-008-101-95, BA-P-077-322-95, BRIG-308-R-2000, CA-S-119-211-97, CASS-109-R-2000, CH-S-012-114-95, HO-P-058-126-97, HO-P-068-220-95, HO-P-068-231-96, HO-P-132-319-97, HO-P-214-311-97, LIBE-102-C-2000, LIBE-102-S-2002, LOCR-102-S-2002, MATT-105-R-2000, MO-P-248-125-96, MO-P-252-323-97, MO-P-366-212-97, MO-P-370-308-97, PATL-124-R-2000, PAXM-121-R-2001, PRWA-101-R-2002, PTOB-002-S-2001, QA-N-066-207-95, RKGR-101-R-2002, SENE-104-R-2001, YOUNG-118-R-2001, ZEKI-104-R-2001

Table 38. The parameter averages of reference sites chosen by BROI for LFLF 301B stream restoration site in the family level

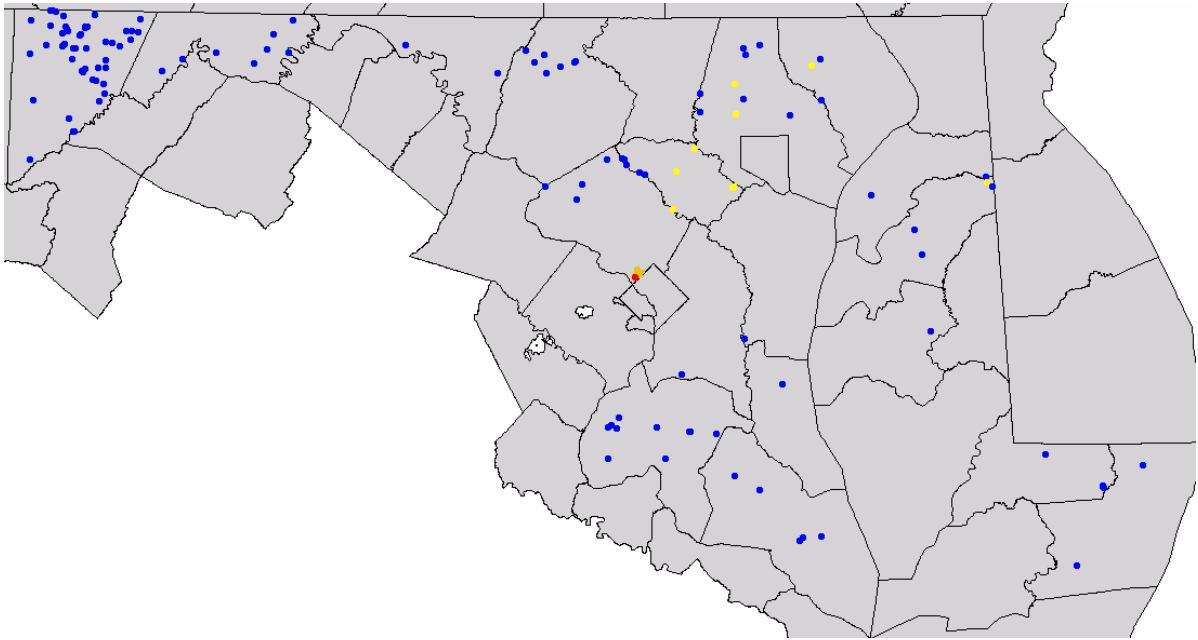
Parameters	LFLF 301B value	Piedmont Sites Average (8)	All Sites Average (8)
LIME	0	0	0
HUC14	40226	N/A	N/A
HUC8	2070008	N/A	N/A
LENGTH	3.35	4.43	4.43
RSB	91.71	68.45	68.45
RSD	7.31	17.68	17.68
SIN	1.33	1.47	1.47
P100	8.44	8.55	8.55
HYPSO	0	0.51	0.51
LAT	0.558916	39.33646	39.33646
X	38.9534	432201.4	432201.4
Y	390381	185482.9	185482.9

Table 39. The parameter averages of reference sites chosen by BROI for LFLF 301B stream restoration site in the genus level

Parameters	LFLF 301B value	Coastal Sites Average (8)	Ridge and Valley Site Average(1)	Appalachian Plateau Site Average (2)	Piedmont Sites Average (17)	All Sites Average (28)
WSLOPE	0.052	0.051	0.122	0.102	0.083	0.09
CSLOPE	59.3	105.61	450.02	104.26	87.47	186.84
RELIEF	143.1	37.86	155.2	152.35	146.18	122.90
LIME	0	0	0	0	0	0.00
HUC14	40226	N/A	N/A	N/A	N/A	N/A
HUC8	2070008	N/A	N/A	N/A	N/A	N/A
PERIM	12.71	4.25	5	12.05	14.16	8.87
RSC	0.98	24.41	57.1	46.05	16.06	35.91
SIN	1.33	1.26	1.19	1.68	1.31	1.36
P100	8.44	8.72	6.34	6.09	8.38	7.38
HIGHELEV	0	0	0	100	0	25.00
HYPSO	0.56	0.60	0.19	0.31	0.53	0.41
LAT	38.9534	38.73388	39.6337	39.5151	39.30119	39.29597
X	390381	429254	315721	200247.5	399593.3	336204
Y	142842	118572.4	218813	207735.5	181460.4	181645.3

The reference sites used for family and genus level model are illustrated in Figures 34(a) and 34(b). The genus level models used 17 sites located in the Piedmont, 8 sites from the Coastal Plain, 1 site from Ridge and Valley, and 2 sites from the Appalachian Plateau. The reference sites used by family level model are all from the Piedmont.

(a) Family Level



(b) Genus Level

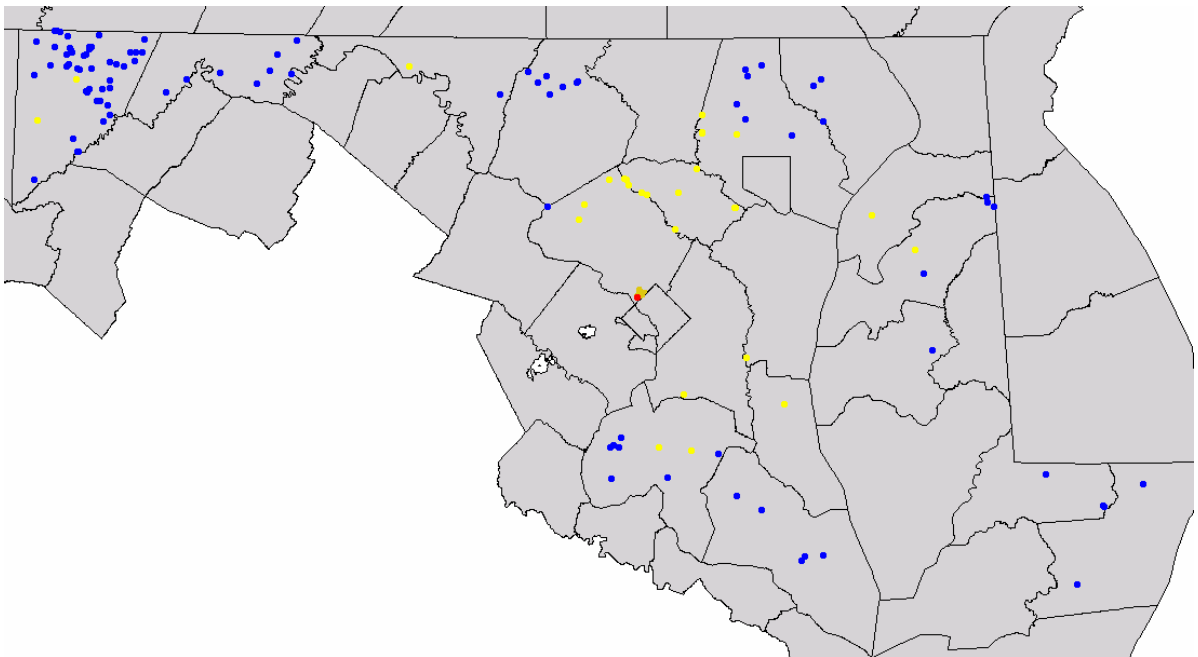


Figure 34. The reference sites used by BROI for site LFLF 301B in Montgomery County

8.1.3 The NROI Scheme

The thresholds and weighting functions for the NROI scheme were defined in Section 2.3.3 and the final weighting function parameters chosen for Maryland were listed in Tables 16 and 17. The Region-style model with NROI scheme used twenty-five reference sites to predict family level taxa and five reference sites to predict genus level taxa for LFLF 301B site. The reference sites selected by the models are listed in Table 40. The reference sites chosen by these two models have five sites in common (HO-P-058-126-97, HO-P-068-220-95, HO-P-068-231-96, PATL-124-R-2000, RKGR-101-R-2002). The family level model used parameters – HUC14, HUC8, LENGTH, RSB, RSD, SIN, HYPISO, LONG, X, and Y as predictors. The genus level model used parameters – WSLOPE, CSLOPE, RELIEF, LIME, HUC14, The parameter averages within different physiography are listed in Table 37 and 38 along with the value at LFLF 301B site. HUC8, PERIM, RSC, SIN, P100, HIGHELEV, HYPISO, LAT, X, and Y as predictors. The family and genus level models used different parameters to measure the similarity, so different reference sites for prediction were used. The reference sites used for family level and genus level model are illustrated in Figures 35(a) and 35(b). The reference sites used by family level models mostly located in the Piedmont and three located in the Coastal Plain. Genus level models use sites located only in the Piedmont as references.

Table 40. The reference sites chosen by NROI for LFLF 301B stream restoration site

Model	Reference Sites
Family Level	BA-P-008-101-95, BA-P-015-120-96, BA-P-077-322-95, BA-P-141-206-96, BA-P-312-309-96, BRIG-308-R-2000, GA-A-505-218-97, HA-P-001-205-96, HA-P-010-103-97, HO-P-058-126-97, HO-P-068-220-95, HO-P-068-231-96, HO-P-132-319-97, HO-P-214-311-97, KE-N-046-226-95, LIBE-102-C-2000, LIBE-102-S-2002, LIGU-312-R-2001, LOCH-404-R-2002, MO-P-252-323-97, MO-P-370-308-97,

	PATL-124-R-2000, PAXM-121-R-2001, QA-N-111-312-95, RKGR-101-R-2002
Genus Level	HO-P-058-126-97, HO-P-068-220-95, HO-P-068-231-96, PATL-124-R-2000, RKGR-101-R-2002

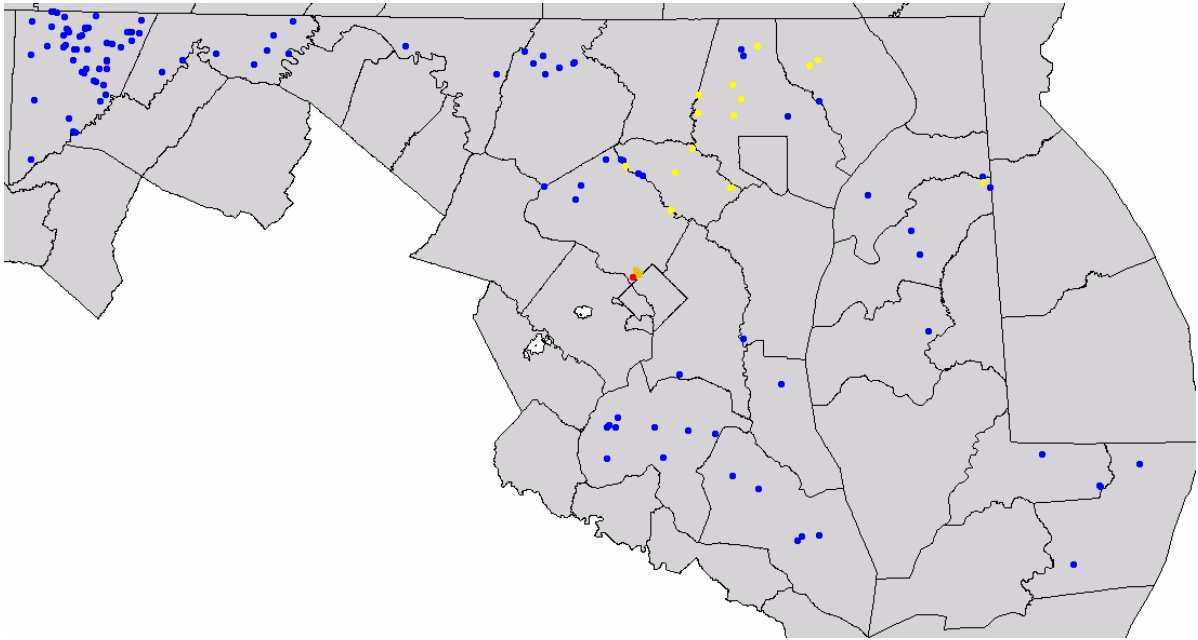
Table 41. The parameter averages of reference sites chosen by NROI for LRLR 301B stream restoration site in the Family level

Parameters	LFLF 301B value	Coastal Sites Average (3)	Piedmont Sites Average (22)	All Sites Average (25)
HUC14	40226	N/A	N/A	N/A
HUC8	2070008	N/A	N/A	N/A
LENGTH	3.35	0.36	5.44	4.83
RSB	91.71	65.17	70.31	69.70
RSD	7.31	17.87	15.05	15.39
SIN	1.33	1.32	1.36	1.35
HYPISO	8.44	0.60	0.52	0.53
LAT	38.9534	39.0819	39.40528	39.36648
X	390381	478896	405951.4	414704.8
Y	142842	157581	193145.5	188877.8

Table 42. The parameter averages of reference sites chosen by NROI for LRLR 301B stream restoration site in the Genus level

Parameters	LFLF 301B value	Piedmont Sites Average (8)	All Sites Average (8)
WSLOPE	0.052	0.060	0.060
CSLOPE	59.3	97.62	97.62
RELIEF	143.1	128.46	128.46
LIME	0	0	0
HUC14	40226	N/A	N/A
HUC8	2070008	N/A	N/A
PERIM	12.71	8.18	8.18
RSC	0.979746	17.36	17.36
SIN	1.33	1.37	1.37
P100	8.44	8.5	8.5
HIGHELEV	0	0	0
HYPISO	0.56	0.53	0.53
LAT	38.9534	39.2381	39.2381
X	390381	412387.2	412387.2
Y	142842	174453	174453

(a) Family Level



(b) Genus Level

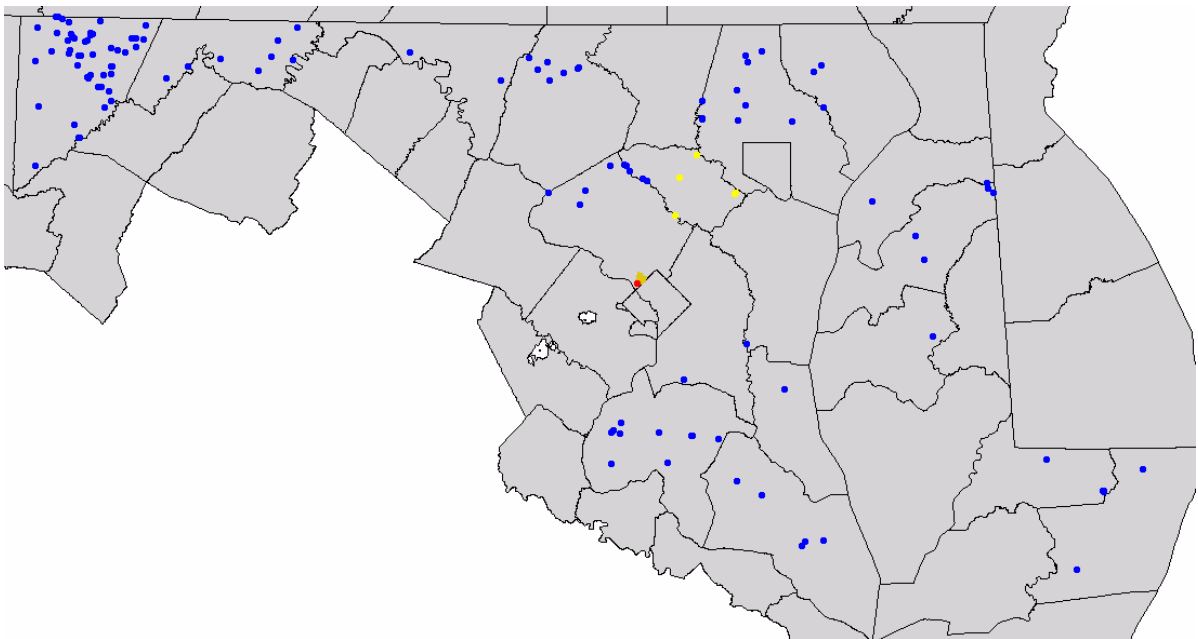


Figure 35. The reference sites used by NROI scheme for site LFLF 301B in Montgomery County

8.2 THE MACROINVERTEBRATE PREDICTIVE RESULTS AT STREAM

RESTORATION SITES

After running the models in GISHydro2000, all the probabilities for taxon presence at each monitoring stream restoration site were predicted. The expected taxon predictive probabilities for each stream restoration site are listed in Appendix E. The predictive probabilities (E) are summarized in Table 43 for family level and Table 45 for genus level. All three Region-style model predictive results are listed along with their comparable observed results (O). The observed data from each monitored restoration site are listed according to the time after restoration. The monitoring times are also listed along with the observed data.

These restoration sites were assumed to be in poor aquatic habitat condition prior to the restoration projects, which was one of the reasons for the restoration plan. Habitat condition should be improving through time after the restoration practice was implemented. Thus, the pre-restoration monitoring were expected to have low observed taxon values (O) compared to our predicted taxon values (E), and the post-restoration monitoring observed taxon values should have increased to more closely approximate the predicted taxon values. If these expectations are fulfilled, these restoration projects have successfully improved the stream habitat conditions.

8.2.1 Family Level

The family level results are summarized in Table 43. The prediction values from different Region-style models are listed together on the left of the table, next to the Site ID column. The three predictions (ANNA, BROI, NROI) from Region-style

models are different but close in values at each site. The pre-construction observed data are listed next to the prediction results.

These observed data results show that half of the restoration sites met or almost met the predicted expected taxon numbers in the family level before stream restoration. The other half of the restoration sites had lower than expected taxon numbers, which implies degraded stream sites. After restoration, those degraded stream sites showed little or no improvement with time. Some of the observed taxon values increased in the most recent monitoring data (LRJB 203B, LRSB 101C). One of the sites kept the same observed taxon value through the monitoring period (LRSB 101A). Some of the monitoring sites that met the predicted taxon number before restoration kept the same observed taxon values in the most recent monitoring data (LRJB 203A, PBPB 305A). Many of them, however, have worse observed data than before restoration (NWLR 102, NWLT 101, NWNW 407A, NWNW 407B, NWNW 407C, NWNW 407D, NWNW 408A, NWNW 408B, PBGS 108, PBPB 305B, PBPB 305C, PBPB 305D, PBPB 305E). It may imply that stream habitat for macroinvertebrate at these sites was still degraded after stream restoration construction. It is also interesting to find that at some monitoring sites the observed taxon values decreased at the first year after construction and then the observed taxon values go back up later (LRJB 203B, LRJB 204, LRSB 101C, NWLR 102, NWLT 101, NWNW 407A, NWNW 407B, NWNW 407C, NWN 407D, PBPB 305F). This may imply that the restoration practices damaged the habitat for benthic macroinvertebrates during construction, but the ecosystem recovered afterward. Many of these observed taxon values that were monitored in later years did not increase to

more than the value from before construction. This may imply that the recovery of ecosystems did not improve from the previously disturbed conditions, or more time is needed to recover. Observed taxon data at two monitoring sites increased in the first year after restoration, but then the values decreased in later year's monitoring (PBPB 305B, PBPB 305D). This may imply that the stream ecosystem was improved after stream restoration practices at the beginning, but the restoration effectiveness did not last through the years, or that external pressures overshadowed the restoration improvements. There are also sites that observed taxon numbers just keep going down even with stream restoration at the stream sites (NWNW 408A, NWNW 408B, PBPB 305E). The restoration effort at these sites demonstrated no positive effects.

The O/E scores for each stream restoration site were also calculated and summarized in Table 44. The O/E scores obtained from different models vary but not by much. The O/E score trends are similar to the results in Table 43. The O/E values indicate if the stream sites are close to the reference condition by how close the O/E scores are to one. The only restoration sites that are close to reference conditions at the latest monitoring are sites PBPB 305A, and maybe PBPB 305C. These sites already had observed taxon values close to expected taxon values prior to restoration. Their O/E scores have little room for improvement and are no better than the scores before construction. Sites LRJB 203B and LRSB 101C have better O/E scores in the latest monitoring data. The O/E score at site LRJB 203A increased from 0.3 to 0.4 in two years. The O/E scores at sites LRSB 101C increased from 0.1 to 0.3 in three years. The O/E scores of the rest of the stream restoration sites either didn't improve from the pre-construction values or decreased.

Note that the predictive model is predicting the expected taxa in reference conditions. The degrading trend in the table does not necessarily mean the stream has fewer and fewer benthic macroinvertebrates, but it does mean the taxa that were expected if the site were in “reference” condition did not show up in the stream. There may be other benthic species, even large numbers of them, in the stream, but they are not expected to be found in reference conditions.

Finally, there are some stream restoration sites that still have the same observed taxon values as they had before restoration. Some restoration sites have better observed taxon values at the latest monitoring than they had before restoration. The increment of the values however is limited (only one taxon increment). The observed taxon values at many sites did not meet the expected taxon values during the few years for which post-construction data are available. Their stream habitat for benthic macroinvertebrate habitat was not improved enough to meet the reference habitat conditions by the restoration project, or the expected taxa haven't shown up after stream restoration. If the stream restoration is taking effect, more time is required for the recovery process to have an effect on the stream macroinvertebrate populations.

Table 43. The summary table of Expected (E) and Observed (O) at the Family Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Family Level

Sites	E			O																				
	Ragan	Ragan	Ragan	pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later		
Soil Data	ANNA	BROI	NROI	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan
Predictive Scheme	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC
LFLF 301B	2001						2002									2005								
	9.05	7.88	8.97				3	3	3							3	3	3						
LFLF 301C	2001			2002																				
	8.48	7.56	8.71	3	3	3																		
LRCR 101A				2002																				
	8.45	8.89	9.51	1	1	1																		
LRCR 101B				2002																				
	7.73	9.2	8.2	3	3	3																		
LRJB 203A	2005			2002			2005			2007														
	9.19	9.71	8.79	3	3	3	3	3	3	3	3	3												
LRJB 203B	2005			2002			2005			2007														
	9.2	9.27	9.5	3	3	3	2	2	2	4	4	4												
LRJB 204	2005			2003			2005			2007														
	9.18	9.28	9.39	3	3	3	1	1	1	3	3	3												
LRSB 101A	2004			2003			2005						2007											
	8.37	10.63	9.07	2	2	2	2	2	2				2	2	2									
LRSB 101C	2004			2002			2005						2007											
	7.68	10.13	8.89	2	2	2	1	1	1				3	3	3									
LRTB 101				2002																				
	7.94	10.65	8.98	3	3	3																		
				2006			2	2	2															
LRTB 202				2002																				
	7.75	10.64	8.43	2	2	2																		
				2006			1	1	1															
LRTB 203A				2002																				
	7.89	9.23	8.54	2	2	2																		
				2006			3	3	3															

Table 43-Continued. The summary table of Expected (E) and Observed (O) at the Family Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Family Level

Sites	E			O pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later		
	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan
	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC
LRTB 203B	7.89	9.23	8.48	2002 5	5	5																		
				2006 2	2	2																		
NWGT 101	9.08	8.94	9.74	2005 2	2	2																		
NWL R 102	2002 9.13	9.67	9.36	2001 4	4	3	2002 2	2	2				2004 3	3	3				2007 3	3	3			
NWL T 101	2002 8.46	9.72	8.73	2001 4	4	4	2002 2	2	2				2004 4	4	4				2007 3	3	3			
NWNW 301	7.92	9.23	8.54	2001 8	8	8																		
				2002 7	7	7																		
				2004 6	7	6																		
NWNW 304	8.46	9.7	8.74	1997 10	10	10	2001 9	9	8															
				2001 9	9	8																		
				2004 6	6	6																		

Table 43-Continued. The summary table of Expected (E) and Observed (O) at the Family Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Family Level

Sites	E			O																																
				pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later														
	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan		
ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI				
Parameter Selection	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC			
NWNW 407A	2002			2002			2003						2005					2007																		
	9.17	8.31	8	8	8	8	6	6	6				4	4	4			5	5	5																
NWNW 407B	2002			2002			2003						2005					2007																		
	9.17	8.31	8	8	8	8	7	6	6				2	2	2			7	7	7																
NWNW 407C	2002			2002			2003						2005					2007																		
	9.18	8.31	7.99	8	8	8	4	4	4				7	6	6			6	6	6																
NWNW 407D	2002			2002			2003						2005					2007																		
	9.11	8.31	7.97	7	7	7	5	5	5				4	4	4			6	6	6																
NWNW 408A	2002			2002			2003						2005					2007																		
	9.08	8.3	7.97	7	7	7	6	6	6				3	3	3			3	3	3																
NWNW 408B	2002			2002			2003						2005					2007																		
	9.19	8.62	7.9	6	6	6	5	4	4				4	4	4			5	4	4																
NWPT 101				2005																																
	8.51	9.6	9.83	7	6	8																														
PBGH 108	2001			2000			2002						2004				2005																			
	8.45	9.25	9.53	8	8	8	8	8	8				7	7	7	7	7	7	7	7																
PBGH 102B				1999									2004																							
	8.53	10.13	10	7	7	7							6	6	6																					
				2004																																
			6	6	6																															
SCFT 101	1994																																			
	8.23	10.64	10.6																									2004								
																												1	1	1						

Table 43-Continued. The summary table of Expected (E) and Observed (O) at the Family Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Family Level

Sites	E			O																										
				pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later								
Soil Data	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan			
Predictive Scheme	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC
PBGS 111				1999			2000			2001			2002			2003			2004			2005								
	8.54	10.13	9.9	8	8	8	9	9	9	9	9	9	7	7	7	7	7	7	7	7	7	8	8	8						
				2000																										
				9	9	9																								
				2001																										
				9	9	9																								
				2002																										
				7	7	7																								
				2003																										
			7	7	7																									
			2004																											
			7	7	7																									
			2005																											
			8	8	8																									
PBPB 305A	9.18	8.08	8.97	2001			2002			2003						2005						2007								
				9	9	10	9	8	9	5	5	6				9	8	9				9	9	9						
PBPB 305B	9.17	7.91	8.41	2001			2002			2003						2005						2007								
				8	8	9	9	9	9	8	8	8				9	8	8				4	4	4						
PBPB 305C	9.19	8.08	8.5	2001			2002			2003			2004			2005						2007								
				9	8	8	8	8	8	8	8	9	6	6	7	5	5	6				7	7	7						
PBPB 305D	2001			2001			2002									2005														
	9.11	8.01	8.47	6	6	6	7	7	7							5	5	5												
PBPB 305E	9.12	8.01	8.48	2001			2002			2003						2005														
				8	8	9	7	7	8	7	7	7				4	4	4												
PBPB 305F	9.21	8.34	8.68	2001			2002			2003						2005														
				9	9	10	7	7	8	5	5	5				9	8	8												

Table 44. The summary table of O/E scores at the Family Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Family Level

Sites	O/E																				
	pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later		
	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan
	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC
LFLF 301B				0.331	0.381	0.334							0.331	0.381	0.334						
LFLF 301C				0.354	0.397	0.344															
LRCR 101A	0.118	0.112	0.105																		
LRCR 101B	0.388	0.326	0.366																		
LRJB 203A	0.326	0.309	0.341	0.326	0.309	0.341	0.326	0.309	0.341												
LRJB 203B	0.326	0.324	0.316	0.217	0.216	0.211	0.435	0.431	0.421												
LRJB 204	0.327	0.323	0.319	0.109	0.108	0.106	0.327	0.323	0.319												
LRSB 101A	0.239	0.188	0.221	0.239	0.188	0.221				0.239	0.188	0.221									
LRSB 101C	0.260	0.197	0.225	0.130	0.099	0.112				0.391	0.296	0.337									
LRTB 101	0.378	0.282	0.334																		
	0.252	0.188	0.223																		
LRTB 202	0.258	0.188	0.237																		
	0.129	0.094	0.119																		
LRTB 203A	0.253	0.217	0.234																		
	0.380	0.325	0.351																		
LRTB 203B	0.634	0.542	0.590																		
	0.253	0.217	0.236																		
NWGT 101	0.220	0.224	0.205																		
NWLR 102	0.438	0.414	0.321	0.219	0.207	0.214				0.329	0.310	0.321				0.329	0.310	0.321			
NWLT 101	0.473	0.412	0.458	0.236	0.206	0.229				0.473	0.412	0.458				0.355	0.309	0.344			
NWNW 301	1.010	0.867	0.937																		
	0.884	0.758	0.820																		
	0.758	0.758	0.703																		
NWNW 304	1.182	1.031	1.144	1.064	0.928	0.915				0.709	0.619	0.686									
	1.064	0.928	0.915																		
	0.709	0.619	0.686																		

Table 44-continued. The summary table of O/E scores at the Family Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Family Level

Sites	O/E																				
	pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later		
	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan
	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC	BIC	Stepwise AIC	AIC
NWNW 407A	0.872	0.963	1	0.654	0.722	0.75				0.436	0.481	0.5				0.545	0.602	0.625			
NWNW 407B	0.872	0.963	1	0.763	0.722	0.75				0.218	0.241	0.25				0.763	0.842	0.875			
NWNW 407C	0.871	0.963	1	0.436	0.481	0.501				0.763	0.722	0.751				0.654	0.722	0.751			
NWNW 407D	0.768	0.842	0.878	0.549	0.602	0.627				0.439	0.481	0.502				0.659	0.722	0.753			
NWNW 408A	0.771	0.843	0.878	0.661	0.723	0.753				0.330	0.361	0.376				0.330	0.361	0.376			
NWNW 408B	0.653	0.696	0.759	0.544	0.464	0.506				0.435	0.464	0.506				0.544	0.464	0.506			
NWPT 101	0.823	0.625	0.814																		
PBGH 108	0.947	0.865	0.839	0.947	0.865	0.839				0.828	0.757	0.735	0.828	0.757	0.735						
PBGS 102B	0.821	0.691	0.700							0.703	0.592	0.600									
	0.703	0.592	0.600																		
PBGS 111	0.937	0.790	0.808																		
	1.054	0.888	0.909																		
	1.054	0.888	0.909																		
	0.820	0.691	0.707																		
	0.820	0.691	0.707																		
	0.820	0.691	0.707																		
	0.937	0.790	0.808																		
PBPB 305A	0.980	1.114	1.115	0.980	0.990	1.003	0.545	0.619	0.669				0.980	0.990	1.003				0.980	1.114	1.003
PBPB 305B	0.872	1.011	1.070	0.981	1.138	1.070	0.872	1.011	0.951				0.981	1.011	0.951				0.436	0.506	0.476
PBPB 305C	0.979	0.990	0.941	0.871	0.990	0.941	0.871	0.990	1.059	0.653	0.743	0.824	0.544	0.619	0.706				0.762	0.866	0.824
PBPB 305D	0.659	0.749	0.708	0.768	0.874	0.826							0.549	0.624	0.590						
PBPB 305E	0.877	0.999	1.061	0.768	0.874	0.943	0.768	0.874	0.825				0.439	0.499	0.472						
PBPB 305F	0.977	1.079	1.152	0.760	0.839	0.922	0.543	0.600	0.576				0.977	0.959	0.922						
SCFT 101																			0.122	0.094	0.094

8.2.2 Genus Level

The genus level taxon data at stream restoration sites generally did not meet the prediction results from the Region-style models. The genus level results are summarized in Table 45. The prediction results from Region-style models have expected taxon values (E) generally above five and some of them are even above ten. The observed taxon values for the genus level are generally below five. Some monitoring sites recorded observed taxon values of zero. At the genus level, the observed taxon values at some monitoring sites kept decreasing after restoration and were worse than the values before construction (LRJB 203B, NWNW 408A, PBGH 108, PBPB 305A, PBPB 305B, PBPB 305C, PBPB 305E, PBPB 305F). This implies that degradation of stream habitats continues even after restoration. Some observed taxon data shows improved values in the latest monitoring data even though the increases were limited (LRJB 203A, LRJB 204, NWLR 102, NWLT 101, NWNW 407B, NWNW 407C, PBPB305D). Some observed taxon values show decreases at the beginning and improvements over time (NWNW 407A, NWNW 407B, NWNW 407D, PBPB 305A, PBPB 305C, PBPB 305E). This implies that the macroinvertebrate population is recovering from the construction impact. Some observed taxon values increased right after restoration but decreased later (PBPB 305D, PBPB 305F), which might imply that the environmental stress overshadowed the stream restoration effect or that the environmental stress increased in later years.

The O/E scores for each stream restoration site were also calculated and summarized in Table 46. The O/E scores obtained from different models vary but not by much. The O/E score trend results are similar to the genus level results in Table

45. The O/E values signify how close the stream sites are to reference conditions. Sites with O/E scores close to 1.00 are close to reference conditions. There is no restoration site that is close to reference conditions in the latest monitoring data. The highest O/E score in that data is 0.434 at site PBPB 305D. This number tells us that the site is not really close to the reference conditions. However, more sites show improvement from the genus level point of view than from family level, including LRJB 203A, LRJB 204, NWLR 102, NWLT 101, NWNW 407C and PBPB 305D. The O/E scores at sites LRJB 203A and LRJB 203 increased from around 0.1 to 0.2 in two years. The O/E score at site NWLT 101 increased from around 0.1 to 0.2 in five years. The O/E score at NWNW 407C increased from around 0 to 0.2 in 5 years as well. The O/E score at site PBPB 305D increased from around 0.3 to 0.4 in four years. These improvements in the genus level are encouraging but at the same time the scores are considered small. The O/E scores of the rest of the stream restoration sites either did not improve from the pre-construction levels or decreased afterward.

From the genus level point of view, a lot of restoration sites begin with poor observed taxon conditions. The results of stream restoration do not show much improvement and some of them show worse taxon observed after restoration. Again, a predictive model predicts the taxon composition in reference conditions. Failure to meet the expectation means the expected taxa were not present in the restoration site. The taxa other than the expected ones may still exist in the stream monitoring sites.

Table 45. The summary table of Expected (E) and Observed (O) for the Genus Level at stream restoration sites in Montgomery County.
Genus Level:

Sites	E			O																				
	pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later					
Soil Data	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan
Predictive scheme	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC
LFLF 301B	2001						2002									2005								
	5.59	4.82	6.45				0	0	0							0	0	0						
LFLF 301C	2001			2002																				
	5.59	4.92	6.38	0	0	0																		
LRRCR 101A				2002																				
	6.21	5.54	9.6	0	0	0																		
LRRCR 101B				2002																				
	6.03	5.46	9.25	0	0	0																		
LRJB 203A	2005			2002			2005			2007														
	4.82	5.55	9.59	0	0	1	0	0	1	0	0	2												
LRJB 203B	2005			2002			2005			2007														
	5.4	5.54	9.84	0	0	1	0	0	1	0	0	0												
LRJB 204	2005			2003			2005			2007														
	5.4	5.54	9.87	0	0	1	0	0	0	0	0	2												
LRSB 101A	2004			2003			2005						2007											
	7.1	5.75	10.72	0	0	0	0	0	0				0	0	1									
LRSB 101C	2004			2002			2005						2007											
	6.25	5.59	11.01	0	0	0	0	0	0				0	0	0									
LRTB 101				2002																				
	6.24	5.71	11.06	0	0	1																		
				2006																				
				0	0	2																		
LRTB 202				2002									2006											
	4.84	4.38	9.93	0	0	1							0	0	0									
				2006																				
				0	0	0																		

Table 45-continued. The summary table of Expected (E) and Observed (O) for the Genus Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Genus Level:

Sites	E			O																							
				pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later					
Soil Data	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan			
Predictive scheme	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC
LRTB 203A	4.84	5.41	10.1	2002 0		1																					
				2006 0	0	1																					
LRTB 203B	4.82	5.41	9.38	2002 1	1	1							2006 0	0	1												
				2006 0	0	1																					
NWGT 101	5.05	4.39	8.68	2005 0	0	0																					
NWLR 102	2002 5.79		11.12	2001 0	0	1	2002 0	0	0				2004 0	0	1				2007 0	0	2						
NWLT 101	2002 7.12	5.77	10.58	2001 0	0	1	2002 0	0	0				2004 0	0	0				2007 0	0	2						
NWNW 301	4.54	3.78	10.65	2001 3	2	5																					
				2002 2	1	4																					
				2004 1	1	1																					
SCFT 101	1994 4.81	5.32	9.07																			2004 0	0	0			

Table 45-continued. The summary table of Expected (E) and Observed (O) for the Genus Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Genus Level:

Sites	E			O																					
				pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later			
	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan		
	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	
AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	
NWNW 304	11.54	12.75	12.51	1997			2001						2004												
				3	2	4	3	2	3				2	1	2										
				2001																					
NWNW 407A	2002			2002			2003						2005							2007					
	4.22	4.41	4.63	1	1	1	0	0	1				1	1	1					1	1	1			
				2002			2003						2005							2007					
NWNW 407B	2002			2002			2003						2005							2007					
	4.22	5.36	4.63	0	0	1	0	0	0				0	0	0					1	1	0			
				2002			2003						2005							2007					
NWNW 407C	2002			2002			2003						2005							2007					
	4.23	4.41	4.63	0	0	0	0	0	0				1	1	0					1	1	0			
				2002			2003						2005							2007					
NWNW 407D	2002			2002			2003						2005							2007					
	4.22	5.36	4.64	1	1	1	0	0	1				1	1	1					1	1	1			
				2002			2003						2005							2007					
NWNW 408A	2002			2002			2003						2005							2007					
	3.76	4.41	4.12	1	1	1	1	0	2				1	0	0					0	0	0			
				2002			2003						2005							2007					
NWNW 408B	2002			2002			2003						2005							2007					
	3.24	5.35	4.08	1	1	0	1	1	0				0	0	0					1	1	0			
				2005																					
NWPT 101	6.24	5.76	10.45	4	4	6																			
PBGH 108	2001			2000			2002						2004						2005						
	5.4	4.76	9.83	4	3	6	3	3	5				2	1	5				0	0	2				
PBGH 102B				1999																					
	7.73	4.45	11.03	6	4	7																			
				2004																					
			3	2	3																				

Table 45-continued. The summary table of Expected (E) and Observed (O) for the Genus Level at stream restoration sites in Montgomery County. (Please refer to Table 34 for Site ID.)

Genus Level:

Sites	E			O																							
				pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later					
Soil Data	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan			
Predictive scheme	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC
PBGS 111	8.17	5.04	10.83	1999																							
PBGP 305A	4.94	4.36	6.86	2000																							
				2	2	1																					
				2001																							
				2	2	1																					
				2002																							
				1	1	1																					
				2003																							
				1	1	1																					
				2004																							
				1	1	0																					
				2005																							
				2	1	2																					
				2001			2002			2003						2005						2007					
	4.94	4.36	6.86	3	2	4	1	1	3	0	0	1				0	0	1				2	1	2			
PBGP 305B	5.03	5.45	6.57	2001			2002			2003						2005						2007					
				2	2	4	2	2	4	2	2	3				1	1	3				0	0	1			
PBGP 305C	5.02	4.88	6.13	2001			2002			2003			2004			2005						2007					
				4	5	5	2	2	4	3	3	3	1	1	2	0	0	1				0	0	1			
PBGP 305D	5.75	5.46	6.88	2001			2002									2005											
				1	1	2	2	2	3							1	1	3									
PBGP 305E	5.75	5.46	6.9	2001			2002			2003						2005											
				3	3	3	2	2	4	1	1	3				2	2	1									
PBGP 305F	5.53	5.69	5.95	2001			2002			2003						2005											
				2	2	4	3	3	3	1	1	2				1	1	2									

Table 46. The summary table of O/E scores for the Genus Level at stream restoration sites in Montgomery County.

Genus Level

Sites	O/E																				
	pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later		
	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan
	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC
LFLF 301B				0	0	0							0	0	0						
LFLF 301C	0	0	0																		
LRCR 101A	0	0	0																		
LRCR 101B	0	0	0																		
LRJB 203A	0	0	0.104	0	0	0.104	0	0	0.207												
LRJB 203B	0	0	0.102	0	0	0.102	0	0	0												
LRJB 204	0	0	0.101	0	0	0	0	0	0.203												
LRSB 101A	0	0	0	0	0	0				0	0	0.093									
LRSB 101C	0	0	0	0	0	0				0	0	0									
LRTB 101	0	0	0.090																		
	0	0	0.181																		
LRTB 202	0	0	0.101																		
	0	0	0																		
LRTB 203A	0	0	0.099																		
	0	0	0.099																		
LRTB 203B	0.207	0.185	0.107																		
	0	0	0.107																		
NWGT 101	0	0	0																		
NWLR 102	0	0	0.090	0	0	0				0	0	0.090				0	0	0.180			
NWLT 101	0	0	0.095	0	0	0				0	0	0				0	0	0.189			
NWNW 301	0.661	0.529	0.469																		
	0.441	0.265	0.376																		
	0.220	0.265	0.094																		
NWNW 304	0.260	0.157	0.320																		
	0.260	0.157	0.240																		
	0.173	0.078	0.160																		

Table 46-continued. The summary table of O/E scores for the Genus Level at stream restoration sites in Montgomery County.

Genus Level

Sites	O/E																				
	pre-construction			1 yr later			2 yr later			3 yr later			4 yr later			5 yr later			6 yr later		
	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan	Ragan
	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI	ANNA	BROI	NROI
Parameter Selection	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC	AIC
NWNW 407A	0.237	0.227	0.216	0	0	0.216				0.237	0.227	0.216				0.237	0.227	0.216			
NWNW 407B	0	0	0.216	0	0	0				0	0	0				0.237	0.187	0			
NWNW 407C	0	0	0	0	0	0				0.236	0.227	0				0.236	0.227	0			
NWNW 407D	0.237	0.187	0.216	0	0	0.216				0.237	0.187	0.216				0.237	0.187	0.216			
NWNW 408A	0.266	0.227	0.243	0.266	0	0.485				0.266	0	0				0	0	0			
NWNW 408B	0.309	0.187	0	0.309	0.187	0				0	0	0				0.309	0.187	0			
NWPT 101	0.641	0.694	0.574																		
PBGH 108	0.741	0.630	0.610	0.556	0.630	0.509				0.370	0.210	0.509	0	0	0.203						
PBGS 102B	0.776	0.899	0.635																		
	0.388	0.449	0.272																		
PBGS 111	0.245	0.397	0.092																		
	0.245	0.397	0.092																		
	0.245	0.397	0.092																		
	0.122	0.198	0.092																		
	0.122	0.198	0.092																		
	0.122	0.198	0																		
	0.245	0.198	0.185																		
PBPB 305A	0.607	0.459	0.583	0.202	0.229	0.437	0.000	0.000	0.146				0.000	0.000	0.146				0.405	0.229	0.292
PBPB 305B	0.398	0.367	0.609	0.398	0.367	0.609	0.398	0.367	0.457				0.199	0.183	0.457				0	0	0.152
PBPB 305C	0.797	1.025	0.816	0.398	0.410	0.653	0.598	0.615	0.489	0.199	0.205	0.326	0.000	0.000	0.163				0	0	0.163
PBPB 305D	0.174	0.183	0.291	0.348	0.366	0.436							0.174	0.183	0.436						
PBPB 305E	0.522	0.5	0.435	0.348	0.366	0.580	0.174	0.183	0.435				0.348	0.366	0.145						
PBPB 305F	0.362	0.351	0.672	0.542	0.527	0.504	0.181	0.176	0.336				0.181	0.176	0.336						
SCFT 101																			0	0	0

8.3 MULTIMETRIC RESULTS FROM STREAM RESTORATION SITES

The development of Index of Biotic Integrity (IBI) using multimetric analysis was described in Section 1.2.1. With the biomonitoring data, these restoration sites can also be examined with the existing benthic IBI for Maryland. The statewide benthic IBI was first developed by Stribling et al. (1998). Using MBSS data (1,098 sites) from 1994 to 1997, they divided Maryland into Coastal and non-Coastal regions and developed the benthic IBI for each region. There were seven and nine genus-level index metrics in the Coastal Plain and Non-Coastal Plain regions, respectively. The problem however was that the number of stream site locations within this database that met reference criteria was fairly small (37) and was not distributed evenly across Maryland's physiographic regions. The implications of a small and clustered distribution of reference sites are that benthic IBI development may not capture all of the variability inherent in the state. A better and more current benthic macroinvertebrate IBI for Maryland was developed by Southerland et al. (2005). With a total of 2508 sites and 196 reference sites with data from 1994 to 2004, they developed new benthic IBI in three regions, Coastal Plain, Piedmont, and Combined Highlands. There are seven, six, and eight metrics for each region, respectively. Our target sites, the stream restoration sites in Montgomery County, are located in the Piedmont region. This new benthic IBI for the Piedmont region is summarized in Table 47. These six metrics have different scoring thresholds that were established as approximately the 10th and 50th (median) percentile values for reference sites. Note that the new benthic macroinvertebrate IBI is intended for use with subsample size ranging from 80 to 120 individuals. The benthic IBI with subsample size outside this

range should be used with caution.

Table 47. The benthic metrics for the Piedmont region in Maryland. (Adapted from Southerland et al. 2005) Note that the new benthic IBI is intended for use with subsample sizes ranging from 80 to 120 individuals.

Benthic IBIs (metrics)	Description	Thresholds		
		5	3	1
Piedmont				
Number of Taxa	Measures the overall variety of the macroinvertebrate assemblage	≥ 25	15 – 24	< 15
Number of EPT	Number of taxa in the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies)	≥ 11	5 – 10	< 5
Number of Ephemeroptera	Number of mayfly taxa	≥ 4	2 – 3	< 2
% Intolerant Urban	Percent of sample considered intolerant to urbanization (tolerance values 0-3)	≥ 51	12 – 50	< 12
% Chironomidae	Percent midge larvae and pupae	≤ 4.6	4.7 – 63	> 63
% Clingers	Percent of sample primarily adapted for inhabiting flowing water, as in riffles	≥ 74	31 – 73	< 31

The number of taxa metric gives us an idea of the macroinvertebrate richness in the stream monitoring site. Taxon richness generally decreases as a stream ecosystem degrades. The insect order Ephemeroptera, Plecoptera, and Trichoptera (EPT) are recognized as taxa that are sensitive to stream condition. Some EPT taxa are less sensitive to pollutants or disturbance, so low taxa counts are usually represented by more tolerant taxa. The richness of Ephemeroptera indicates the ability of a stream to support this generally intolerant insect order. Number of EPT and number of Ephemeroptera thus tell us the health condition of a stream site. According to the master taxon list by Southerland et al. (2005), most taxa at the genus level are

given a tolerant value scaled from zero to ten. Zero means extremely intolerant to perturbation and urbanization. Ten means very tolerant to perturbation and urbanization. Taxa with tolerance values from 0-3 are those species intolerant to urbanization. The percentage of intolerance to urbanization shows what percentage of the subsample is composed of intolerant species. Chironomidae family taxa generally have cosmopolitan distributions and may occur even in highly-polluted streams. The percentage of Chironomidae is expected to increase as environmental stressors increase. The taxa which cling to surfaces in fast moving water by means of morphological adaptations or construction of fixed retreats increase in abundance in the absence of environmental stressors. The stressors which most adversely affect this metric are those that directly disturb or produce loss of high quality habitat, such as clean gravel riffles. The percentage of Clingers shows the habitat composition in the stream by habitat preference of benthic taxa.

According to the definitions in Table 47, the benthic IBI metrics for each stream restoration monitoring site were calculated using the biomonitoring data and determined the IBI scores using their designed threshold. The results of benthic IBI metrics at each site are listed in Table 49 along with their IBI at the time the data were collected. The total individual numbers of each subsample are listed along with the date that the data were taken. Those subsamples that are not in the defined IBI subsample range (80 to 120 individuals) are marked with an asterisk (*). The calculation of the benthic IBI metrics, however, still uses the respective total individual number from the subsample as the denominator. The interpretation of these benthic IBI should be used with caution. At the summary results in Table 49, the

benthic metrics and IBI scores show no improvement or even degradation in the restoration sites that have been monitored. Strictly speaking, only stream site NWLT 101 shows improvement in the benthic metric at the latest monitoring. The benthic IBI metrics at site NWLT 101 still remain at 1 and 3, showing limit improvement in habitat conditions. Some sites, NWNW407B, NWNW407B, PBGH108, PBPB 305D and PBPB305F, show decreasing IBI metrics during the monitoring period and go back up at the latest data collection. The benthic metrics at site PBPB 305C through time show degrading stream conditions. The benthic metrics values at other monitoring sites either keep the same IBI values or fluctuate during the monitoring period. From a multimetric point of view, most of these restoration projects did not improve habitat for macroinvertebrates.

Table 48. IBI score ranges and corresponding narrative ratings. (From Stribling et al. 1998)

IBI Score Range	Narrative Rating
4.0 - 5.0	Good
3.0 - 3.9	Fair
2.0 - 2.9	Poor
1.0 - 1.9	Very Poor

The benthic IBI scores were adjusted to a common scale ranging from 1 to 5 for statewide comparisons by Stribling et al (1998). The corresponding narrative rating is listed in Table 48. The scores among the benthic IBI were averaged, and the final benthic IBI score was obtained for each stream site. The results are summarized in Table 50. The results from the final benthic IBI scores agree with the results in Table 49. Only at site NWLT 101 did the benthic IBI score increase from 1 to 1.33,

but it is still considered very poor according to the narrative rating for Maryland. The benthic IBI scores at PBPB 305B, PBPB 305C decreased from 3 (3.67) to 1 (1.33) which lowered their ratings from fair to very poor.

Table 49. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individual not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
LFLF301B	2001			4/26/2002	* Total: 130					4/12/2005	* Total : 29				
	Number of Taxa			9	1					8	1				
	Number of EPT			1	1					1	1				
	Number of Ephemeroptera			0	1					0	1				
	% Intolerant Urban			10	1					3.45	1				
	% Chironomidae			70.77	1					68.97	1				
	% Clingers			14.62	1					10.34	1				
LFLF301C		4/26/2002	* Total : 141												
	Number of Taxa	8	1												
	Number of EPT	1	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	4.26	1												
	% Chironomidae	75.89	1												
	% Clingers	17.02	1												
LRCR101A		3/21/2002	Total : 111												
	Number of Taxa	9	1												
	Number of EPT	0	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	0.90	1												
	% Chironomidae	58.56	3												
	% Clingers	0	1												
LRCR101B		3/22/2002	* Total :143												
	Number of Taxa	10	1												
	Number of EPT	1	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	0	1												
	% Chironomidae	90.91	1												
	% Clingers	0.70	1												

Table 49-continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
LRJB203A	Jan-05	3/28/2002	* Total : 68	4/26/2005	* Total : 131	4/5/2007	* Total : 128								
	Number of Taxa	8	1	10	1	11	1								
	Number of EPT	2	1	0	1	2	1								
	Number of Ephemeroptera	0	1	0	1	0	1								
	% Intolerant Urban	1.47	1	0	1	0	1								
	% Chironomidae	80.88	1	90.84	1	75.78	1								
	% Clingers	10.29	1	1.53	1	13.28	1								
LRJB203B	Jan-05	3/28/2002	Total : 112	4/26/2005	* Total : 42	4/5/2007	* Total : 35								
	Number of Taxa	9	1	6	1	9	1								
	Number of EPT	2	1	0	1	1	1								
	Number of Ephemeroptera	0	1	0	1	0	1								
	% Intolerant Urban	2.68	1	0	1	0.00	1								
	% Chironomidae	82.14	1	83.33	1	65.71	1								
	% Clingers	10.71	1	0	1	5.71	1								
LRJB204	Jan-05	3/24/2003	* Total : 13	4/26/2005	* Total : 57	4/18/2007	* Total : 124								
	Number of Taxa	7	1	5	1	9	1								
	Number of EPT	2	1	0	1	1	1								
	Number of Ephemeroptera	0	1	0	1	0	1								
	% Intolerant Urban	0	1	0	1	0	1								
	% Chironomidae	53.85	3	94.74	1	83.87	1								
	% Clingers	30.77	1	1.75	1	12.10	1								
LRSB101A	2004	3/25/2003	* Total : 39	4/25/2005	* Total : 65			4/5/2007	Total : 104						
	Number of Taxa	7	1	9	1			9	1						
	Number of EPT	0	1	0	1			0	1						
	Number of Ephemeroptera	0	1	0	1			0	1						
	% Intolerant Urban	10.26	1	0	1			0	1						
	% Chironomidae	0	5	83.08	1			14.42	3						
	% Clingers	2.56	1	4.62	1			0.96	1						

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
LRSB101C	2004	4/4/2002	* Total : 50	4/25/2005	* Total : 10			4/5/2007	* Total : 77						
	Number of Taxa	7	1	5	1			13	1						
	Number of EPT	0	1	0	1			0	1						
	Number of Ephemeroptera	0	1	0	1			0	1						
	% Intolerant Urban	2	1	0	1			0	1						
	% Chironomidae	38	3	80	1			55.84	3						
	% Clingers	8	1	0	1			5.19	1						
LRTB101		4/17/2002	Total : 105							4/21/2006	Total : 120				
	Number of Taxa	9	1							9	1				
	Number of EPT	2	1							1	1				
	Number of Ephemeroptera	0	1							0	1				
	% Intolerant Urban	2.86	1							0	1				
	% Chironomidae	83.81	1							80.83	1				
	% Clingers	3.81	1							3.33	1				
LRTB202		4/18/2002	* Total : 25												
	Number of Taxa	7	1												
	Number of EPT	1	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	0	1												
	% Chironomidae	40	3												
	% Clingers	4	1												
		4/26/2006	* Total : 72												
	Number of Taxa	8	1												
	Number of EPT	0	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	0	1												
	% Chironomidae	30.55555556	3												
	% Clingers	1.388888889	1												

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
LRTB203A		4/17/2002	* Total : 68												
	Number of Taxa	8	1												
	Number of EPT	0	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	1.47	1												
	% Chironomidae	45.59	3												
	% Clingers	1.47	1												
		3/28/2006	Total : 115												
	Number of Taxa	9	1												
	Number of EPT	3	1												
	Number of Ephemeroptera	1	1												
	% Intolerant Urban	0	1												
	% Chironomidae	45.22	3												
	% Clingers	4.35	1												
LRTB203B		4/19/2002	* Total : 141												
	Number of Taxa	13	1												
	Number of EPT	3	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	0	1												
	% Chironomidae	83.69	1												
	% Clingers	9.93	1												
		3/28/2006	* Total : 76												
	Number of Taxa	11	1												
	Number of EPT	1	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	0	1												
	% Chironomidae	64.47	1												
	% Clingers	1.32	1												

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
NWGT101		4/12/2005	* Total : 21												
	Number of Taxa	7	1												
	Number of EPT	1	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	0	1												
	% Chironomidae	76.19	1												
	% Clingers	9.52	1												
NWLR102	2002	3/28/2001	* Total : 313	4/19/2002	Total : 112			3/30/2004	* Total : 126	4/8/2007	Total : 95				
	Number of Taxa	12	1	10	1			12	1	8	1				
	Number of EPT	2	1	1	1			2	1	1	1				
	Number of Ephemeroptera	0	1	0	1			0	1	0	1				
	% Intolerant Urban	0	1	0.89	1			2.38	1	0	1				
	% Chironomidae	27.16	3	92.86	1			76.19	1	78.95	1				
	% Clingers	5.43	1	0.89	1			9.52	1	4.21	1				
NWLT101	2002	5/3/2001	* Total : 208	4/19/2002	* Total : 181			3/30/2004	* Total : 146	4/18/2007	Total : 115				
	Number of Taxa	6	1	8	1			8	1	13	1				
	Number of EPT	3	1	0	1			1	1	2	1				
	Number of Ephemeroptera	0	1	0	1			0	1	0	1				
	% Intolerant Urban	2.88	1	0	1			0	1	0	1				
	% Chironomidae	94.71	1	98.90	1			96.58	1	44.35	3				
	% Clingers	4.81	1	0.55	1			1.37	1	29.57	1				
NWNW301		4/14/2001	Total : 109												
	Number of Taxa	23	3												
	Number of EPT	12	5												
	Number of Ephemeroptera	6	5												
	% Intolerant Urban	41.28	3												
	% Chironomidae	25.69	3												
	% Clingers	65.14	3												

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
		3/29/2002	Total : 108												
	Number of Taxa	23	3												
	Number of EPT	10	3												
	Number of Ephemeroptera	8	5												
	% Intolerant Urban	40.74	3												
	% Chironomidae	46.30	3												
	% Clingers	48.15	3												
		4/14/2004	Total : 105												
	Number of Taxa	20	3												
	Number of EPT	6	3												
	Number of Ephemeroptera	2	3												
	% Intolerant Urban	4.76	1												
	% Chironomidae	77.14	1												
	% Clingers	13.33	1												
NWNW304		5/7/1997	* Total : 277												
	Number of Taxa	22	3												
	Number of EPT	13	5												
	Number of Ephemeroptera	6	5												
	% Intolerant Urban	27.44	3												
	% Chironomidae	45.49	3												
	% Clingers	50.54	3												
		4/5/2001	* Total : 164												
	Number of Taxa	22	3												
	Number of EPT	9	3												
	Number of Ephemeroptera	5	5												
	% Intolerant Urban	33.54	3												
	% Chironomidae	37.20	3												
	% Clingers	50.61	3												

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
		4/14/2004	Total : 106												
	Number of Taxa	13	1												
	Number of EPT	3	1												
	Number of Ephemeroptera	1	1												
	% Intolerant Urban	1.89	1												
	% Chironomidae	72.64	1												
	% Clingers	24.53	1												
NWNW407A	2002	4/8/2002	Total : 114	4/29/2003	Total : 112			4/6/2005	* Total : 65			4/18/2007	Total : 90		
	Number of Taxa	23	3	11	1			10	1			13	1		
	Number of EPT	7	3	2	1			3	1			2	1		
	Number of Ephemeroptera	4	5	1	1			1	1			0	1		
	% Intolerant Urban	3.51	1	2.68	1			0	1			0	1		
	% Chironomidae	45.61	3	91.07	1			83.08	1			36.67	3		
	% Clingers	39.47	3	5.36	1			12.31	1			3.33	1		
NWNW407B	2002	4/8/2002	Total : 112	4/29/2003	Total : 101			4/6/2005	* Total : 21			4/18/2007	Total : 99		
	Number of Taxa	19	3	19	3			10	1			16	3		
	Number of EPT	6	3	5	3			2	1			5	3		
	Number of Ephemeroptera	5	5	4	5			2	3			4	5		
	% Intolerant Urban	17.86	3	9.90	1			4.76	1			9.09	1		
	% Chironomidae	60.71	3	62.38	3			57.14	3			67.68	1		
	% Clingers	17.86	1	9.90	1			14.29	1			12.12	1		
NWNW407C	2002	4/8/2002	Total : 111	4/29/2003	Total : 114			4/6/2005	Total : 120			4/18/2007	* Total : 128		
	Number of Taxa	22	3	21	3			19	3			20	3		
	Number of EPT	4	1	4	1			8	3			4	1		
	Number of Ephemeroptera	3	3	4	5			3	3			2	3		
	% Intolerant Urban	9.91	1	5.26	1			16.67	3			5.47	1		
	% Chironomidae	55.86	3	66.67	1			36.67	3			61.72	3		
	% Clingers	10.81	1	18.42	1			38	3			20.31	1		

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
NWNW407D	2002	4/8/2002	Total : 111	4/20/2003	Total : 95			4/6/2005	* Total : 49			4/11/2007	Total : 114		
	Number of Taxa	23	3	18	3			8	1			17	3		
	Number of EPT	5	3	2	1			2	1			3	1		
	Number of Ephemeroptera	3	3	2	3			1	1			2	3		
	% Intolerant Urban	11.71	1	13.68	3			2.04	1			3.51	1		
	% Chironomidae	47.75	3	53.68	3			79.59	1			48.25	3		
	% Clingers	12.61	1	12.63	1			16.33	1			12.28	1		
NWNW408A	2002	4/8/2002	Total : 106	4/29/2003	Total : 105			4/6/2005	* Total : 39			4/11/2007	* Total : 143		
	Number of Taxa	21	3	19	3			10	1			11	1		
	Number of EPT	6	3	4	1			2	1			1	1		
	Number of Ephemeroptera	5	5	2	3			0	1			1	1		
	% Intolerant Urban	6.60	1	6.67	1			2.56	1			0	1		
	% Chironomidae	58.49	3	70.48	1			71.79	1			51.75	3		
	% Clingers	22.64	1	9.52	1			17.95	1			8.39	1		
NWNW408B	2002	4/8/2002	Total : 110	4/29/2003	* Total : 52			4/6/2005	* Total : 69			4/11/2007	Total : 101		
	Number of Taxa	17	3	14	1			13	1			15	3		
	Number of EPT	5	3	5	3			3	1			4	1		
	Number of Ephemeroptera	5	5	2	3			3	3			2	3		
	% Intolerant Urban	8.18	1	3.85	1			15.94	3			0.99	1		
	% Chironomidae	75.45	1	53.85	3			43.48	3			56.44	3		
	% Clingers	16.36	1	19.23	1			18.84	1			14.85	1		
NWPT101		4/12/2005	* Total : 77												
	Number of Taxa	21	3												
	Number of EPT	8	3												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	27.27	3												
	% Chironomidae	22.08	3												
	% Clingers	54.55	3												

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
PBGH108	2001	3/24/2000	* Total : 393	4/22/2002	Total : 119			4/20/2004	* Total : 166	4/29/2005	Total : 120				
	Number of Taxa	20	3	23	3			18	3	19	3				
	Number of EPT	6	3	5	3			5	3	6	3				
	Number of Ephemeroptera	3	3	2	3			0	1	0	1				
	% Intolerant Urban	50.89	3	50.42	3			7.83	1	30.83	3				
	% Chironomidae	36.90	3	36.97	3			66.87	1	32.50	3				
	% Clingers	58.02	3	53.78	3			27.71	1	52.50	3				
PBGS102B		4/5/1999	* Total : 227												
	Number of Taxa	18	3												
	Number of EPT	8	3												
	Number of Ephemeroptera	2	3												
	% Intolerant Urban	72.69	5												
	% Chironomidae	17.18	3												
	% Clingers	81.50	5												
		4/13/2004	Total : 92												
	Number of Taxa	16	3												
	Number of EPT	2	1												
	Number of Ephemeroptera	0	1												
	% Intolerant Urban	10.87	1												
	% Chironomidae	63.04	1												
	% Clingers	26.09	1												
PBGS111		4/5/1999	* Total : 386												
	Number of Taxa	15	3												
	Number of EPT	9	3												
	Number of Ephemeroptera	3	3												
	% Intolerant Urban	76.17	5												
	% Chironomidae	11.14	3												
	% Clingers	88.34	5												

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
		3/27/2000	* Total : 267												
	Number of Taxa	18	3												
	Number of EPT	10	3												
	Number of Ephemeroptera	3	3												
	% Intolerant Urban	45.32	3												
	% Chironomidae	23.97	3												
	% Clingers	74.53	5												
		4/2/2001	* Total : 185												
	Number of Taxa	18	3												
	Number of EPT	9	3												
	Number of Ephemeroptera	3	3												
	% Intolerant Urban	69.19	5												
	% Chironomidae	10.27	3												
	% Clingers	86.49	5												
		4/22/2002	* Total : 154												
	Number of Taxa	17	3												
	Number of EPT	9	3												
	Number of Ephemeroptera	3	3												
	% Intolerant Urban	64.94	5												
	% Chironomidae	13.64	3												
	% Clingers	83.77	5												
		3/26/2003	Total : 103												
	Number of Taxa	15	3												
	Number of EPT	9	3												
	Number of Ephemeroptera	5	5												
	% Intolerant Urban	73.79	5												
	% Chironomidae	5.83	3												
	% Clingers	90.29	5												
		4/13/2004	* Total : 124												
	Number of Taxa	11	1												
	Number of EPT	5	3												
	Number of Ephemeroptera	1	1												
	% Intolerant Urban	34.68	3												
	% Chironomidae	45.97	3												
	% Clingers	50.81	3												

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
		4/11/2005	Total : 107												
	Number of Taxa	22	3												
	Number of EPT	9	3												
	Number of Ephemeroptera	3	3												
	% Intolerant Urban	38.32	3												
	% Chironomidae	28.97	3												
	% Clingers	62.62	3												
PBPB305A	2001	4/26/2001	Total : 118	4/23/2002	* Total : 130	4/30/2003	* Total : 36			4/6/2005	Total : 89			4/10/2007	Total : 115
	Number of Taxa	32	5	27	5	14	1			27	5			18	3
	Number of EPT	10	3	9	3	5	3			7	3			6	3
	Number of Ephemeroptera	5	5	4	5	2	3			4	5			2	3
	% Intolerant Urban	27.12	3	6.92	1	16.67	3			2.25	1			4.35	1
	% Chironomidae	38.14	3	64.62	1	36.11	3			30.34	3			70.43	1
	% Clingers	53.39	3	30.00	1	47.22	3			51.69	3			20.87	1
PBPB305B	2001	4/26/2001	* Total : 173	4/23/2002	* Total : 122	4/30/2003	Total : 97			4/6/2005	* Total : 139			4/10/2007	Total : 111
	Number of Taxa	24	3	30	5	22	3			28	5			10	1
	Number of EPT	6	3	7	3	6	3			6	3			2	1
	Number of Ephemeroptera	3	3	4	5	3	3			3	3			0	1
	% Intolerant Urban	23.70	3	15.57	3	21.65	3			6.47	1			0.00	1
	% Chironomidae	40.46	3	30.33	3	43.30	3			32.37	3			70.27	1
	% Clingers	52.60	3	54.10	3	49.48	3			48.20	3			27.93	1
PBPB305C	2001	4/3/2001	*Total : 137	4/29/2002	* Total : 195	4/14/2003	* Total : 154	4/16/2004	* Total : 354	4/6/2005	* Total : 44			4/10/2007	Total : 101
	Number of Taxa	25	5	27	5	25	5	21	3	12	1			15	3
	Number of EPT	7	3	8	3	7	3	7	3	4	1			3	1
	Number of Ephemeroptera	3	3	5	5	1	1	2	3	2	3			1	1
	% Intolerant Urban	23.36	3	12.31	3	9.74	1	4.80	1	2.27	1			0.99009901	1
	% Chironomidae	20.44	3	54.87	3	54.55	3	67.80	1	59.09	3			71.28712871	1
	% Clingers	78.83	5	36.41	3	37.01	3	30.23	1	36.36	3			20.79207921	1

Table 49 - continued. The summary table of multimetrics at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction		1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
		Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI	Metrics values	IBI
PBPB305D	2001	4/26/2001	Total : 103	4/23/2002	Total : 110					4/6/2005	* Total : 53				
	Number of Taxa	16	3	15	3					17	3				
	Number of EPT	6	3	5	3					5	3				
	Number of Ephemeroptera	3	3	3	3					2	3				
	% Intolerant Urban	17.48	3	2.73	1					5.66	1				
	% Chironomidae	44.66	3	85.45	1					45.28	3				
	% Clingers	49.51	3	9.09	1					35.85	3				
PBPB305E	2001	4/26/2001	* Total : 128	4/23/2002	* Total : 180	4/30/2003	Total : 96			4/6/2005	* Total : 13				
	Number of Taxa	20	3	25	5	20	3			5	1				
	Number of EPT	7	3	6	3	4	1			2	1				
	Number of Ephemeroptera	3	3	3	3	3	3			0	1				
	% Intolerant Urban	21.88	3	3.89	1	5.21	1			15.38	3				
	% Chironomidae	38.28	3	72.78	1	60.42	3			38.46	3				
	% Clingers	59.38	3	25.00	1	29.17	1			61.54	3				
PBPB305F	2001	4/26/2001	* Total : 126	4/23/2002	* Total : 153	4/30/2003	* Total : 67			4/6/2005	Total : 86				
	Number of Taxa	23	3	18	3	20	3			22	3				
	Number of EPT	9	3	7	3	2	1			7	3				
	Number of Ephemeroptera	4	5	3	3	1	1			2	3				
	% Intolerant Urban	21.43	3	5.23	1	1.49	1			6.98	1				
	% Chironomidae	31.75	3	83.01	1	68.66	1			16.28	3				
	% Clingers	65.08	3	15.69	1	23.88	1			72.09	3				
SCFT101	1994													4/13/2004	Total : 117
	Number of Taxa													9	1
	Number of EPT													0	1
	Number of Ephemeroptera													0	1
	% Intolerant Urban													0	1
	% Chironomidae													38.46	3
	% Clingers													0.85	1

Table 50. The summary table of benthic IBI at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction	1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
			BIBI		BIBI		BIBI		BIBI		BIBI		BIBI	
LFLF301B	2001			4/26/2002	1*					4/12/2005	1*			
LFLF301C		4/26/2002	1*											
LRCR101A		3/21/2002	1.33											
LRCR101B		3/22/2002	1*											
LRJB203A	Jan-05	3/28/2002	1*	4/26/2005	1*	4/5/2007	1*							
LRJB203B	Jan-05	3/28/2002	1	4/26/2005	1*	4/5/2007	1*							
LRJB204	Jan-05	3/24/2003	1.33*	4/26/2005	1*	4/18/2007	1*							
LRSB101A	2004	3/25/2003	1.67*	4/25/2005	1*			4/5/2007	1.33					
LRSB101C	2004	4/4/2002	1.33*	4/25/2005	1*			4/5/2007	1.33*					
LRTB101		4/17/2002	1											
		4/21/2006	1											
LRTB202		4/18/2002	1.33*											
		4/26/2006	1.33*											
LRTB203A		4/17/2002	1.33*											
		3/28/2006	1.33											
LRTB203B		4/19/2002	1*											
		3/28/2006	1*											
NWGT101		4/12/2005	1*											
NWLR102	2002	3/28/2001	1.33*	4/19/2002	1			3/30/2004	1*	4/8/2007	1			
NWLT101	2002	5/3/2001	1*	4/19/2002	1*			3/30/2004	1*	4/18/2007	1.33			
NWNW301		4/14/2001	3.67											
		3/29/2002	3.33											
		4/14/2004	2											
NWNW304		5/7/1997	3.67*											
		4/5/2001	3.33*											
		4/14/2004	1											

Table 50-continued. The summary table of benthic IBI at stream restoration sites in Montgomery County. (*: total number of individuals not within the IBI defined range) (Please refer to Table 34 for Site ID.)

Station ID	Completed	Pre-construction	1 yr later		2 yr later		3 yr later		4 yr later		5 yr later		6 yr later	
			BIBI		BIBI		BIBI		BIBI		BIBI		BIBI	
NWNW407A	2002	4/8/2002	3	4/29/2003	1			4/6/2005	1*			4/18/2007	1.33	
NWNW407B	2002	4/8/2002	3	4/29/2003	2.67			4/6/2005	1.67*			4/18/2007	2.33	
NWNW407C	2002	4/8/2002	2	4/29/2003	2			4/6/2005	3			4/18/2007	2*	
NWNW407D	2002	4/8/2002	2.33	4/20/2003	2.33			4/6/2005	1*			4/11/2007	2	
NWNW408A	2002	4/8/2002	2.67	4/29/2003	1.67			4/6/2005	1*			4/11/2007	1.33*	
NWNW408B	2002	4/8/2002	2.33	4/29/2003	2*			4/6/2005	2*			4/11/2007	2	
NWPT101		4/12/2005	2.67*											
PBGH108	2001	3/24/2000	3*	4/22/2002	3			4/20/2004	1.67*	4/29/2005	2.67			
PBG102B		4/5/1999	3.67*											
		4/13/2004	1.33											
PBG111		4/5/1999	3.67*											
		3/27/2000	3.33*											
		4/2/2001	3.67*											
		4/22/2002	3.67*											
		3/26/2003	4											
		4/13/2004	2.33*											
	4/11/2005	3												
PBPB305A	2001	4/26/2001	3.67	4/23/2002	2.67*	4/30/2003	2.67*			4/6/2005	3.33		4/10/2007	2
PBPB305B	2001	4/26/2001	3*	4/23/2002	3.67*	4/30/2003	3			4/6/2005	3*		4/10/2007	1
PBPB305C	2001	4/3/2001	3.67*	4/29/2002	3.67*	4/14/2003	2.67*	4/16/2004	2*	4/6/2005	2*		4/10/2007	1.33
PBPB305D	2001	4/26/2001	3	4/23/2002	2					4/6/2005	2.67*			
PBPB305E	2001	4/26/2001	3*	4/23/2002	2.33*	4/30/2003	2*			4/6/2005	2*			
PBPB305F	2001	4/26/2001	3.33*	4/23/2002	2*	4/30/2003	1.33*			4/6/2005	2.67			
SCFT101	1994												4/13/2004	1.33

CHAPTER 9: DISCUSSION AND CONCLUSIONS

The stream restoration sites in Montgomery County provided biomonitoring data to apply to multivariate and multimetric analysis. The macroinvertebrate predictive models using multivariate analysis can be treated as family level or genus level models. The benthic IBI scores from multimetric analysis, on the other hand, don't discriminate among taxon levels but depends on the composition of different taxa. The differences between these two analyses as applied to stream restoration sites are discussed below.

9.1 DISCUSSION

9.1.1 The Multivariate Analysis

When using Region-style models in GISHydro2000 to predict macroinvertebrate presence probabilities at stream restoration sites in Montgomery County, some of the models select not only the reference sites within the Piedmont but also sites in the Coastal Plain and Ridge and Valley according the physical similarity. It is interesting to find that not all the reference sites within Montgomery County were picked as references for the stream restoration sites. This indicates that the sites in close proximity are not necessarily appropriate reference sites to predict the presence of macroinvertebrates. From the biological point of view, those reference sites that are not nearby might give us information about good stream habitat for benthic macroinvertebrates. In additions, the attribute table of "Reference Site" theme in ArcView, as described in Section "Manual for using Region-Style

Models in GISHydro2000”, listed not only the site’s name and location, but also the physical parameters in each sites. The used reference sites were also highlighted in the attribute table, after running the Region-style model in GISHydro2000. These highlighted reference site in the table provide the target site the reference physical characteristic information other than biodiversity prediction. Other information of these selected reference sites, such as temperature, pH value, dissolved organic carbon, dissolved oxygen, etc, can be obtained by looking up the stream site ID in the MBSS data set, <http://mddnr.chesapeakebay.net/mbss/search.cfm>. The environmental stressors at the target site may be identified by comparing the difference between these reference sites and the target site.

In Chapter Eight, the taxon presence probabilities were predicted at the family level and at the genus level for each stream restoration site. The results have similar trends but are not the same in values. The family level results show some good stream conditions at some restoration sites, while genus level results generally show that streams are in poor conditions. The sites showing improvements at the family level do not show the same improvements at the genus level. But generally the sites that have higher O/E scores at the family level have higher O/E scores at the genus level as well. Some restored sites retained earlier O/E scores for the family level but dropped a lot in genus level. Note that some taxon individuals in the biomonitoring data couldn’t be identified to the genus level but only to the family level or higher. Those individuals were ignored in the observed data, which may cause inaccuracies in the O/E scores at the genus level. The O/E scores at the genus level may underestimate stream conditions, but more sites at the genus level show improvements in O/E scores

than at the family level, which indicate the improvement in stream habitat condition.

After consulting with Dr. William Lamp, the results show that the models predict more family taxa than genus taxa in the same stream site. This seems to be unreasonable. More than one genus taxa are included in one family taxon. There should be more genus taxa predicted than family taxa at the same site. But due to the mechanisms of the predictive model, only the probabilities higher than $P_c = 0.5$ were counted and included in the probability prediction. Thus, even though there are more genus taxa having occurrence probabilities larger than zero, only those with probabilities larger than 0.5 were counted. So, the predicted probabilities for genus taxa were lower than those for family taxa. The predicted probabilities also rely on the data quality from the reference sites. If the data quality of reference sites is improved, the predictive model precision in genus level will be improved.

9.1.2 The Multimetric Analysis

The benthic IBI metrics were also calculated using the biomonitoring data at restoration sites. The IBI metrics definitely give us more specific idea in those sensitive taxon composition, from which the stream conditions are determined. The IBI scores however only vary between 1, 3 and 5, depending on the threshold values, which may limit their ability to explain conditions. According to the rating for Maryland, the final IBI scores range from 1 to 5. And the scores from reference sites run from around 3 to 5. The final benthic IBI scores for restoration sites show no sites had a final score above 3 (Good or Fair condition) in the latest monitoring data. Sites PBGH 108, PBPB 305D, and PBPB 305F had the highest final IBI score at 2.67

(Poor condition). The rest of the restoration sites are still in poor or very poor condition according to the latest monitoring data. Note that the benthic IBI is designed for use with a total number of individuals between 80 and 120. Many of the observed taxon samples at restoration sites were not within this range. The benthic IBI metrics with percentages, such as the Percentage of Intolerant Urban, Percentage of Chironomidae, and Percentage of Clingers, will over- or under-estimate the benthic IBI scores. Besides, not every taxon in the master taxon list (Southerland et al., 2005) has a tolerance value. Those taxa without tolerance values were ignored in the metrics calculations. These undesirable factors may cause problems for IBI credibility.

9.1.3 The Comparison

The concepts of multimetric and multivariate analysis for biomonitoring are very different. Multimetric analysis pays more attention to information from sensitive taxon species, while multivariate analysis treats the information from each taxon equally. The stream conditions determined by the final benthic IBI scores at restoration sites are similar to, but not necessarily the same as, the conditions determined by O/E scores. Using site PBGS 111 as an example, the restoration project at this site was not yet completed as reflected in the available data. This is a pre-construction site. The O/E scores are illustrated in Figure 36 along with the final benthic IBI scores. The O/E scores from different Region-style models are shown according to the left vertical axis. The final benthic IBI scores are shown on the right vertical axis. The highest benthic IBI score, 4, of the entire Table 50 is at pre-construction monitoring year 2003 at this site. This benthic IBI score shows the stream site condition as good. The O/E score at the same time and the same site

however is 0.7 to 0.8 at the family level, which is high but not as close to 1, the preferred reference condition. The final benthic IBI scores at site PBGS 111 seem more easily to be affected by two out of six metrics that have lower scores. If looking into the raw observed taxon data for year 2003 monitoring, there were four fewer in Number of Taxa metrics and four fewer in Number of Ephemeroptera metrics than the previous year. The stream condition dropped from “Good” at year 2003 to “Poor” at year 2004 in the final benthic IBI. On the other hand, the O/E scores shows no difference between years 2003 and 2004. In other words, the same biomonitoring data set was interpreted differently by different analyses.

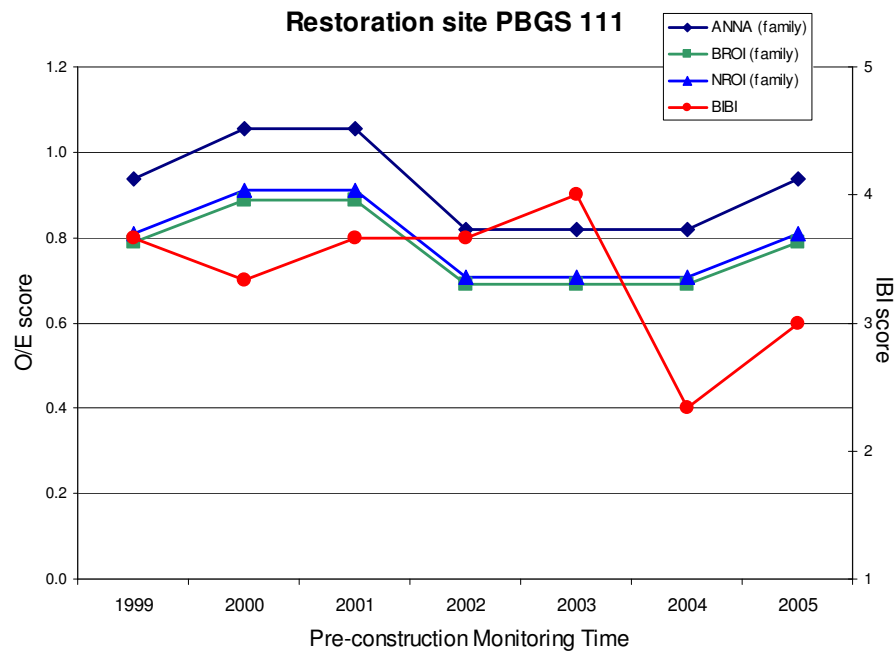
The comparison graph was also drawn for site PBPB 305C in Figure 37. This is the site with six years of post-restoration monitoring data. The benthic IBI scores decreased beginning at 1-year post-restoration and never increased again. The O/E scores in the family level remained the same for approximately two years, decreased during 3rd and 4th years, and then rose back up at the latest monitoring. The O/E scores in the genus level continually decreased since restoration. The latest benthic IBI scores are low because five out of six metrics have values of 1. The latest O/E score in the family level was, however, between 0.76 and 0.87, close to reference condition. The latest O/E score in the genus level was as low as 0.163 or 0. From Figure 33, it seems that predictive model at genus level captures similar trend at restoration site as benthic IBI scores. Since the predictive model checks individual taxon appearance, the O/E scores at family level may find the increased biodiversity in the stream site that benthic IBI score and model in genus level didn't capture. The benthic IBI scores seem to be rougher in scale than O/E scores. The metric scores

drop from 5 to 3 or 3 to 1 once the metric value pass the threshold value and final benthic IBI scores thus drop from fair to very poor. On the other hand, this quick change in IBI scores may give better warning of degrading stream conditions. The comparison graphs from the restoration sites other than PBGS 111 and PBPB 305C were also drawn if there were more than one year post-restoration data available. These graphs are in Appendix F for reference.

When the biomonitoring raw data were analyzed and summarized, regardless of whether it was benthic IBI from multimetrics analysis or O/E scores from multivariate analysis, much information was simplified and lost. In multimetrics analysis, final benthic IBI scores describe the stream condition as good, fair, poor, or very poor. But, only when individual metrics are studied, the sensitive species composition information is obtained for the stream conditions. In multivariate analysis, the taxon expected probability is predicted individually. The O/E values related how close the stream site values were to reference condition. But only when the presence of taxa are compared with the expected taxa, the missing or unexpected macroinvertebrates at the stream site provide information of potential environmental stressors. Final benthic IBI and O/E ratio offer integrated stream conditions, which reflect the water quality and stream site conditions in general. In order to improve the aquatic habitat for desired benthic species, the increased and decreased benthic metrics and the presence or absence of benthic species under different environmental stressors can provide quidence for appropriate remediation. I suggest to use the benthic metrics and presence or absence of benthic species as indicators of in-stream (or stream side) structural designs. Thus, the effectiveness of restorationprojects can

be verified.

(a) Family level model



(b) Genus level model

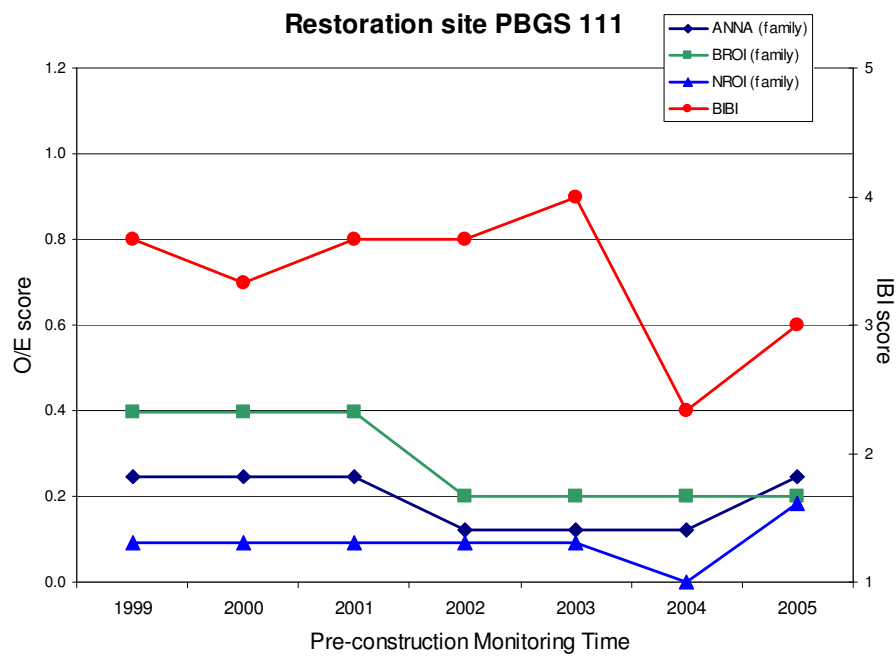
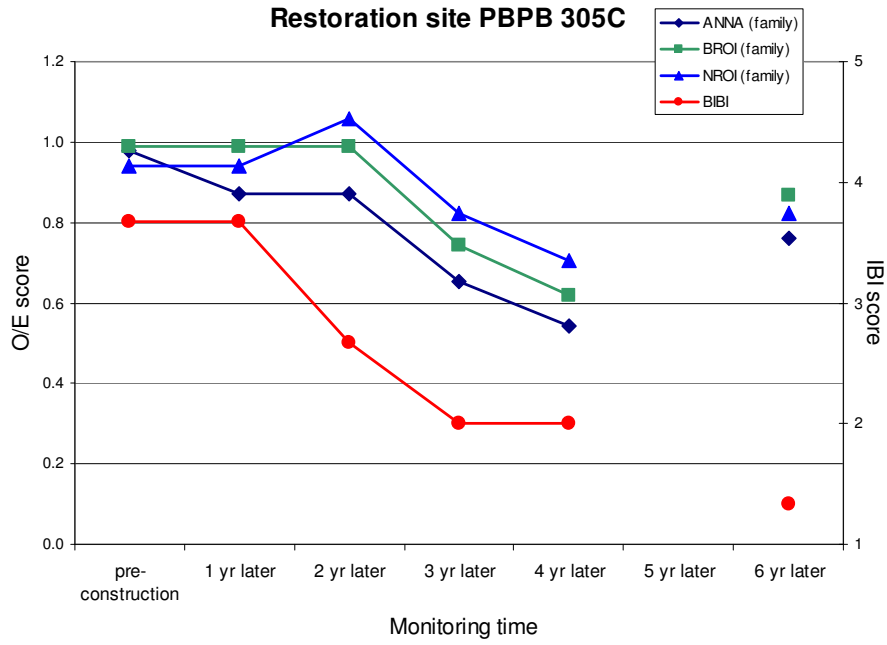


Figure 36. The stream condition trends at restoration site PBGS 111 captured by different analysis methods

(a) Family level model



(b) Genus level model

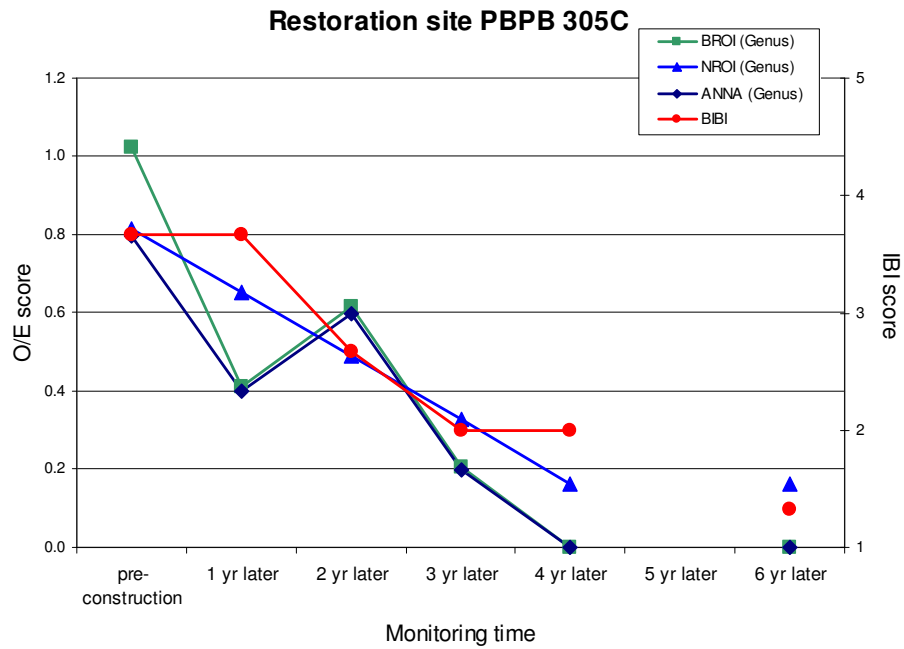


Figure 37. The stream condition trends at restoration site PBPB 305C captured by different analysis methods

9.2 CONCLUSIONS

The Region-style models perform better than previous RIVPACS-style models in Maryland. They offer alternative choices when building predictive models. In-stream physical information, especially substrate information describing the stream bed, can improve the benthic predictive models. The rational multiple regression function checking along with different parameter selection methods improved the parameter selection in the multiple regression for Region-style models. The rational multiple regression function checking is suggested for future use in building predictive models.

According to the evaluation results from both IBI and O/E scores, the stream restoration site conditions for macroinvertebrate populations in Montgomery County have improved only to a limited extent. Some of them showed major degradation through time. Some showed no improvement from previous conditions. The benthic scores at certain sites decreased after restoration construction for one to two years, which indicated that restoration practices might have damaged stream conditions at the beginning. There are up to five or six years of post-restoration monitoring data, but the stream benthic macroinvertebrates didn't seem to recover from previous perturbations. Even though the monitoring result on all sites did not go back to a good or better condition than pre-restoration in the available data, the O/E or benthic IBI scores of some restoration sites started to increase at the 3rd- or 4th- year post-restoration (Figure 33, and Appendix F). If the trend continues, the stream ecosystem in those sites may recover back to or even better than pre-restoration. The monitoring time in the available data are not long enough to show the recovery. I consulted many

entomologists regarding the recovery time for macroinvertebrate or the entire stream ecosystem. There is no consensus answer for the recovery time. Some research showed the macroinvertebrate recover to the “end-point” within three years after disturbance (i.e., pesticide contamination) (Niemi et al., 1990). The literature also suggested that longer recovery time is required if the disturbance resulted in physical alteration of existing habitats, residual pollution remained in the system, or the system was isolated and recolonization was suppressed (Niemi et al., 1990; Yount, 1990). The recovery process is highly affected by the life history of macroinvertebrates, the availabilities of unaffected stream areas, the spatial scale of disturbance, and the variety of stream flushing rate (Reice et al., 1990; Wallace, 1990; Yount, 1990). Research that has been done in Europe (Jähnig et al., 2008) shows no improvement in macroinvertebrate communities when the in-stream functional habitats has been improved after 10 years. So, the required recovery time for macroinvertebrate or stream ecosystem is not certain and varies.

The inconsistent analysis results from benthic IBI and O/E score shows the strength and weakness of these two evaluation methods. The benthic IBI metrics indicate the stream condition with sensitive species. When an environmental stressor level raises the alarm, benthic IBI metrics respond immediately. Appropriate stream restoration or protection should be immediately applied. However, these sensitive metrics react to natural variance as well. For example, number of Ephemeroptera (mayfly) is one of the benthic IBI metrics for Piedmont region in Maryland. Ephemeroptera are usually active in riffles. Riffles in the stream are significantly and rapidly affected by the stream water level. During a drought year, the number of

Ephemeroptera in the monitoring sample will drop, which does not necessarily mean the environmental stressors raise. In this situation, O/E values with the biodiversity information can provide valuable supplemental information to determine the stream conditions.

One thing that can't be explained in the data is that some of the restoration sites began with good or fair stream conditions (O/E between 0.8 and 0.9 and benthic IBI at 3.67). What were the reasons for initiating the restoration projects in the first place? And the assessment data show that stream conditions at these sites didn't become better or became worse. Maybe low impact restoration practice should be considered when stream restoration projects were initiated for other than biological reasons.

This study built the alternative macroinvertebrate predictive models for Maryland. It also shows that macroinvertebrate predictive models set up a standard for evaluating stream restoration effectiveness and provide supplement information other than using only multimetric analysis. The advantage of using Region-style models is that there is no need to find reference site for the target site and the models provide stream physical information for reference.

9.3 FUTURE RESEARCH

The Region-style models for Maryland have been built as an extension for GISHydro2000. The alternative soil data source from STATSGO is added to preference choices. Also, these predictive models can be adapted to any other region if the required biomonitoring and physical data are available. The programs for

building Region-style models have been written in R (Appendix B). The built models can be applied easily in any region with a biomonitoring data base.

I would like to obtain more restoration biomonitoring data to support the results of using my proposed system, especially this study shows that five or six year post-restoration monitoring may not be enough. Research in the literature has found that not enough post-restoration monitoring data are collected nationwide (Bernhardt et al., 2005), the existing information in Maryland can be a start and trigger more data collection. I want to compare not just taxon data differences in these restoration sites, but also evaluate their project rehabilitation results, for example hydraulic design and replanting designs. This will help us understand which designs benefit an ecosystem and which don't. That should lead to more systematic and constructive ideas about restoration methods. By talking to Daniel Harper, the manager of Water Restoration Program in Montgomery County DEP, these design plans should be available.

According to the personal consultation with Dr. Raymond Morgan, the multimetric analysis has its weakness as described in the Section 9.2. He used an example that in some stream the fish IBI scores is very low but the stream actually has no problem and is healthy. Applying to the case of macroinvertebrates, the predictive model that emphasizes on biodiversity composition can provide additional values to determine the stream health condition. The model has been built as an extension for GISHydro2000. This will be helpful when obtaining supplement information in other water quality research in Maryland.

From talking to Sean Smith from Department of Natural Resources, Maryland, some concerns about the predictive model were arised and some future

research ideas were suggested.

1. Different causes that contribute to the uncertainties in the predictive model :

The O/E ratio's variance was not studied, thus the reliability of the model was not certain. The required difference between two O/E scores that means really dissimilar stream conditions need to be studied.

Many natural uncertainties can affect the presence and absence of macroinvertebrates and thus the model prediction. The weather or climate variability, e.g., wet or dry year, will affect the occurrence of macroinvertebrates, and thus the predictive model. The different sampling date or time, e.g., before or after storm, varies the monitoring results. This various results will be added to the model uncertainties.

Different sampling efforts e.g., different protocols will add more uncertainty to the model. Errors in the sample cause the uncertainties in the model.

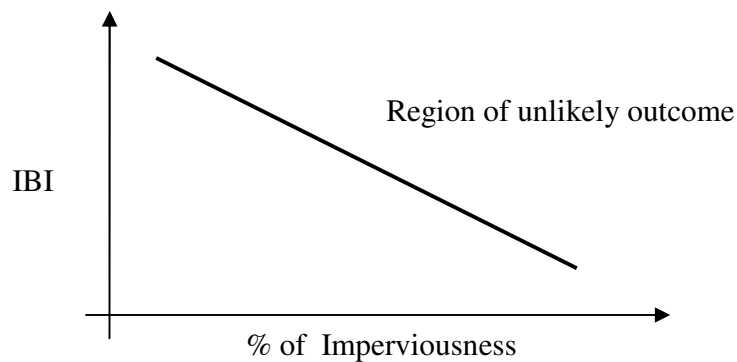
2. The possibility to include other important factors that affect the macroinvertebrates.

The urban development increases the percentage of imperviousness and cause the increasement of environmental stressors, e.g., temperature, flow variability, pollutions, etc. It seems to be odd to him to try to predict the benthic with only physical parameters. I explained to him that the predictive model is trying to statistically link the physical characteristics to the presence of benthic macroinvertebrate. It is not designed of to explain the relationship

between macroinvertebrates and physical characteristics. The specific physical feature only partially explain the presence of macroinvertebrates. Sean is concerned these factors overshadowed the physical characteristics to affect the occurrence of macroinvertebrates.

3. The limitation of stream restoration in urban region

Sean showed a graph that was proposed as below:



It means that when an area increasing urban development (imperviousness), there will be an impossible barrier for stream restoration to improve stream ecosystem (IBI scores) to a better condition. The higher the development is, the lower the limitation (IBI scores) is. Sean is concerned about the comparison is between the two different watershed conditions, the pristine stream and the high urban development.. It will never achieve the goal if the restoration project is in the urban streams to the reference condition. I think that even though it maybe impossible to restore the impaired urban streams to reference stream condition, it will be interesting to see if the benthic species composition has been moving toward the reference stream condition.

4. Potential comparison between the different imperviousness

Sean suggested to find imperviousness difference between the MBSS reference sites and the Montgomery County restoration sites. Thus the difference will be known while doing comparison. Also, he suggested looking into the imperviousness before and after stream restoration project. If the benthic monitoring data didn't improve after stream restoration project, it can be either

- a. the stream restoration practice didn't work
- b. the imperviousness of the watershed has increased after restoration.

The improvement done by the project can not balance out the effect of more imperviousness.

5. Evaluate the geomorphic design in stream restoration project

Most of the stream restoration has been emphasized on the channel lateral or longitudinal modification. (and some of vegetation planting). Sean suggested looking into if these modifications really improve the benthic population and diversity. Or which kinds of modification really improve biologically. It will be more interesting if it can be found that how it relates to the benthic condition.

6. The similarity between the the selection results from Region-style models and the Geology District Map

Sean is now in a group effort in developing a Geology District Map to guide the stream restoration. They first discriminate the geology district by using cluster analysis, and then this the district map will guide stream

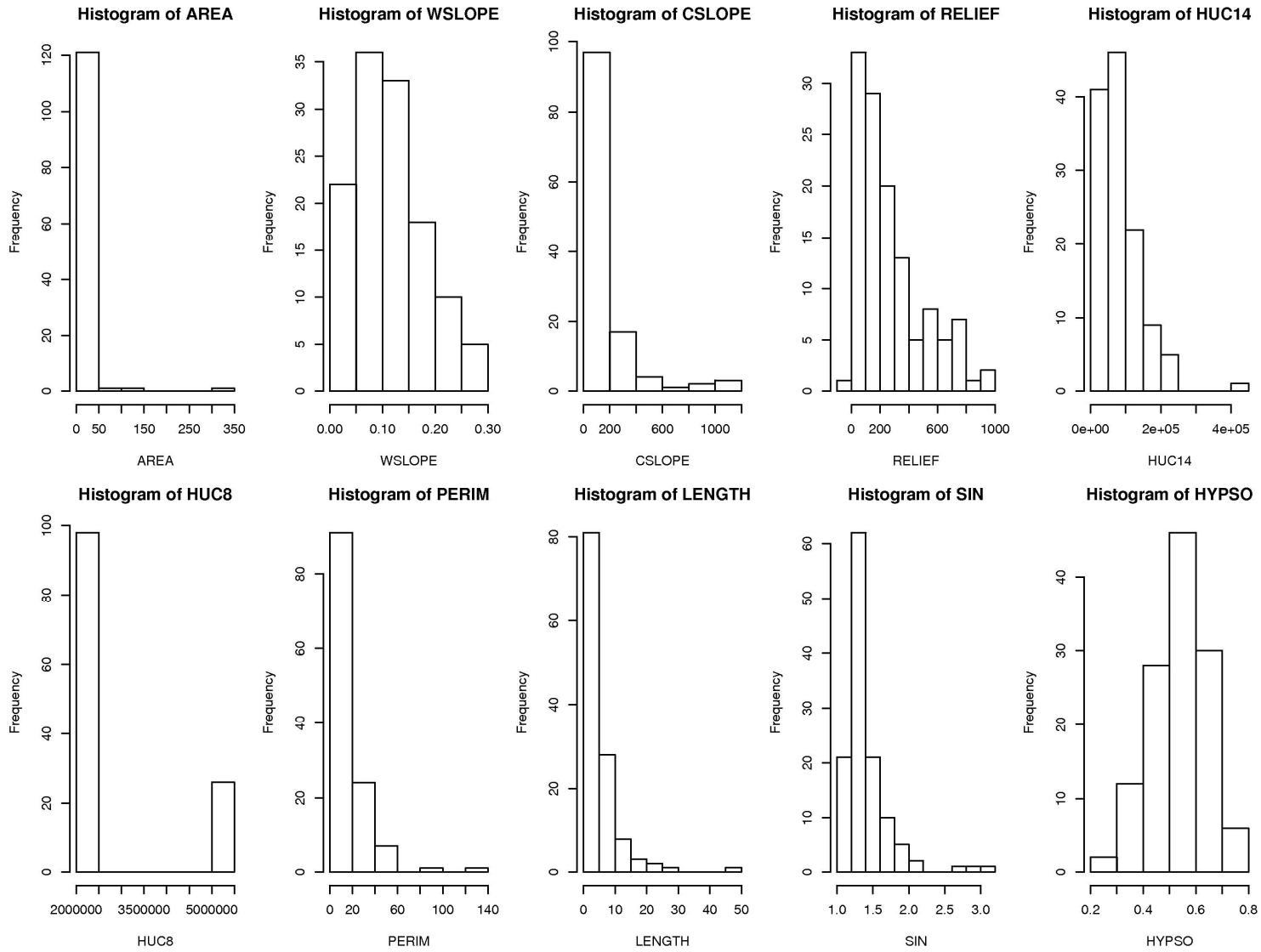
restoration practitioners to know more about the location and geology of the stream site. For example, if the stream site is in Coastal Plain, adding gravel to the stream bed to create habitat will be the effort in vain.

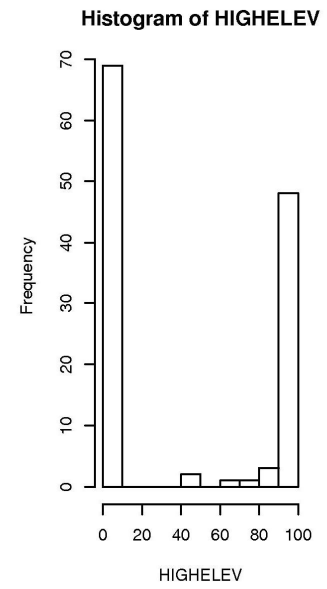
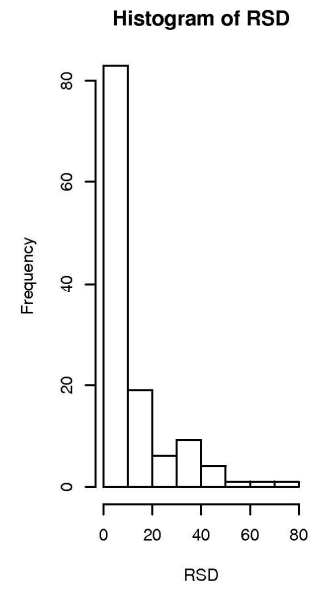
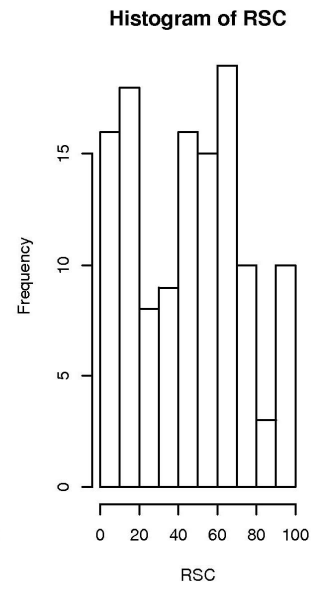
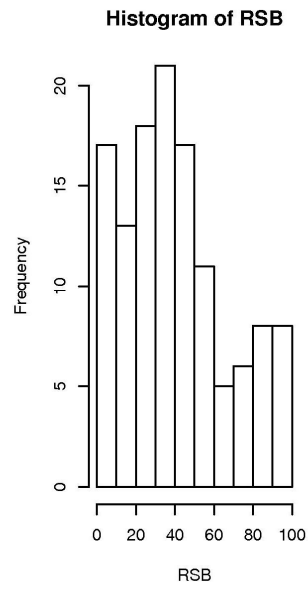
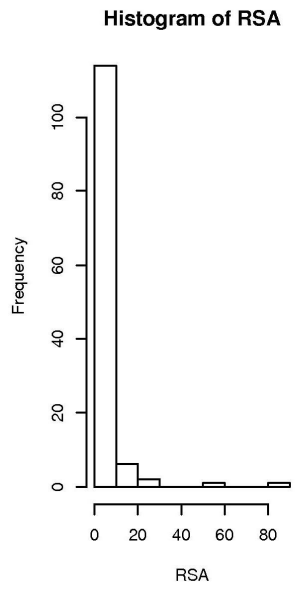
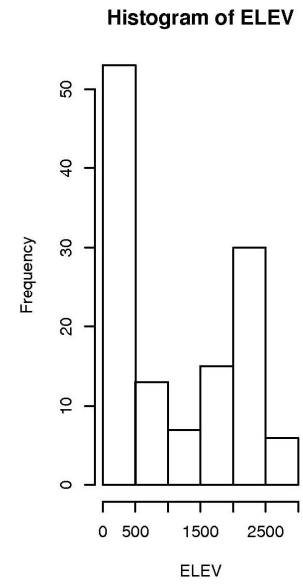
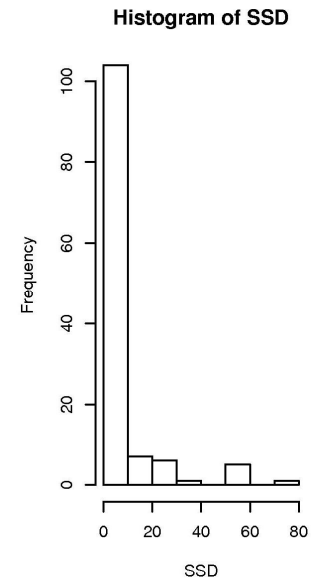
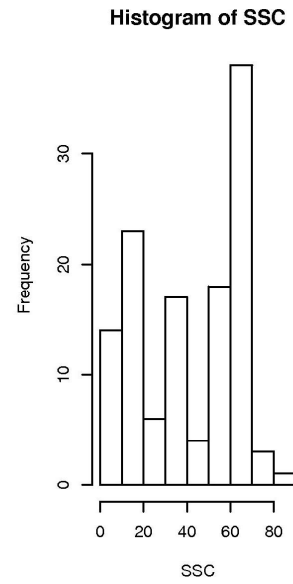
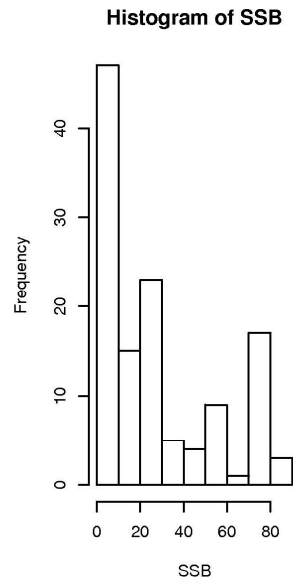
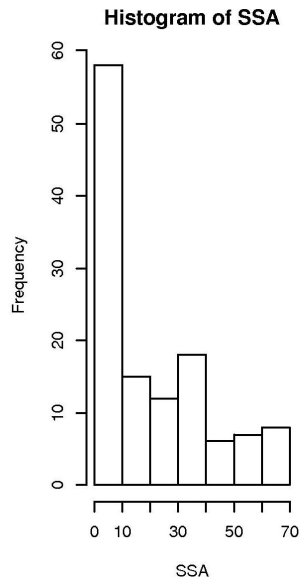
The predictive model link the mostly geologic characteristics to the macroinvertebrate composition. The Region-style models particularly pick out the similar physical characteristic reference sites to perform the prediction. It is not surprise to find the sites chosen by the Region-style models matched the Geology District Map.

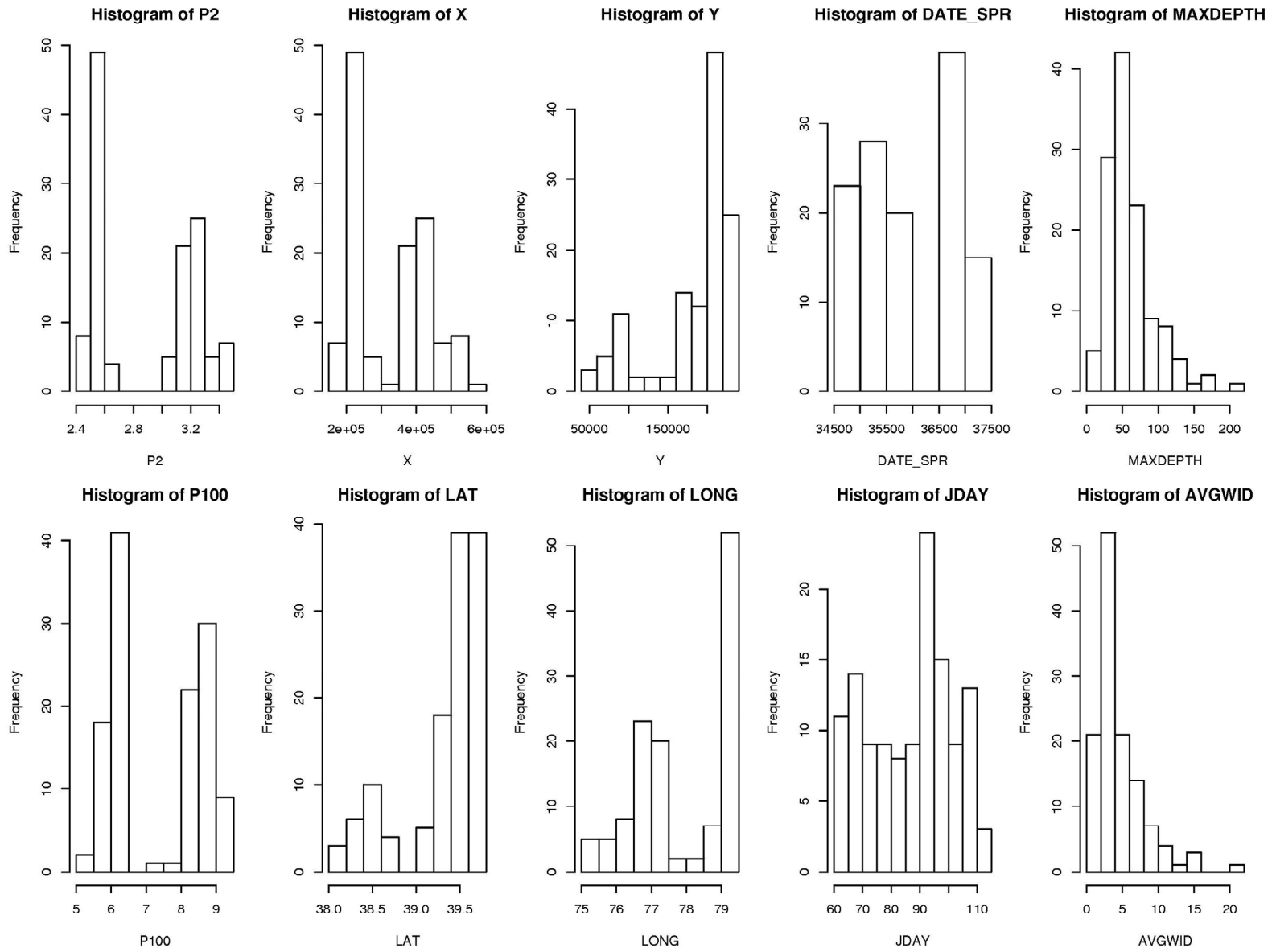
The time-base in Tables 43, 44, 45, and 46 were set up as the years after the restoration project completed. This was designed to see the recovery time scale after restoration. It was suggested by Dr. Glenn Moglen that if the time is based on the monitoring year for all the sites, then the cause of degrading condition can be determined to be either natural variance for all the site, e.g., drought effect, or the individual site effect.

In predictive model, the same method was used to develop the model in family and genus level. However, the family and genus taxon data are different, thus should be treat differently. It was suggested by all the committees to develop the genus level model in a hierarchical approach, so the precision can be improved by relating to family level model.

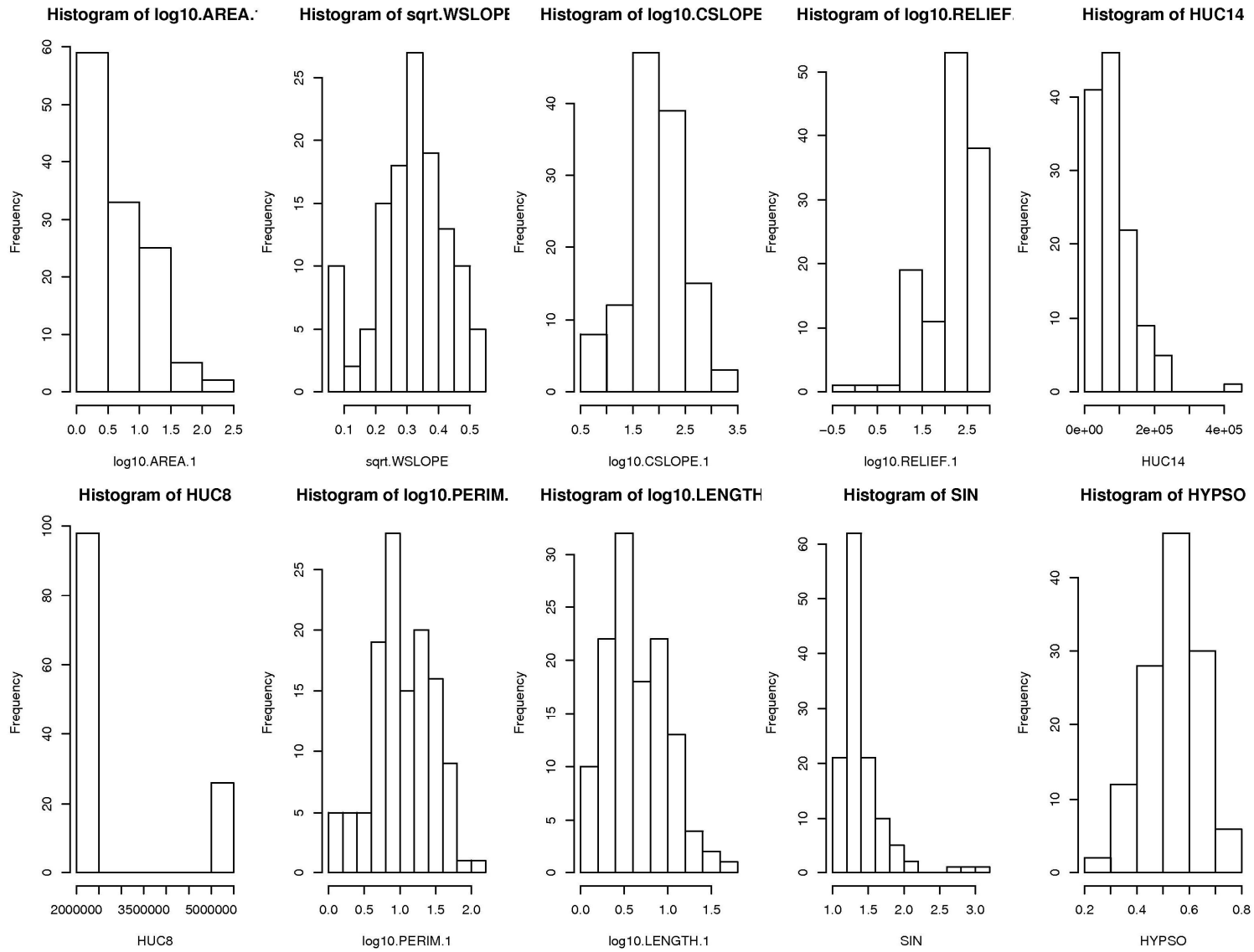
Appendix A-1. Potential Predictor Distribution

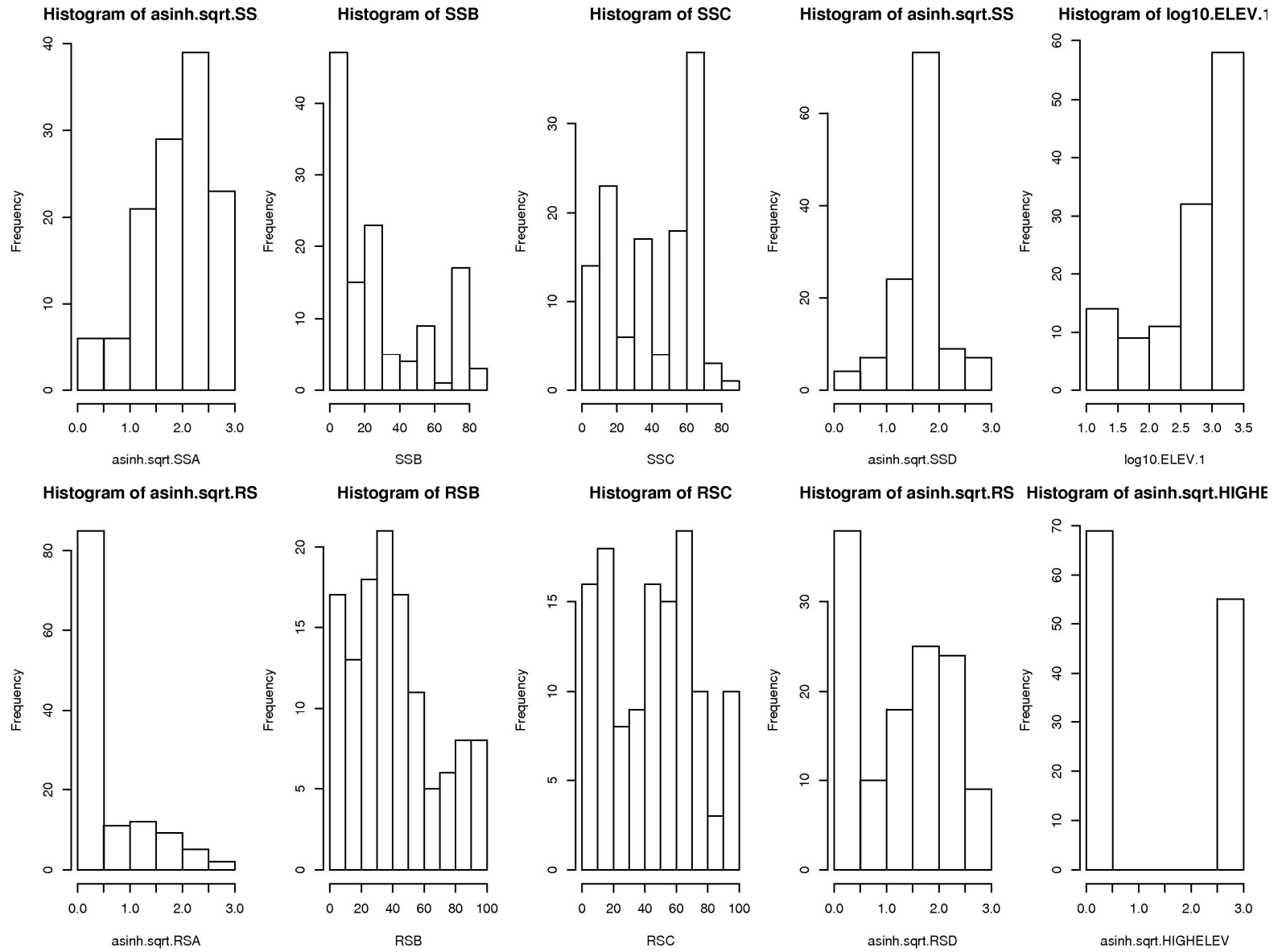


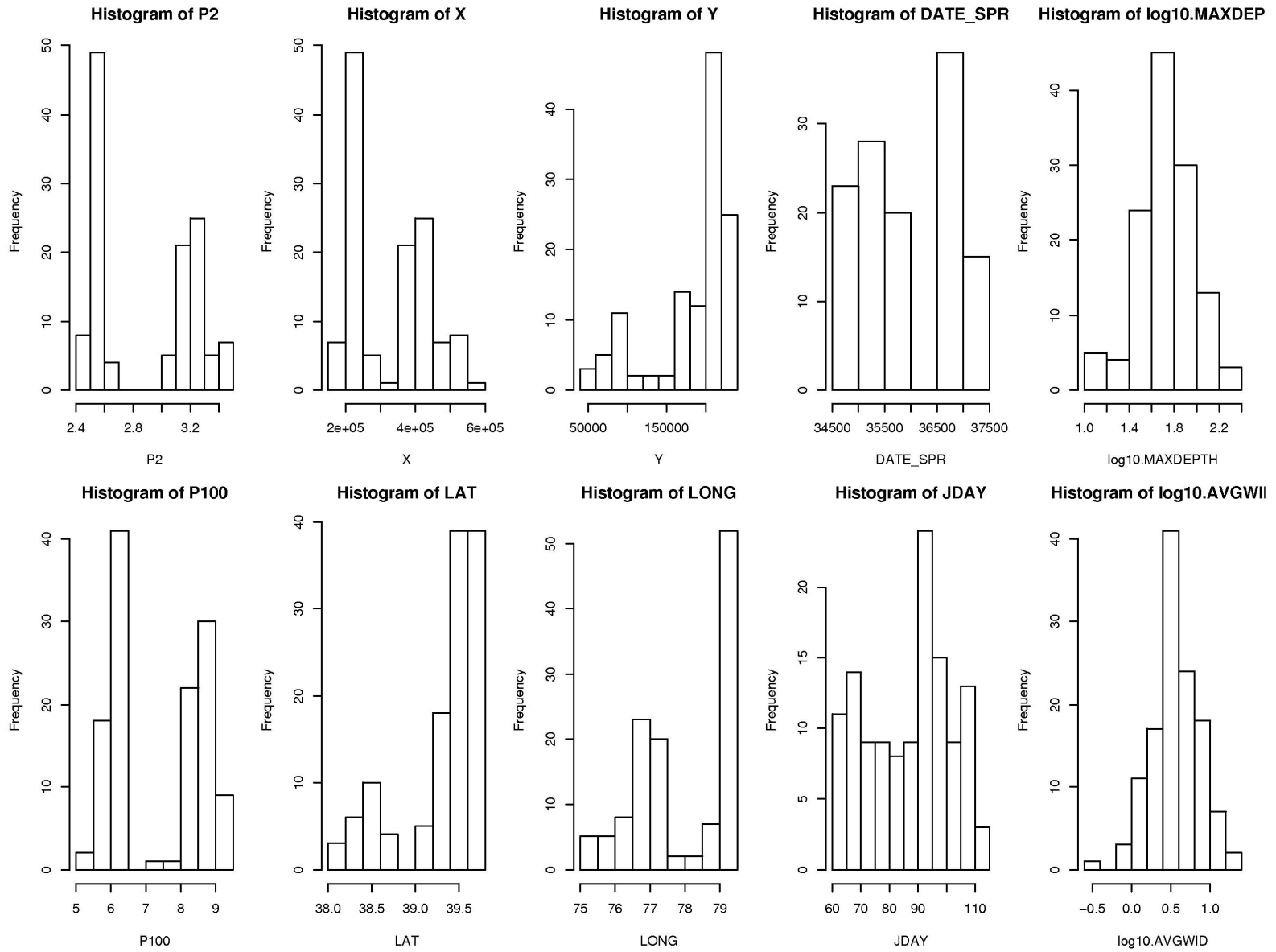




Appendix A-2. Transformed Potential Predictor Distribution







Appendix B. R-Code for Region-Style Models

```
#####  
# ROI Method For MBSS data taxa level with predictors  
# --- All predictors / GISHydro2000 predictors ---  
#  
#                               last update 01/15/08 -- Yin-Phan Tsang  
#  
# Input files:  
# "bugdata.txt"--MBSS bio data to taxa level, from all the sites,  
#               Data is in present/absent (1/0) with species names at the first row  
#               The first column is StationID, and the second is REF  
#               The REF column contains "R" as reference site, "V" as validation site  
#               "T" as test site  
# "pre.txt"--physical parameters generated from GISHydro for all the site  
#               The first column is StationID, and the second is REF  
#               The REF column contains "R" as reference site, "V" as validation site  
#               "T" as test site  
# plotname -- Give the name to desired plot  
#  
# Output files:  
# "Result.txt"--All the results including predators and coefficients for each dimension,  
#               and the final predicted taxon occurring probabilities with different Pc,  
#               and the Observed(O), Expected(E), and O/E ration  
# p. s. other results, like NMDS, calculated Euclidean distance...etc, can be printed  
#               just remove the "#" sign in front of those "write.table" lines  
#####
```

Best AIC and BIC

```
##-- save current workspace  
  
#input files & plot name  
fname <- "Genus.AllPreditors"  
bugdata <- read.delim("bugallgenus.txt")  
pre <- read.delim("AllPredictors.txt")  
  
write.table(c(fname, "AIC BIC selection"), file=paste(fname, "RMSE AICBICselection.txt"))  
  
# load library & R code & workspace  
  
#load(paste(fname, "subset555.RData"))  
library(labdsv)  
library(leaps)  
source("d.ij.r")  
source("ROIInull.prob.r")  
source("O.E5only.r")  
source("RMSE5only.r")  
  
##calculate draw rss,AIC&BIC for different model order  
  
mo<-vector()  
for(i in 1:(ncol(preRCAL)-1)) {  
  mo<-c(mo, rep(i, 5))  
}  
mo<-c(mo, ncol(preRCAL))  
mo<-matrix(data=rep(mo, dn), nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)  
parAIC<-nrow(preRCAL)*log(bestsubrss/nrow(preRCAL))+(2*nrow(preRCAL)*(mo+2))/(nrow(preRCAL)-(mo+2)-1)  
parBIC<-nrow(preRCAL)*log(bestsubrss/nrow(preRCAL))+(mo+2)*log(nrow(preRCAL))  
  
minmo<-matrix(data=0, nrow=2, ncol=dn)  
minmorow<-matrix(data=0, nrow=2, ncol=dn)
```



```

for(k in 1:dn){
  for(i in 1:(5*(ncol(preRCAL)-1)+1)){
    if(parAIC[i,k] == min(parAIC[,k])) {minmo[1,k]<-mo[i,k];minmorow[1,k]<-i}
    if(parBIC[i,k] == min(parBIC[,k])) {minmo[2,k]<-mo[i,k];minmorow[2,k]<-i}
  }
}

RMSE.AICBIC<-matrix(data=0,nrow=2,ncol=(dn+2))
m<-0
subpre<-vector("list",dn)
axislm<-vector("list",dn)
for(nn in 1:2){

  if(nn == 1) write.table("AIC",file=paste(fname,"RMSE AICBICselection.txt"),append=TRUE)
  if(nn == 2) write.table("BIC",file=paste(fname,"RMSE AICBICselection.txt"),append=TRUE)

  i<-minmorow[nn,1]
  ii<-minmorow[nn,2]
  iii<-minmorow[nn,3]
  subpre[[1]]<-vector()
  for(j in 2:ncol(bestsubpar[[1]])){
    if (bestsubpar[[1]][i,j] == TRUE) subpre[[1]] <-
c(subpre[[1]],names(NMDSpreRCAL[[1]][j]))
    write.table(subpre[[1]],file=paste(fname,"RMSE AICBICselection.txt"),append=TRUE)

    subpre[[2]]<-vector()
    for(j in 2:ncol(bestsubpar[[2]])){
      if (bestsubpar[[2]][ii,j] == TRUE) subpre[[2]] <-
c(subpre[[2]],names(NMDSpreRCAL[[2]][j]))
      write.table(subpre[[2]],file=paste(fname,"RMSE AICBICselection.txt"),append=TRUE)

      subpre[[3]]<-vector()
      for(j in 2:ncol(bestsubpar[[3]])){
        if (bestsubpar[[3]][iii,j] == TRUE) subpre[[3]] <-
c(subpre[[3]],names(NMDSpreRCAL[[3]][j]))
        write.table(subpre[[3]],file=paste(fname,"RMSE AICBICselection.txt"),append=TRUE)

        for(k in 1:dn){
          subsetpre <- NMDSpreRCAL[[k]][c("axis",subpre[[k]])]
          axislm[[k]]<- lm(axis ~ ., data=subsetpre)}

        ### Internal Validation (Calibration
Data)=====
        ## calculate the distance between test sites and reference sites
sdi j <- d.ij(dn,axislm,preRCAL,preRCAL)

        ## calculate the Pc for the ROI null model
n<-2
p <- ROIInull.prob(n,sdi j,RCAL)
OE.results <- O.E5only(RCAL,p)

O <- OE.results$O
E <- OE.results$E
RMSEi <-RMSE5only(O,E)

        ### External Validation (Validation
Data)=====
sdi j <- d.ij(dn,axislm,preRCAL,preRVAL)
p <- ROIInull.prob(n,sdi j,RCAL)
OE.results <- O.E5only(RVAL,p)

O <- OE.results$O

```

```

E <- OE.results$E
RMSEe <-RMSE5only(0, E)

m<-m+1
RMSE. AICBIC[m, ]<- c(i, ii, iii, RMSEi, RMSEe)

save.image(paste(fname, "RMSE AICBIC.RData"))
}
RMSE. AICBIC<-as.data.frame(RMSE. AICBIC)
names(RMSE. AICBIC)<-c("axis1", "axis2", "axis3", "internal", "external")
row.names(RMSE. AICBIC)<-c("AIC", "BIC")
write.table(RMSE. AICBIC, file=paste(fname, "RMSE AICBICselection.txt"), append=TRUE)
#####
#####

##draw RMSE internal&external
postscript(paste(fname, "RMSE AICBIC.ps"))

mo<-vector()
for(i in 1:(ncol(preRCAL)-1)) {
  mo<-c(mo, rep(i, (5^dn)))}
mo<-as.matrix(c(mo, ncol(preRCAL)))
RMSEi.m<-vector();RMSEe.m<-vector()
RMSEi<-matrix(data=RMSE.sum[-nrow(RMSE.sum), 4], nrow=(5*5*5), ncol=(ncol(preRCAL)-1))
RMSEe<-matrix(data=RMSE.sum[-nrow(RMSE.sum), 5], nrow=(5*5*5), ncol=(ncol(preRCAL)-1))
for(i in 1:(ncol(preRCAL)-1)) {
  RMSEi.m<-c(RMSEi.m, mean(RMSEi[, i]))
  RMSEe.m<-c(RMSEe.m, mean(RMSEe[, i]))}
RMSEi.m<-c(RMSEi.m, RMSE.sum[nrow(RMSE.sum), 4])
RMSEe.m<-c(RMSEe.m, RMSE.sum[nrow(RMSE.sum), 5])

par(mfrow=c(2, 1))
par(mar=c(0, 4, 4, 2), cex=0.8)
plot(mo, RMSE.sum[, 4], axes=FALSE, ylab="RMSE", main=fname)
axis(1, lab=NA);axis(2);box()
lines(1:ncol(preRCAL), RMSEi.m, col="blue", lwd=2)
lines(c(min(as.integer(RMSE.AICBIC[1, 1:3]/5)), max(as.integer(RMSE.AICBIC[1, 1:3]/5))), c(rep(RMSE.AICBIC[1, 4], 2)), col="red", lwd=2)
lines(c(min(as.integer(RMSE.AICBIC[2, 1:3]/5)), max(as.integer(RMSE.AICBIC[2, 1:3]/5))), c(rep(RMSE.AICBIC[2, 4], 2)), col="purple", lwd=2)
text((max(mo)-5), (min(RMSE.sum[, 4])+0.02), "internal")
par(mar=c(5, 4, 0, 2))
plot(mo, RMSE.sum[, 5], ylab="RMSE", xlab="Model order")
lines(1:ncol(preRCAL), RMSEe.m, col="blue", lwd=2)
lines(c(min(as.integer(RMSE.AICBIC[1, 1:3]/5)), max(as.integer(RMSE.AICBIC[1, 1:3]/5))), c(rep(RMSE.AICBIC[1, 5], 2)), col="red", lwd=2)
lines(c(min(as.integer(RMSE.AICBIC[2, 1:3]/5)), max(as.integer(RMSE.AICBIC[2, 1:3]/5))), c(rep(RMSE.AICBIC[2, 5], 2)), col="purple", lwd=2)
text((max(mo)-5), (min(RMSE.sum[, 5])+0.02), "external")
dev.off()

save.image(paste(fname, "RMSE AICBIC.RData"))

```

Subset Best 5 at Each Axis

```

##-- save current workspace

#input files & plot name
fname <- "Genus.SummaryPredictors"
bugdata <- read.delim("bugallgenus.txt")

```

```

pre <- read.delim("SummaryPredictors.txt")

# load library & R code

library(labdsv)
library(leaps)
source("d.ij.r")
source("ROInull.prob.r")
source("O.E5only.r")
source("RMSE5only.r")

#omit the NA
merg <- merge(bugdata, pre, by = intersect(names(bugdata), names(pre)))
om.merg <- na.omit(merg, subset=TRUE)
bugdata <- subset(om.merg, select = names(bugdata))
pre <- subset(om.merg, select = names(pre))

#extract reference sites bio-data and physical parameters
RCAL<-subset(bugdata, REF == "R")[,3:ncol(bugdata)]
RVAL<-subset(bugdata, REF == "V")[,3:ncol(bugdata)]
preRCAL<-subset(pre, REF == "R")[,3:ncol(pre)]
preRVAL<-subset(pre, REF == "V")[,3:ncol(pre)]

#Bray-Curtis distance matrix
RCAL.bc <- dsvdis(RCAL, index="bray/curtis")

#Non-metric dimensional scaling
stress <- 100
dn <- 1
while(stress > 25){
dn <- dn+1
RCAL.nmds <- isoMDS(RCAL.bc, k=dn)
stress <- as.numeric(RCAL.nmds[2])}

#write.table("RCAL NMDS", file="Result.txt", append=TRUE)
#write.table(RCAL.nmds, file="Result.txt", append=TRUE)

NMDS<- data.frame(RCAL.nmds[1])

NMDSpreRCAL <- list(1:dn)
bestsubpar <- list(1:dn)
bestsubrss <-matrix(data=0, nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)

for(k in 1:dn){
NMDSpreRCAL[[k]] <- data.frame(NMDS[k], preRCAL)
names(NMDSpreRCAL[[k]]) <- c("axis", names(preRCAL)) # rename the NMDSpreRCAL data frame columns
findbest5sub<-regsubsets(axis~., data=NMDSpreRCAL[[k]], nbest=5, nvmax=ncol(preRCAL))
bestsubpar[[k]]<-summary(findbest5sub)$which
bestsubrss[,k]<-summary(findbest5sub)$rss
}

##calculate draw rss,AIC&BIC for different model order
postscript(paste(fname, "rss AIC BIC.ps"))
mo<-vector()
for(i in 1:(ncol(preRCAL)-1)){
mo<-c(mo, rep(i, 5))}
mo<-c(mo, ncol(preRCAL))
mo<-matrix(data=rep(mo, dn), nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)
parAIC<-nrow(preRCAL)*log(bestsubrss/nrow(preRCAL))+(2*nrow(preRCAL)*(mo+2))/(nrow(preRCAL)-(mo+2)-1)
parBIC<-nrow(preRCAL)*log(bestsubrss/nrow(preRCAL))+(mo+2)*log(nrow(preRCAL))

axisno<-vector()

```

```

minmo<-matrix(data=0,nrow=2,ncol=dn)
for(k in 1:dn){
  for(i in 1:(5*(ncol(preRCAL)-1)+1)){
    if(parAIC[i,k] == min(parAIC[,k])) minmo[1,k]<-mo[i,k]
    if(parBIC[i,k] == min(parBIC[,k])) minmo[2,k]<-mo[i,k]}
  axisno<-c(axisno,rep(k,(5*(ncol(preRCAL)-1)+1)))}
par(mfrow=c(3,1))
par(mar=c(1,4,2,2))
pbestsubrssl<-data.frame(axisno,c(mo),c(bestsubrssl))
plot(pbestsubrssl[,2],pbestsubrssl[,3],cex=0.8,ylab="RSS",main=fname,
     pch=as.numeric(pbestsubrssl[,1]),col=as.numeric(pbestsubrssl[,1]))
legend((max(mo)-max(mo)/4),(max(bestsubrssl)),legend=c("axis 1","axis 2","axis 3"),
     pch=1:dn,col=1:dn)
par(mar=c(2,4,1,2))
pparAIC<-data.frame(axisno,c(mo),c(parAIC))
plot(pparAIC[,2],pparAIC[,3],cex=0.8,ylab="AIC",
     pch=as.numeric(pparAIC[,1]),col=as.numeric(pparAIC[,1]))
text(5,(max(parAIC)-10),adj=0,
     paste("axis 1-",minmo[1,1],"\naxis 2-",minmo[1,2],"\naxis 3-",minmo[1,3]))
par(mar=c(5,4,0,2))
pparBIC<-data.frame(axisno,c(mo),c(parBIC))
plot(pparBIC[,2],pparBIC[,3],cex=0.8,xlab=c("Model order"),ylab="BIC",
     pch=as.numeric(pparBIC[,1]),col=as.numeric(pparBIC[,1]))
text(5,(max(parBIC)-20),adj=0,
     paste("axis 1-",minmo[2,1],"\naxis 2-",minmo[2,2],"\naxis 3-",minmo[2,3]))
dev.off()

#####
*****
## try best 5 from each orders,
##

RMSE.sum<-matrix(data=0,nrow=(5*5*5*(ncol(preRCAL)-1)+1),ncol=5)
axislm <-vector("list",dn)
subpre<-vector("list",dn)
m<-0
for(nn in seq(1,(5*(ncol(preRCAL)-1)+1),5)){
  if(nn<(5*(ncol(preRCAL)-1)+1)){
    for(i in nn:(nn+4)){
      subpre[[1]]<-vector()
      for(j in 2:ncol(bestsubpar[[1]])){
        if(bestsubpar[[1]][i,j] == TRUE) subpre[[1]] <-
c(subpre[[1]],names(NMDSpreRCAL[[1]])[j])}
      for(ii in nn:(nn+4)){
        subpre[[2]]<-vector()
        for(j in 2:ncol(bestsubpar[[2]])){
          if(bestsubpar[[2]][ii,j] == TRUE) subpre[[2]] <-
c(subpre[[2]],names(NMDSpreRCAL[[2]])[j])}
        for(iii in nn:(nn+4)){
          subpre[[3]]<-vector()
          for(j in 2:ncol(bestsubpar[[3]])){
            if(bestsubpar[[3]][iii,j] == TRUE) subpre[[3]] <-
c(subpre[[3]],names(NMDSpreRCAL[[3]])[j])}

          for(k in 1:dn){
            subsetpre <- NMDSpreRCAL[[k]][c("axis",subpre[[k]])]
            axislm[[k]]<-lm(axis ~ ., data=subsetpre)}

          ### Internal Validation (Calibration
Data)=====
          ## calculate the distance between test sites and reference sites

```

```

sdi j <- d.ij(dn, axislm, preRCAL, preRCAL)

## calculate the Pc for the ROI null model
n<-2
p <- ROIInull.prob(n, sdi j, RCAL)
OE.results <- O.E5only(RCAL, p)

O <- OE.results$O
E <- OE.results$E
RMSEi <-RMSE5only(O, E)

### External Validation (Validation
Data)=====
sdi j <- d.ij(dn, axislm, preRCAL, preRVAL)
p <- ROIInull.prob(n, sdi j, RCAL)
OE.results <- O.E5only(RVAL, p)

O <- OE.results$O
E <- OE.results$E
RMSEe <-RMSE5only(O, E)

m<-m+1
RMSE.sum[m, ] <- c(i, ii, iii, RMSEi, RMSEe)
}}}}
if(nn==(5*(ncol(preRCAL)-1)+1)) {
i<-nn
ii<-nn
iii<-nn
subpre[[1]]<-vector()
for(j in 2:ncol(bestsubpar[[1]])) {
if (bestsubpar[[1]][i, j] == TRUE) subpre[[1]] <-
c(subpre[[1]], names(NMDSpreRCAL[[1]])[j])

subpre[[2]]<-vector()
for(j in 2:ncol(bestsubpar[[2]])) {
if (bestsubpar[[2]][ii, j] == TRUE) subpre[[2]] <-
c(subpre[[2]], names(NMDSpreRCAL[[2]])[j])

subpre[[3]]<-vector()
for(j in 2:ncol(bestsubpar[[3]])) {
if (bestsubpar[[3]][iii, j] == TRUE) subpre[[3]] <-
c(subpre[[3]], names(NMDSpreRCAL[[3]])[j])

for(k in 1:dn) {
subsetpre <- NMDSpreRCAL[[k]][c("axis", subpre[[k]])]
axislm[[k]]<- lm(axis ~ ., data=subsetpre)}

### Internal Validation (Calibration
Data)=====
## calculate the distance between test sites and reference sites
sdi j <- d.ij(dn, axislm, preRCAL, preRCAL)

## calculate the Pc for the ROI null model
n<-2
p <- ROIInull.prob(n, sdi j, RCAL)
OE.results <- O.E5only(RCAL, p)

O <- OE.results$O
E <- OE.results$E
RMSEi <-RMSE5only(O, E)

```

```

##### External Validation (Validation
Data)=====
      sdi.j <- d.i.j(dn, axis1m, preRCAL, preRVAL)
      p <- ROI.null.prob(n, sdi.j, RCAL)
      OE.results <- O.E5only(RVAL, p)

      O <- OE.results$O
      E <- OE.results$E
      RMSEe <-RMSE5only(O, E)

      m<-m+1
      RMSE.sum[m, ] <- c(i, ii, iii, RMSEi, RMSEe)
    }
  print(nn)
  save.image(paste(fname, "subset555.RData"))
}
write.table(RMSE.sum, file=paste(fname, "RMSE.txt"))
#####
#####

##draw RMSE internal&external
postscript(paste(fname, "RMSE.ps"))
mo<-vector()
for(i in 1:(ncol(preRCAL)-1)) {
  mo<-c(mo, rep(i, (5^dn)))}
mo<-c(mo, ncol(preRCAL))
axisie<-c(rep(1, nrow(RMSE.sum)), rep(2, nrow(RMSE.sum)))
pRMSE<-data.frame(axisie, rep(mo, 2), c(as.matrix(RMSE.sum[, 4:5])))
plot(pRMSE[, 2], pRMSE[, 3], cex=0.5, ylab="RMSE", xlab="Model order", main=fname,
     pch=as.numeric(pRMSE[, 1]), col=as.numeric(pRMSE[, 1]))
legend((max(mo)-max(mo)/4), (max(pRMSE[, 3])), legend=c("internal", "external"),
      pch=1:2, col=1:2)
dev.off()

```

Tij and Correlation

```

##-- save current workspace

#input files & plot name
fname <- "Genus.Allpredictors"
bugdata <- read.delim("bugallgenus.txt")
pre <- read.delim("AllPredictors.txt")

# load library & R code & workspace

library(labds)
library(leaps)
source("d.i.j.r")
source("O.E.r")
source("OE.plot.r")
source("result.5sortn.r")

#omit the NA
merg <- merge(bugdata, pre, by = intersect(names(bugdata), names(pre)))
om.merg <- na.omit(merg, subset=TRUE)
bugdata <- subset(om.merg, select = names(bugdata))
pre <- subset(om.merg, select = names(pre))

```

```

#transform predictor variables
# for 11/09/2007 new transform, mixed with dr. Paul and mine
#attach(pre)
#pre <- transform(pre, "log10. AREA. 1" = log10(AREA+1))
#pre <- transform(pre, "sqrt. WSLOPE" = sqrt(WSLOPE))
#pre <- transform(pre, "log10. CSLOPE. 1" = log10(CSLOPE+1))
#pre <- transform(pre, "log10. RELIEF. 1" = log10(RELIEF+1))
#pre <- transform(pre, "log10. PERIM. 1" = log10(PERIM+1))
#pre <- transform(pre, "log10. LENGTH. 1" = log10(LENGTH+1))
#pre <- transform(pre, "asinh. sqrt. RSA" = asinh(sqrt(RSA)))
#pre <- transform(pre, "asinh. sqrt. RSD" = asinh(sqrt(RSD)))
#pre <- transform(pre, "log10. ELEV. 1" = log10(ELEV+1))
#pre <- transform(pre, "asinh. sqrt. HIGHELEV" = asinh(sqrt(HIGHELEV)))
#if((fname == "Family.GISHydropredictors") | (fname == "Genus.GISHydropredictors")){
#   pre <- data.frame(pre[-c(3, 4, 5, 6, 10, 11, 12, 15, 16, 20)])}
#if((fname == "Family.Allpredictors") | (fname == "Genus.Allpredictors")){
#   pre <- transform(pre, "log10. MAXDEPTH" = log10(MAXDEPTH))
#   pre <- transform(pre, "log10. AVGWID" = log10(AVGWID))
#   pre <- data.frame(pre[-c(5, 6, 7, 8, 12, 13, 14, 17, 18, 22, 29, 30)])}
#detach(pre)

#extract reference sites bio-data and physical parameters
RCAL<-subset(bugdata, REF == "R")[, 3:ncol(bugdata)]
RVAL<-subset(bugdata, REF == "V")[, 3:ncol(bugdata)]
preRCAL<-subset(pre, REF == "R")[, 3:ncol(pre)]
preRVAL<-subset(pre, REF == "V")[, 3:ncol(pre)]

#Bray-Curtis distance matrix
RCAL.bc <- dsvdis(RCAL, index="bray/curtis")

#Non-metric dimensional scaling
stress <- 100
dn <- 1
while(stress > 25){
  dn <- dn+1
  RCAL.nmDS <- isoMDS(RCAL.bc, k=dn)
  stress <- as.numeric(RCAL.nmDS[2])}

NMDS<- data.frame(RCAL.nmDS[1])

#write.table(NMDS, file=paste(fname, "forTij.txt"), append=TRUE)
#write.table(preRCAL, file=paste(fname, "forTij.txt"), append=TRUE)

Sy<-as.data.frame(t(sd(NMDS)))
Sx<-as.data.frame(t(sd(preRCAL)))

NMDSpreRCAL <- list(1:dn)
stepaxis <- list(1:dn)
bestsubpar <- list(1:dn)
bestsubrSS<-matrix(data=100, nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)
write.table(paste(fname, "stepAIC"), file=paste(fname, "forTij.txt"), append=TRUE)

for(k in 1:dn){
  NMDSpreRCAL[[k]] <- data.frame(NMDS[k], preRCAL)
  names(NMDSpreRCAL[[k]]) <- c(paste("axis", k), names(preRCAL)) # rename the NMDSpreRCAL data frame
  columns

  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(14, 11)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 31, 11)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(31, 2, 4, 11)]
}

```

```

#plotname<-paste(fname, "axis", k)
#write.table(cor(NMDSpreRCAL[[k]]), file=paste(fname, "correlation.txt"), append=TRUE)
#postscript(paste(plotname, ".ps"))
#pairs(NMDSpreRCAL[[k]], gap=0, cex=0.4, main=plotname)
#dev.off()

##StepwiseAIC
names(NMDSpreRCAL[[k]][1] <- "axis" # rename the NMDSpreRCAL data frame columns
axislm <- lm(axis ~ ., data= NMDSpreRCAL[[k]])
stepaxis[[k]]<- stepAIC(axislm, direction="both", trace=FALSE)
axiscoef<-as.data.frame(t(coef(stepaxis[[k]])))
for(jj in 2:ncol(axiscoef)) {
axiscoef[2, jj]<-Sx[1, names(axiscoef)[jj]]
axiscoef[3, jj]<-axiscoef[1, jj]*axiscoef[2, jj]/Sy[k]
rownames(axiscoef)<-c("coef", "Sx", "Tij")
write.table(axiscoef, file=paste(fname, "forTij.txt"), append=TRUE)

# exhaustive AIC, exhaustive BIC
findbest5sub<-regsubsets(axis~, data=NMDSpreRCAL[[k]], nbest=5, nvmax=ncol(preRCAL))
bestsubpar[[k]]<-summary(findbest5sub)$which
bestsubrssi[1: (5*(ncol(NMDSpreRCAL[[k]])-2)+1), k]<-summary(findbest5sub)$rss
}

mo<-vector()
for(i in 1:(ncol(preRCAL)-1)) {
mo<-c(mo, rep(i, 5))}
mo<-c(mo, ncol(preRCAL))
mo<-matrix(data=rep(mo, dn), nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)

parAIC<-nrow(preRCAL)*log(bestsubrssi/nrow(preRCAL))+(2*nrow(preRCAL)*(mo+2))/(nrow(preRCAL)-(mo+2)-1)
parBIC<-nrow(preRCAL)*log(bestsubrssi/nrow(preRCAL))+(mo+2)*log(nrow(preRCAL))

minmo<-matrix(data=0, nrow=2, ncol=dn)
minmorow<-matrix(data=0, nrow=2, ncol=dn)
for(k in 1:dn) {
for(i in 1: (5*(ncol(preRCAL)-1)+1)) {
if(parAIC[i, k] == min(parAIC[, k])) {minmo[1, k]<-mo[i, k]; minmorow[1, k]<-i}
if(parBIC[i, k] == min(parBIC[, k])) {minmo[2, k]<-mo[i, k]; minmorow[2, k]<-i}
}
}

m<-0
subpre<-vector("list", dn)
axislm<-vector("list", dn)
for(nn in 1:2) {

if(nn == 1) plotname<-paste(fname, "best AIC")
if(nn == 2) plotname<-paste(fname, "best BIC")

i<-minmorow[nn, 1]
ii<-minmorow[nn, 2]
iii<-minmorow[nn, 3]
subpre[[1]]<-vector()
for(j in 2:ncol(bestsubpar[[1]])) {
if (bestsubpar[[1]][i, j] == TRUE) subpre[[1]] <-
c(subpre[[1]], names(as.data.frame(bestsubpar[[1]])) [j])}

subpre[[2]]<-vector()
for(j in 2:ncol(bestsubpar[[2]])) {

```



```

        if (bestsubpar[[2]][ii, j] == TRUE) subpre[[2]] <-
c(subpre[[2]], names(as.data.frame(bestsubpar[[2]])[j]))

        subpre[[3]]<-vector()
        for(j in 2:ncol(bestsubpar[[3]])){
            if (bestsubpar[[3]][iii, j] == TRUE) subpre[[3]] <-
c(subpre[[3]], names(as.data.frame(bestsubpar[[3]])[j]))

            write.table(plotname, file=paste(fname, "forTij.txt"), append=TRUE)
            for(k in 1:dn){
                subsetpre <- NMDSPreRCAL[[k]][c("axis", subpre[[k]])]
                axislm[[k]]<- lm(axis ~ ., data=subsetpre)
                axiscoef<-as.data.frame(t(coef(axislm[[k]])))
                for(jj in 2:ncol(axiscoef)){
                    axiscoef[2, jj]<-Sx[1, names(axiscoef)[jj]]
                    axiscoef[3, jj]<-axiscoef[1, jj]*axiscoef[2, jj]/Sy[k]
                    rownames(axiscoef)<-c("coef", "Sx", "Tij")}
                write.table(axiscoef, file=paste(fname, "forTij.txt"), append=TRUE)

            }

        }
}

```

ANNA Select Neighbor Number with RMSE

```

##-- save current workspace

# input files & plot name
mname<- "ANNA"
fname <- paste("Genus.Allpredictors")
bugdata <- read.delim("bugallgenus.txt")
pre <- read.delim("AllPredictors.txt")

# load library & R code & workspace

library(labdsv)
source("d.ij.r")
source("ANNA.prob.r")
source("O.E.r")
source("OE.plot.r")
source("result.RMSEsortanna.r")
source("stepwise.AIC.r")
source("ex.AICBIC.r")

#omit the NA
merg <- merge(bugdata, pre, by = intersect(names(bugdata), names(pre)))
om.merg <- na.omit(merg, subset=TRUE)
bugdata <- subset(om.merg, select = names(bugdata))
pre <- subset(om.merg, select = names(pre))

#transform predictor variables
# for 11/09/2007 new transform, mixed with dr. Paul and mine
#attach(pre)
#pre <- transform(pre, "log10.AREA.1" = log10(AREA+1))
#pre <- transform(pre, "sqrt.WSLOPE" = sqrt(WSLOPE))
#pre <- transform(pre, "log10.CSLOPE.1" = log10(CSLOPE+1))
#pre <- transform(pre, "log10.RELIEF.1" = log10(RELIEF+1))
#pre <- transform(pre, "log10.PERIM.1" = log10(PERIM+1))
#pre <- transform(pre, "log10.LENGTH.1" = log10(LENGTH+1))
#pre <- transform(pre, "asinh.sqrt.RSA" = asinh(sqrt(RSA)))

```

```

#pre <- transform(pre, "asinh.sqrt.RSD" = asinh(sqrt(RSD)))
#pre <- transform(pre, "log10.ELEV.1" = log10(ELEV+1))
#pre <- transform(pre, "asinh.sqrt.HIGHELEV" = asinh(sqrt(HIGHELEV)))
#if((fname == "Family.GISHydropredictors") | (fname == "Genus.GISHydropredictors")){
#   pre <- data.frame(pre[-c(3, 4, 5, 6, 10, 11, 12, 15, 16, 20)])}
#if((fname == "Family.Allpredictors") | (fname == "Genus.Allpredictors")){
#   pre <- transform(pre, "log10.MAXDEPTH" = log10(MAXDEPTH))
#   pre <- transform(pre, "log10.AVGWID" = log10(AVGWID))
#   pre <- data.frame(pre[-c(5, 6, 7, 8, 12, 13, 14, 17, 18, 22, 29, 30)])}
#detach(pre)

#extract reference sites bio-data and physical parameters
RCAL<-subset(bugdata, REF == "R")[, 3:ncol(bugdata)]
RVAL<-subset(bugdata, REF == "V")[, 3:ncol(bugdata)]
preRCAL<-subset(pre, REF == "R")[, 3:ncol(pre)]
preRVAL<-subset(pre, REF == "V")[, 3:ncol(pre)]

#eliminate two outlier in the validation data
RVAL<-RVAL[-c(10, 19), ]
preRVAL<-preRVAL[-c(10, 19), ]

#Bray-Curtis distance matrix
RCAL.bc <- dsvdis(RCAL, index="bray/curtis")

#Non-metric dimensional scaling
stress <- 100
dn <- 1
while(stress > 25){
dn <- dn+1
RCAL.nmds <- isoMDS(RCAL.bc, k=dn)
stress <- as.numeric(RCAL.nmds[2])}

NMDS<- data.frame(RCAL.nmds[1])

NMDSpreRCAL <- list(1:dn)

for(k in 1:dn){
NMDSpreRCAL[[k]] <- data.frame(NMDS[k], preRCAL)
names(NMDSpreRCAL[[k]]) <- c("axis", names(preRCAL)) # rename the NMDSpreRCAL data frame columns

# After Rational Checking
if(fname == "Family.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(11, 9, 17, 23, 21)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 24, 31, 21, 17, 12)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(23, 19, 21, 25)]}

if(fname == "Family.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(15, 21, 19)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(24, 23, 15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(17, 21)]}

if(fname == "Genus.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(14, 11)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 31, 11)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(31, 2, 4, 11)]}

if(fname == "Genus.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21, 22, 19, 8, 2)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21, 22, 17, 3, 2, 15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(19, 24, 17, 23, 22)]}

```

```

}

# different selecting parameter method
# nn=1:Exhaustive AIC, nn=2:Exhaustive BIC, nn=3:StepwiseAIC

for(nn in 1:3){

# pick the selecting parameter method
if(nn == 1) {
  plotname<-paste(fname,mname,"best AIC")
  axislm<-ex.AICBIC(nn,dn,NMDSpreRCAL)}
if(nn == 2) {
  plotname<-paste(fname,mname,"best BIC")
  axislm<-ex.AICBIC(nn,dn,NMDSpreRCAL)}
if(nn == 3) {
  plotname <- paste(fname,mname,"stepwiseAIC")
  axislm<-stepwise.AIC(dn,NMDSpreRCAL)}
write.table(plotname,file=paste(plotname,".txt"), append=TRUE)
for(k in 1:dn) write.table(coef(axislm[[k]]), file=paste(plotname,".txt"), append=TRUE)

  m<-0
  results.int<-matrix(data=0,nr=1,nc=16)
  results.int<-as.data.frame(results.int)
  names(results.int)<-c("NeiNo",
    "Pc0.mean","Pc0.sd","Pc0.slope","Pc0.intercept","Pc0.r2",
    "Pc0.01.mean","Pc0.01.sd","Pc0.01.slope","Pc0.01.intercept","Pc0.01.r2",
    "Pc0.5.mean","Pc0.5.sd","Pc0.5.slope","Pc0.5.intercept","Pc0.5.r2")

  ##check different Neighbor number

  for(neino in 2:120){

    m<-m+1
    ### Internal Validation (Calibration
Data)=====
    ## calculate the distance between test sites and reference sites
    sdi j <- d.ij(dn,axislm,preRCAL,preRCAL)

    ## calculate the Pc for the ROI null model
    p <- ANNA.prob(neino,sdi j,RCAL)
    OE.results <- O.E(RCAL,p)

    OEsum <- OE.results$results
    results.int[m,]<-c(neino,OEsum)
    if(m%%50==0) print(m)

  }

  ## sort the results depends on
  ## Pc>0.5's mean,sd,slope,intercept,and r^2 geometric mean
  # results of internal
  sort.int<-result.RMSEsortanna(results.int)
  write.table("internal sort results", file=paste(plotname,".txt"),append=TRUE)
  write.table(sort.int[1:10,], file=paste(plotname,".txt"),append=TRUE)
  save.image(paste(plotname,".RData"))

  m<-0
  results.ext<-matrix(data=0,nr=1,nc=16)

```

```

results.ext<-as.data.frame(results.ext)
names(results.ext)<-c("NeiNo",
                    "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",
                    "Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",
                    "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2")

## draw the top 10 combination graphs
for(top in 1:10){

    postscript(paste(plotname, "top", top, ".ps"))

    ### Internal Validation (Calibration
Data)=====
    ## calculate the distance between test sites and reference sites
    sdi j <- d. i j(dn, axislm, preRCAL, preRCAL)

    neino<-sort.int$NeiNo[top]

    ## calculate the Pc for the ROI null model
    p <- ANNA. prob(neino, sdi j, RCAL)
    OE. results <- O. E(RCAL, p)

    #draw the graphs
    op<-par(mfrow=c(4, 3), mar=c(4, 4, 5, 1))
    O <- OE. results$O
    E <- OE. results$E
    OEsu m <- OE. results$results
    plot. main<-paste(plotname, "\nNeighbor number =", neino, "\nInternal")
    OE. plot(O, E, OEsu m, plot. main)

    ### External Validation (Validation
Data)=====
    m <- m+1
    sdi j <- d. i j(dn, axislm, preRCAL, preRVAL)
    p <- ANNA. prob(neino, sdi j, RCAL)
    OE. results <- O. E(RVAL, p)
    O <- OE. results$O
    E <- OE. results$E
    OEsu m <- OE. results$results
    results. ext[m, ]<-c(neino, OEsu m)

    plot. main<-c("external")
    par(mar=c(4, 4, 2, 1))
    OE. plot(O, E, OEsu m, plot. main)

    dev. off()
}

write. table("external results", file=paste(plotname, ".txt"), append=TRUE)
write. table(results. ext, file=paste(plotname, ".txt"), append=TRUE)
## sort the results depends on
## Pc>0.5's mean, sd, slope, intercept, and r^2 geometric mean
# results of external
write. table("external sort results", file=paste(plotname, ".txt"), append=TRUE)
sort. ext<-result. RMSEsortanna(results. ext)
write. table(sort. ext, file=paste(plotname, ".txt"), append=TRUE)
save. image(paste(plotname, ".RData"))

}

```

ANNA Model

```
##-- save current workspace

# input files & plot name
fname <- "Genus.GISHydropredictors"
bugdata <- read.delim("bugallgenus.txt")
pre <- read.delim("GISHydroPredictors.txt")
Smeth <- "StepwiseAIC"
neino <- 18

if(Smeth == "AIC") {nn<-1
                    plotname<-paste(fname,"ANNA best AIC")}
if(Smeth == "BIC") {nn<-2
                    plotname<-paste(fname,"ANNA best BIC")}
if(Smeth == "StepwiseAIC") {nn<-3
                             plotname <- paste(fname,"ANNA stepwiseAIC")}

# load library & R code & workspace

library(labdsv)
library(leaps)
source("d.ij.r")
source("ANNA.prob.r")
source("O.E.r")
source("OE.plot.r")
source("result.5sortanna.r")

#omit the NA
merg <- merge(bugdata, pre, by = intersect(names(bugdata), names(pre)))
om.merg <- na.omit(merg, subset=TRUE)
bugdata <- subset(om.merg, select = names(bugdata))
pre <- subset(om.merg, select = names(pre))

#extract reference sites bio-data and physical parameters
RCAL<-subset(bugdata, REF == "R")[,3:ncol(bugdata)]
RVAL<-subset(bugdata, REF == "V")[,3:ncol(bugdata)]
preRCAL<-subset(pre, REF == "R")[,3:ncol(pre)]
preRVAL<-subset(pre, REF == "V")[,3:ncol(pre)]

# modify RVAL data, pick out 2 point
RVAL<-RVAL[-c(10,19),]
preRVAL<-preRVAL[-c(10,19),]

#Bray-Curtis distance matrix
RCAL.bc <- dsvdis(RCAL, index="bray/curtis")

#Non-metric dimensional scaling
stress <- 100
dn <- 1
while(stress > 25){
dn <- dn+1
RCAL.nmds <- isoMDS(RCAL.bc, k=dn)
stress <- as.numeric(RCAL.nmds[2])}

NMDS<- data.frame(RCAL.nmds[1])

NMDSpreRCAL <- list(1:dn)
stepaxis <- list(1:dn)
bestsubpar <- list(1:dn)
```

```

bestsubrssl <-matrix(data=100, nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)

for(k in 1:dn){
  NMDSpreRCAL[[k]] <- data.frame(NMDS[k], preRCAL)
  names(NMDSpreRCAL[[k]]) <- c("axis", names(preRCAL)) # rename the NMDSpreRCAL data frame columns

# After Rational Checking
if(fname == "Family.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(11, 9, 17, 23, 21)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 24, 31, 21, 17, 12)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(23, 19, 21, 25)]}

if(fname == "Family.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(15, 21, 19)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(24, 23, 15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(17, 21)]}

if(fname == "Genus.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(14, 11)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 31, 11)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(31, 2, 4, 11)]}

if(fname == "Genus.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21, 22, 19, 8, 2)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21, 22, 17, 3, 2, 15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(19, 24, 17, 23, 22)]}

##AIC, BIC
if(Smeth == "AIC" | Smeth == "BIC"){
  findbest5sub<-regsubsets(axis ~ ., data=NMDSpreRCAL[[k]], nbest=5, nvmax=ncol(preRCAL))
  bestsubpar[[k]]<-summary(findbest5sub)$which
  bestsubrssl[1:(5*(ncol(NMDSpreRCAL[[k]])-2)+1), k]<-summary(findbest5sub)$rss}

##StepwiseAIC
if(Smeth == "StepwiseAIC"){
  names(NMDSpreRCAL[[k]])[1] <- "axis" # rename the NMDSpreRCAL data frame columns
  axislm <- lm(axis ~ ., data= NMDSpreRCAL[[k]])
  stepaxis[[k]]<- stepAIC(axislm, direction="both")
  write.table(coef(stepaxis[[k]]), file=paste(plotname, "final.txt"), append=TRUE)}
}

if(nn == 1 | nn==2){

mo<-vector()
for(i in 1:(ncol(preRCAL)-1)){
  mo<-c(mo, rep(i, 5))}
mo<-c(mo, ncol(preRCAL))
mo<-matrix(data=rep(mo, dn), nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)
parAIC<-nrow(preRCAL)*log(bestsubrssl/nrow(preRCAL))+(2*nrow(preRCAL)*(mo+2))/(nrow(preRCAL)-(mo+2)-1)
parBIC<-nrow(preRCAL)*log(bestsubrssl/nrow(preRCAL))+(mo+2)*log(nrow(preRCAL))

minmo<-matrix(data=0, nrow=2, ncol=dn)
minmorow<-matrix(data=0, nrow=2, ncol=dn)
for(k in 1:dn){
  for(i in 1:(5*(ncol(preRCAL)-1)+1)){
    if(parAIC[i, k] == min(parAIC[, k])) {minmo[1, k]<-mo[i, k]; minmorow[1, k]<-i}
    if(parBIC[i, k] == min(parBIC[, k])) {minmo[2, k]<-mo[i, k]; minmorow[2, k]<-i}
  }
}
}

```

```

}

subpre<-vector("list", dn)
axislm<-vector("list", dn)

      i<-minmrow[nn, 1]
      ii<-minmrow[nn, 2]
      iii<-minmrow[nn, 3]
      subpre[[1]]<-vector()
      for(j in 2:ncol(bestsubpar[[1]])) {
        if (bestsubpar[[1]][i, j] == TRUE) subpre[[1]] <-
c(subpre[[1]], names(NMDSpreRCAL[[1]])[j])

      subpre[[2]]<-vector()
      for(j in 2:ncol(bestsubpar[[2]])) {
        if (bestsubpar[[2]][ii, j] == TRUE) subpre[[2]] <-
c(subpre[[2]], names(NMDSpreRCAL[[2]])[j])

      subpre[[3]]<-vector()
      for(j in 2:ncol(bestsubpar[[3]])) {
        if (bestsubpar[[3]][iii, j] == TRUE) subpre[[3]] <-
c(subpre[[3]], names(NMDSpreRCAL[[3]])[j])

      write.table(paste(plotname, "Axis Coefficients"), file=paste(plotname, "final.txt"),
append=TRUE)
      for (k in 1:dn) {
        subsetpre <- NMDSpreRCAL[[k]][c("axis", subpre[[k]])]
        axislm[[k]]<- lm(axis ~ ., data=subsetpre)
        write.table(coef(axislm[[k]]), file=paste(plotname, "final.txt"), append=TRUE) } }

if(Smeth == "StepwiseAIC") axislm<-stepaxis

postscript(paste(plotname, "final.ps"))

### Internal Validation (Calibration
Data)=====
      ## calculate the distance between test sites and reference sites
      sdi j <- d. i j(dn, axislm, preRCAL, preRCAL)

      ## calculate the Pc for the ROI null model
      p <- ANNA. prob(neino, sdi j, RCAL)
      OE. results <- O. E(RCAL, p)
      O <- OE. results$O
      E <- OE. results$E
      OEsum <- OE. results$results
      write.table("Calibration (Internal)", file=paste(plotname, "final.txt"), append=TRUE)

      write.table(O, file=paste(plotname, "final.txt"), append=TRUE)
      write.table(E, file=paste(plotname, "final.txt"), append=TRUE)
      write.table(OEsum, file=paste(plotname, "final.txt"), append=TRUE)

      #draw the graphs
      op<-par(mfrow=c(4, 3), mar=c(4, 4, 5, 1))
      plot. main<-paste(plotname, "\nNeighbor number =", neino, "\nInternal")
      OE. plot(O, E, OEsum, plot. main)

### External Validation (Validation
Data)=====
      sdi j <- d. i j(dn, axislm, preRCAL, preRVAL)
      p <- ANNA. prob(neino, sdi j, RCAL)

```

```

OE.results <- O.E(RVAL, p)
O <- OE.results$O
E <- OE.results$E
OESum <- OE.results$results
write.table("Validation (External)", file=paste(plotname, "final.txt"), append=TRUE)

write.table(O, file=paste(plotname, "final.txt"), append=TRUE)
write.table(E, file=paste(plotname, "final.txt"), append=TRUE)
write.table(OESum, file=paste(plotname, "final.txt"), append=TRUE)

plot.main<-c("external")
par(mar=c(4, 4, 2, 1))
OE.plot(O, E, OESum, plot.main)

dev.off()

```

BROI Select Parameters with RMSE

```

##-- save current workspace

# input files & plot name
mname<- "ROI option1"
fname <- paste("Genus.Allpredictors")
bugdata <- read.delim("bugallgenus.txt")
pre <- read.delim("AllPredictors.txt")

# load library & R code & workspace

library(labdsv)
source("d.ij.r")
source("O.E.r")
source("OE.plot.r")
source("stepwise.AIC.r")
source("ex.AICBIC.r")
source("ROIopl.prob.r")
source("result.RMSEsort.r")

#omit the NA
merg <- merge(bugdata, pre, by = intersect(names(bugdata), names(pre)))
om.merg <- na.omit(merg, subset=TRUE)
bugdata <- subset(om.merg, select = names(bugdata))
pre <- subset(om.merg, select = names(pre))

#transform predictor variables
# for 11/09/2007 new transform, mixed with dr. Paul and mine
#attach(pre)
#pre <- transform(pre, "log10.AREA.1" = log10(AREA+1))
#pre <- transform(pre, "sqrt.WSLOPE" = sqrt(WSLOPE))
#pre <- transform(pre, "log10.CSLOPE.1" = log10(CSLOPE+1))
#pre <- transform(pre, "log10.RELIEF.1" = log10(RELIEF+1))
#pre <- transform(pre, "log10.PERIM.1" = log10(PERIM+1))
#pre <- transform(pre, "log10.LENGTH.1" = log10(LENGTH+1))
#pre <- transform(pre, "asinh.sqrt.RSA" = asinh(sqrt(RSA)))
#pre <- transform(pre, "asinh.sqrt.RSD" = asinh(sqrt(RSD)))
#pre <- transform(pre, "log10.ELEV.1" = log10(ELEV+1))
#pre <- transform(pre, "asinh.sqrt.HIGHELEV" = asinh(sqrt(HIGHELEV)))
#if((fname == "Family.GISHydropredictors") | (fname == "Genus.GISHydropredictors")){

```



```

#       pre <- data.frame(pre[-c(3, 4, 5, 6, 10, 11, 12, 15, 16, 20)])}
#if((fname == "Family.Allpredictors") | (fname == "Genus.Allpredictors")){
#       pre <- transform(pre, "log10.MAXDEPTH" = log10(MAXDEPTH))
#       pre <- transform(pre, "log10.AVGWID" = log10(AVGWID))
#       pre <- data.frame(pre[-c(5, 6, 7, 8, 12, 13, 14, 17, 18, 22, 29, 30)])}
#detach(pre)

#extract reference sites bio-data and physical parameters
RCAL<-subset(bugdata, REF == "R")[, 3:ncol(bugdata)]
RVAL<-subset(bugdata, REF == "V")[, 3:ncol(bugdata)]
preRCAL<-subset(pre, REF == "R")[, 3:ncol(pre)]
preRVAL<-subset(pre, REF == "V")[, 3:ncol(pre)]

#eliminate two outlier in the validation data
RVAL<-RVAL[-c(10, 19), ]
preRVAL<-preRVAL[-c(10, 19), ]

#Bray-Curtis distance matrix
RCAL.bc <- dsvdis(RCAL, index="bray/curtis")

#Non-metric dimensional scaling
stress <- 100
dn <- 1
while(stress > 25){
  dn <- dn+1
  RCAL.nmnds <- isoMDS(RCAL.bc, k=dn)
  stress <- as.numeric(RCAL.nmnds[2])}

NMDS<- data.frame(RCAL.nmnds[1])

NMDSpreRCAL <- list(1:dn)

for(k in 1:dn){
  NMDSpreRCAL[[k]] <- data.frame(NMDS[k], preRCAL)
  names(NMDSpreRCAL[[k]]) <- c("axis", names(preRCAL)) # rename the NMDSpreRCAL data frame columns

# After Rational Checking
if(fname == "Family.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(11, 9, 17, 23, 21)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 24, 31, 21, 17, 12)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(23, 19, 21, 25)]}

if(fname == "Family.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(15, 21, 19)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(24, 23, 15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(17, 21)]}

if(fname == "Genus.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(14, 11)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 31, 11)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(31, 2, 4, 11)]}

if(fname == "Genus.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21, 22, 19, 8, 2)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21, 22, 17, 3, 2, 15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(19, 24, 17, 23, 22)]}
}

# different selecting parameter method

```

```

# nn=1:Exhaustive AIC, nn=2:Exhaustive BIC, nn=3:StepwiseAIC

for(nn in 1:3){

# pick the selecting parameter method
if(nn == 1) {
  plotname<-paste(fname,mname,"best AIC")
  axislm<-ex.AICBIC(nn, dn, NMDSpreRCAL)}
if(nn == 2) {
  plotname<-paste(fname,mname,"best BIC")
  axislm<-ex.AICBIC(nn, dn, NMDSpreRCAL)}
if(nn == 3) {
  plotname <- paste(fname,mname,"stepwiseAIC")
  axislm<-stepwise.AIC(dn, NMDSpreRCAL)}
write.table(plotname,file=paste(plotname, ". txt"), append=TRUE)
for(k in 1:dn) write.table(coef(axislm[[k]]), file=paste(plotname, ". txt"), append=TRUE)

  ##check different n, NST, ThetaU combination

  m<-0
  results.int<-matrix(data=0, nr=1, nc=20)
  results.int<-as.data.frame(results.int)
  names(results.int)<-c("ThetaL", "ThetaU", "TP", "NST", "n",
    "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",
    "Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",
    "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2")

  for(n in seq(2, 3.5, 0.5)) {
    for(NST in 8:15) {
      for(ThetaL in seq(5, 15, 5)) {
        for(ThetaU in c(25, 30, 50)) {
          for(TP in c(50, 75, 85)) {

            m<-m+1
            ### Internal Validation (Calibration
Data)=====
            ## calculate the distance between test sites and reference sites
            sdi j <- d. ij(dn, axislm, preRCAL, preRCAL)

            TL<-quantile(sdi j, p=ThetaL/100)
            TU<-quantile(sdi j, p=ThetaU/100)
            tp<-quantile(sdi j, p=TP/100)

            ## calculate the Pc for the ROI null model
            p <- ROIop1. prob(n, NST, TL, TU, tp, sdi j, RCAL)
            OE. results <- O. E(RCAL, p)

            OEsu m <- OE. results$results
            results.int[m, ]<-c(ThetaL, ThetaU, TP, NST, n, OEsu m)
            if(m%100==0) print(m)
          }}}
        }
      }
    }

    ## sort the results depends on
    ## Pc>0.5's mean, sd, slope, intercept, and r^2 geometric mean
    # results of internal
    sort.int<-result.RMSEsort(results.int)
    write.table("internal sort results", file=paste(plotname, ". txt"), append=TRUE)
    write.table(sort.int[1:10, ], file=paste(plotname, ". txt"), append=TRUE)
    save.image(paste(plotname, ". RData"))

```

```

m<-0
results.ext<-matrix(data=0,nr=1,nc=20)
results.ext<-as.data.frame(results.ext)
names(results.ext)<-c("ThetaL","ThetaU","TP","NST","n",
                    "Pc0.mean","Pc0.sd","Pc0.slope","Pc0.intercept","Pc0.r2",
                    "Pc0.01.mean","Pc0.01.sd","Pc0.01.slope","Pc0.01.intercept","Pc0.01.r2",
                    "Pc0.5.mean","Pc0.5.sd","Pc0.5.slope","Pc0.5.intercept","Pc0.5.r2")

## draw the top 10 combination graphs
for(top in 1:10) {

  ThetaL<-sort.int$ThetaL[top]
  ThetaU<-sort.int$ThetaU[top]
  TP <-sort.int$TP[top]
  NST<-sort.int$NST[top]
  n<-sort.int$n[top]

  postscript(paste(plotname,"top",top,".ps"))

  ### Internal Validation (Calibration
Data)=====
  ## calculate the distance between test sites and reference sites
  sdi.j <- d.i.j(dn,axislm,preRCAL,preRCAL)

  TL<-quantile(sdi.j,p=ThetaL/100)
  TU<-quantile(sdi.j,p=ThetaU/100)
  tp<-quantile(sdi.j,p=TP/100)

  ## calculate the Pc for the ROI null model
  p <- ROIopl.prob(n,NST,TL,TU,tp,sdi.j,RCAL)
  OE.results <- O.E(RCAL,p)

  #draw the graphs
  op<-par(mfrow=c(4,3),mar=c(4,4,4,1))
  O <- OE.results$O
  E <- OE.results$E
  OEsu.m <- OE.results$results
  plot.main<-
paste(plotname,"\nm=",n,"NST=",NST,"ThetaL=",ThetaL,"ThetaU=",ThetaU,"TP=",TP,"\ninternal")
  OE.plot(O,E,OEsu.m,plot.main)

  ### External Validation (Validation
Data)=====
  m<-m+1
  sdi.j <- d.i.j(dn,axislm,preRCAL,preRVAL)
  p <- ROIopl.prob(n,NST,TL,TU,tp,sdi.j,RCAL)
  OE.results <- O.E(RVAL,p)
  O <- OE.results$O
  E <- OE.results$E
  OEsu.m <- OE.results$results
  results.ext[m,]<-c(ThetaL,ThetaU,TP,NST,n,OEsu.m)

  plot.main<-c("external")
  par(mar=c(4,4,2,1))
  OE.plot(O,E,OEsu.m,plot.main)

  dev.off()

```

```

    }

    write.table("external results", file=paste(plotname, ".txt"), append=TRUE)
    write.table(results.ext, file=paste(plotname, ".txt"), append=TRUE)
    ## sort the results depends on
    ## Pc>0.5's mean, sd, slope, intercept, and r^2 geometric mean
    # results of external
    write.table("external results", file=paste(plotname, ".txt"), append=TRUE)
    sort.ext<-result.RMSEsort(results.ext)
    write.table(sort.ext, file=paste(plotname, ".txt"), append=TRUE)
    save.image(paste(plotname, ".RData"))

}

```

BROI Model

```

##-- save current workspace

#input files & plot name
fname <- "Genus.GISHydropredictors"
bugdata <- read.delim("bugallgenus.txt")
pre <- read.delim("GISHydroPredictors.txt")
Smeth <- "StepwiseAIC"
ThetaL<-10
ThetaU<-25
TP<-50
NST<-8
n<-2.5

if(Smeth == "AIC") {nn<-1
  plotname<-paste(fname, "ROI option 1 best AIC")}
if(Smeth == "BIC") {nn<-2
  plotname<-paste(fname, "ROI option 1 best BIC")}
if(Smeth == "StepwiseAIC") {nn<-3
  plotname <- paste(fname, "ROI option 1 stepwiseAIC")}

# load library & R code & workspace

library(labdsv)
library(leaps)
source("d.ij.r")
source("ROIop1.prob.r")
source("O.E.r")
source("OE.plot.r")
source("result.5sort.r")

#omit the NA
merg <- merge(bugdata, pre, by = intersect(names(bugdata), names(pre)))
om.merg <- na.omit(merg, subset=TRUE)
bugdata <- subset(om.merg, select = names(bugdata))
pre <- subset(om.merg, select = names(pre))

#extract reference sites bio-data and physical parameters
RCAL<-subset(bugdata, REF == "R")[, 3:ncol(bugdata)]
RVAL<-subset(bugdata, REF == "V")[, 3:ncol(bugdata)]
preRCAL<-subset(pre, REF == "R")[, 3:ncol(pre)]
preRVAL<-subset(pre, REF == "V")[, 3:ncol(pre)]

# modify RVAL data, pick out 2 point

```

```

RVAL<-RVAL[-c(10,19),]
preRVAL<-preRVAL[-c(10,19),]

#Bray-Curtis distance matrix
RCAL.bc <- dsvdis(RCAL, index="bray/curtis")

#Non-metric dimensional scaling
stress <- 100
dn <- 1
while(stress > 25){
  dn <- dn+1
  RCAL.nmnds <- isoMDS(RCAL.bc, k=dn)
  stress <- as.numeric(RCAL.nmnds[2])}

NMDS<- data.frame(RCAL.nmnds[1])

NMDSpreRCAL <- list(1:dn)
stepaxis <- list(1:dn)
bestsubpar <- list(1:dn)
bestsubrsts <-matrix(data=100,nrow=(5*(ncol(preRCAL)-1)+1),ncol=dn)

for(k in 1:dn){
  NMDSpreRCAL[[k]] <- data.frame(NMDS[k], preRCAL)
  names(NMDSpreRCAL[[k]]) <- c("axis", names(preRCAL)) # rename the NMDSpreRCAL data frame columns

# After Rational Checking
if(fname == "Family.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(11,9,17,23,21)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30,24,31,21,17,12)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(23,19,21,25)]}

if(fname == "Family.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(15,21,19)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(24,23,15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(17,21)]}

if(fname == "Genus.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(14,11)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30,31,11)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(31,2,4,11)]}

if(fname == "Genus.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21,22,19,8,2)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21,22,17,3,2,15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(19,24,17,23,22)]}

##AIC, BIC
if(Smeth == "AIC" |Smeth == "BIC"){
  findbest5sub<-regsubsets(axis~., data=NMDSpreRCAL[[k]], nbest=5, nvmax=ncol(preRCAL))
  bestsubpar[[k]]<-summary(findbest5sub)$which
  bestsubrsts[1:(5*(ncol(NMDSpreRCAL[[k]])-2)+1),k]<-summary(findbest5sub)$rsts}

##StepwiseAIC
if(Smeth == "StepwiseAIC"){
  names(NMDSpreRCAL[[k]])[1] <- "axis" # rename the NMDSpreRCAL data frame columns
  axislm <- lm(axis ~ ., data= NMDSpreRCAL[[k]])
  stepaxis[[k]]<- stepAIC(axislm, direction="both")
  write.table(coef(stepaxis[[k]]), file=paste(plotname,"final.txt"), append=TRUE)}
}

```

```

if (nn == 1 | nn==2) {

mo<-vector()
for(i in 1:(ncol(preRCAL)-1)) {
  mo<-c(mo, rep(i, 5))}
mo<-c(mo, ncol(preRCAL))
mo<-matrix(data=rep(mo, dn), nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)
parAIC<-nrow(preRCAL)*log(bestsubrss/nrow(preRCAL))+(2*nrow(preRCAL)*(mo+2))/(nrow(preRCAL)-(mo+2)-1)
parBIC<-nrow(preRCAL)*log(bestsubrss/nrow(preRCAL))+(mo+2)*log(nrow(preRCAL))

minmo<-matrix(data=0, nrow=2, ncol=dn)
minmorow<-matrix(data=0, nrow=2, ncol=dn)
for(k in 1:dn) {
  for(i in 1:(5*(ncol(preRCAL)-1)+1)) {
    if(parAIC[i, k] == min(parAIC[, k])) {minmo[1, k]<-mo[i, k]; minmorow[1, k]<-i}
    if(parBIC[i, k] == min(parBIC[, k])) {minmo[2, k]<-mo[i, k]; minmorow[2, k]<-i}
  }
}

subpre<-vector("list", dn)
axislm<-vector("list", dn)

  i<-minmorow[nn, 1]
  ii<-minmorow[nn, 2]
  iii<-minmorow[nn, 3]
  subpre[[1]]<-vector()
  for(j in 2:ncol(bestsubpar[[1]])) {
    if (bestsubpar[[1]][i, j] == TRUE) subpre[[1]] <-
c(subpre[[1]], names(NMDSpreRCAL[[1]][j]))}

  subpre[[2]]<-vector()
  for(j in 2:ncol(bestsubpar[[2]])) {
    if (bestsubpar[[2]][ii, j] == TRUE) subpre[[2]] <-
c(subpre[[2]], names(NMDSpreRCAL[[2]][j]))}

  subpre[[3]]<-vector()
  for(j in 2:ncol(bestsubpar[[3]])) {
    if (bestsubpar[[3]][iii, j] == TRUE) subpre[[3]] <-
c(subpre[[3]], names(NMDSpreRCAL[[3]][j]))}

  write.table(paste(plotname, "Axis Coefficients"), file=paste(plotname, "final.txt"),
append=TRUE)

  for (k in 1:dn) {
subsetpre <- NMDSpreRCAL[[k]][c("axis", subpre[[k]])]
axislm[[k]]<- lm(axis ~ ., data=subsetpre)
write.table(coef(axislm[[k]]), file=paste(plotname, "final.txt"), append=TRUE) } }

if(Smeth == "StepwiseAIC") axislm <- stepaxis

postscript(paste(plotname, "final.ps"))

### Internal Validation (Calibration
Data)=====
## calculate the distance between test sites and reference sites
sdi j <- d. ij(dn, axislm, preRCAL, preRCAL)

TL<-quantile(sdi j, p=ThetaL/100)
TU<-quantile(sdi j, p=ThetaU/100)

```

```

tp<-quantile(sdi,j,p=TP/100)

## calculate the Pc for the ROI null model
p <- ROIop1.prob(n, NST, TL, TU, tp, sdi,j, RCAL)
OE.results <- O.E(RCAL, p)
O <- OE.results$O
E <- OE.results$E
OEsun <- OE.results$results
write.table("Calibration (Internal)", file=paste(plotname, "final.txt"), append=TRUE)

write.table(O, file=paste(plotname, "final.txt"), append=TRUE)
write.table(E, file=paste(plotname, "final.txt"), append=TRUE)
write.table(OEsun, file=paste(plotname, "final.txt"), append=TRUE)

#draw the graphs
op<-par(mfrow=c(4, 3), mar=c(4, 4, 4, 1))
plot.main<-
paste(plotname, "\nn=", n, "NST=", NST, "ThetaL=", ThetaL, "ThetaU=", ThetaU, "TP=", TP, "\ninternal")
OE.plot(O, E, OEsun, plot.main)

### External Validation (Validation
Data)=====
sdi,j <- d.i.j(dn, axislm, preRCAL, preRVAL)
p <- ROIop1.prob(n, NST, TL, TU, tp, sdi,j, RCAL)
OE.results <- O.E(RVAL, p)
O <- OE.results$O
E <- OE.results$E
OEsun <- OE.results$results
write.table("Validation (External)", file=paste(plotname, "final.txt"), append=TRUE)

write.table(O, file=paste(plotname, "final.txt"), append=TRUE)
write.table(E, file=paste(plotname, "final.txt"), append=TRUE)
write.table(OEsun, file=paste(plotname, "final.txt"), append=TRUE)

plot.main<-c("external")
par(mar=c(4, 4, 2, 1))
OE.plot(O, E, OEsun, plot.main)

dev.off()

```

NROI Select Parameters with RMSE

```

##-- save current workspace

# input files & plot name
mname<- "newROI"
fname <- paste("Genus.Allpredictors")
bugdata <- read.delim("bugallgenus.txt")
pre <- read.delim("AllPredictors.txt")

# load library & R code & workspace

library(labdsv)
source("d.i.j.r")
source("O.E.r")
source("OE.plot.r")
source("stepwise.AIC.r")
source("ex.AICBIC.r")

```

```

source("ROI_prob.r")
source("result_RMSEsortn.r")

#omit the NA
merg <- merge(bugdata, pre, by = intersect(names(bugdata), names(pre)))
om.merg <- na.omit(merg, subset=TRUE)
bugdata <- subset(om.merg, select = names(bugdata))
pre <- subset(om.merg, select = names(pre))

#transform predictor variables
# for 11/09/2007 new transform, mixed with dr. Paul and mine
#attach(pre)
#pre <- transform(pre, "log10.AREA.1" = log10(AREA+1))
#pre <- transform(pre, "sqrt.WSLOPE" = sqrt(WSLOPE))
#pre <- transform(pre, "log10.CSLOPE.1" = log10(CSLOPE+1))
#pre <- transform(pre, "log10.RELIEF.1" = log10(RELIEF+1))
#pre <- transform(pre, "log10.PERIM.1" = log10(PERIM+1))
#pre <- transform(pre, "log10.LENGTH.1" = log10(LENGTH+1))
#pre <- transform(pre, "asinh.sqrt.RSA" = asinh(sqrt(RSA)))
#pre <- transform(pre, "asinh.sqrt.RSD" = asinh(sqrt(RSD)))
#pre <- transform(pre, "log10.ELEV.1" = log10(ELEV+1))
#pre <- transform(pre, "asinh.sqrt.HIGHELEV" = asinh(sqrt(HIGHELEV)))
#if((fname == "Family.GISHydropredictors") | (fname == "Genus.GISHydropredictors")){
#   pre <- data.frame(pre[-c(3, 4, 5, 6, 10, 11, 12, 15, 16, 20)])}
#if((fname == "Family.Allpredictors") | (fname == "Genus.Allpredictors")){
#   pre <- transform(pre, "log10.MAXDEPTH" = log10(MAXDEPTH))
#   pre <- transform(pre, "log10.AVGWID" = log10(AVGWID))
#   pre <- data.frame(pre[-c(5, 6, 7, 8, 12, 13, 14, 17, 18, 22, 29, 30)])}
#detach(pre)

#extract reference sites bio-data and physical parameters
RCAL<-subset(bugdata, REF == "R")[, 3:ncol(bugdata)]
RVAL<-subset(bugdata, REF == "V")[, 3:ncol(bugdata)]
preRCAL<-subset(pre, REF == "R")[, 3:ncol(pre)]
preRVAL<-subset(pre, REF == "V")[, 3:ncol(pre)]

#eliminate two outlier in the validation data
RVAL<-RVAL[-c(10, 19),]
preRVAL<-preRVAL[-c(10, 19),]

#Bray-Curtis distance matrix
RCAL.bc <- dsvdis(RCAL, index="bray/curtis")

#Non-metric dimensional scaling
stress <- 100
dn <- 1
while(stress > 25){
  dn <- dn+1
  RCAL.nmds <- isoMDS(RCAL.bc, k=dn)
  stress <- as.numeric(RCAL.nmds[2])}

NMDS<- data.frame(RCAL.nmds[1])

NMDSpreRCAL <- list(1:dn)

for(k in 1:dn){
  NMDSpreRCAL[[k]] <- data.frame(NMDS[k], preRCAL)
  names(NMDSpreRCAL[[k]]) <- c("axis", names(preRCAL)) # rename the NMDSpreRCAL data frame columns

# After Rational Checking
if(fname == "Family.Allpredictors"){

```



```

        if(k == 1) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(11, 9, 17, 23, 21)]
        if(k == 2) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(30, 24, 31, 21, 17, 12)]
        if(k == 3) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(23, 19, 21, 25)]

if(fname == "Family. GISHydropredictors") {
  if(k == 1) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(15, 21, 19)]
  if(k == 2) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(24, 23, 15)]
  if(k == 3) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(17, 21)]

if(fname == "Genus. Allpredictors") {
  if(k == 1) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(14, 11)]
  if(k == 2) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(30, 31, 11)]
  if(k == 3) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(31, 2, 4, 11)]

if(fname == "Genus. GISHydropredictors") {
  if(k == 1) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(21, 22, 19, 8, 2)]
  if(k == 2) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(21, 22, 17, 3, 2, 15)]
  if(k == 3) NMDSPreRCAL[[k]]<-NMDSPreRCAL[[k]][-c(19, 24, 17, 23, 22)]

}

# different selecting parameter method
# nn=1:Exhaustive AIC, nn=2:Exhaustive BIC, nn=3:StepwiseAIC

for(nn in 1:3){

# pick the selecting parameter method
if(nn == 1) {
  plotname<-paste(fname, mname, "best AIC")
  axislm<-ex. AICBIC(nn, dn, NMDSPreRCAL)}
if(nn == 2) {
  plotname<-paste(fname, mname, "best BIC")
  axislm<-ex. AICBIC(nn, dn, NMDSPreRCAL)}
if(nn == 3) {
  plotname <- paste(fname, mname, "stepwiseAIC")
  axislm<-stepwise. AIC(dn, NMDSPreRCAL)}
write.table(plotname, file=paste(plotname, ". txt"), append=TRUE)
for(k in 1:dn) write.table(coef(axislm[[k]]), file=paste(plotname, ". txt"), append=TRUE)

##check different n, NST, ThetaU combination
m<-0
results.int<-matrix(data=0, nr=1, nc=19)
results.int<-as.data.frame(results.int)
names(results.int)<-c("ThetaL", "ThetaU", "NST", "n",
  "Pc0. mean", "Pc0. sd", "Pc0. slope", "Pc0. intercept", "Pc0. r2",
  "Pc0. 01. mean", "Pc0. 01. sd", "Pc0. 01. slope", "Pc0. 01. intercept", "Pc0. 01. r2",
  "Pc0. 5. mean", "Pc0. 5. sd", "Pc0. 5. slope", "Pc0. 5. intercept", "Pc0. 5. r2")

for(n in seq(2, 4, 0.5)) {
  for(NST in 4:8) {
    for(ThetaL in seq(4, 10, 2)) {
      for(ThetaU in seq(30, 90, 10)) {

          m<-m+1
          ### Internal Validation (Calibration
Data)=====
          ## calculate the distance between test sites and reference sites
          sdi.j <- d. i.j(dn, axislm, preRCAL, preRCAL)
          TL<-quantile(sdi.j, p=ThetaL/100)

```

```

TU<-quantile(sdi j,p=ThetaU/100)

## calculate the Pc for the ROI null model
p <- ROI. prob(n, NST, TL, TU, sdi j, RCAL)
OE. results <- O. E(RCAL, p)

OEsum <- OE. results$results
results. int[m, ]<-c(ThetaL, ThetaU, NST, n, OEsum)
if( (m%%100) == 0) print(m)
}}}}

## sort the results depends on
## Pc>0. 5' s mean, sd, slope, intercept, and r^2 geometric mean
# results of internal
sort. int<-result. RMSEsortn(results. int)
write. table("internal sort results", file=paste(plotname, ". txt"), append=TRUE)
write. table(sort. int[1:10, ], file=paste(plotname, ". txt"), append=TRUE)
save. image(paste(plotname, ". RData"))

m<-0
results. ext<-matrix(data=0, nr=1, nc=19)
results. ext<-as. data. frame(results. ext)
names(results. ext)<-c("ThetaL", "ThetaU", "NST", "n",
"Pc0. mean", "Pc0. sd", "Pc0. slope", "Pc0. intercept", "Pc0. r2",
"Pc0. 01. mean", "Pc0. 01. sd", "Pc0. 01. slope", "Pc0. 01. intercept", "Pc0. 01. r2",
"Pc0. 5. mean", "Pc0. 5. sd", "Pc0. 5. slope", "Pc0. 5. intercept", "Pc0. 5. r2")

## draw the top 10 combination graphs
for(top in 1:10){
ThetaL<-sort. int$ThetaL[top]
ThetaU<-sort. int$ThetaU[top]
NST<-sort. int$NST[top]
n<-sort. int$n[top]

postscript(paste(plotname, "top", top, ". ps"))

### Internal Validation (Calibration
Data)=====
## calculate the distance between test sites and reference sites
sdi j <- d. ij(dn, axislm, preRCAL, preRCAL)
TL<-quantile(sdi j, p=ThetaL/100)
TU<-quantile(sdi j, p=ThetaU/100)

## calculate the Pc for the ROI null model
p <- ROI. prob(n, NST, TL, TU, sdi j, RCAL)
OE. results <- O. E(RCAL, p)

#draw the graphs
op<-par(mfrow=c(4, 3), mar=c(4, 4, 5, 1))
O <- OE. results$O
E <- OE. results$E
OEsum <- OE. results$results
plot. main<-paste(plotname, "\nn=", n, "NST=", NST, "ThetaL=", ThetaL, "%
ThetaU=", ThetaU, "%", "\nInternal")
OE. plot(O, E, OEsum, plot. main)

### External Validation (Validation
Data)=====
m<-m+1
sdi j <- d. ij(dn, axislm, preRCAL, preRVAL)
p <- ROI. prob(n, NST, TL, TU, sdi j, RCAL)

```

```

        OE.results <- O.E(RVAL, p)
        O <- OE.results$O
        E <- OE.results$E
        OEsum <- OE.results$results
        results.ext[m, ]<-c(ThetaL, ThetaU, NST, n, OEsum)

        plot.main<-c("external")
        par(mar=c(4, 4, 2, 1))
        OE.plot(O, E, OEsum, plot.main)

        dev.off()
    }

    write.table("external results", file=paste(plotname, ".txt"), append=TRUE)
    write.table(results.ext, file=paste(plotname, ".txt"), append=TRUE)
    ## sort the results depends on
    ## Pc>0.5's mean, sd, slope, intercept, and r^2 geometric mean
    # results of external
    write.table("external sort results", file=paste(plotname, ".txt"), append=TRUE)
    sort.ext<-result.RMSEsortn(results.ext)
    write.table(sort.ext, file=paste(plotname, ".txt"), append=TRUE)
    save.image(paste(plotname, ".RData"))
}

```

NROI Model

```

##-- save current workspace

#input files & plot name
fname <- "Genus.GISHydropredictors"
bugdata <- read.delim("bugallgenus.txt")
pre <- read.delim("GISHydroPredictors.txt")
Smeth <- "StepwiseAIC"
ThetaL<- 4
ThetaU<- 30
NST <- 8
n <- 3.5

if(Smeth == "AIC") {nn<-1
                    plotname<-paste(fname, "ROI new threshold best AIC")}
if(Smeth == "BIC") {nn<-2
                    plotname<-paste(fname, "ROI new threshold best BIC")}
if(Smeth == "StepwiseAIC") {nn<-3
                             plotname <- paste(fname, "ROI new threshold stepwiseAIC")}

# load library & R code & workspace

library(labdsv)
library(leaps)
source("d.ij.r")
source("ROI.prob.r")
source("O.E.r")
source("OE.plot.r")
source("result.5sortn.r")

#omit the NA
merg <- merge(bugdata, pre, by = intersect(names(bugdata), names(pre)))
om.merg <- na.omit(merg, subset=TRUE)
bugdata <- subset(om.merg, select = names(bugdata))

```

```

pre <- subset(om.merg, select = names(pre))

#extract reference sites bio-data and physical parameters
RCAL<-subset(bugdata, REF == "R")[,3:ncol(bugdata)]
RVAL<-subset(bugdata, REF == "V")[,3:ncol(bugdata)]
preRCAL<-subset(pre, REF == "R")[,3:ncol(pre)]
preRVAL<-subset(pre, REF == "V")[,3:ncol(pre)]

# modify RVAL data, pick out 2 point
RVAL<-RVAL[-c(10,19),]
preRVAL<-preRVAL[-c(10,19),]

#Bray-Curtis distance matrix
RCAL.bc <- dsvdis(RCAL, index="bray/curtis")

#Non-metric dimensional scaling
stress <- 100
dn <- 1
while(stress > 25){
  dn <- dn+1
  RCAL.nmnds <- isoMDS(RCAL.bc, k=dn)
  stress <- as.numeric(RCAL.nmnds[2])}

NMDS<- data.frame(RCAL.nmnds[1])

NMDSpreRCAL <- list(1:dn)
stepaxis <- list(1:dn)
bestsubpar <- list(1:dn)
bestsubrss <-matrix(data=100,nrow=(5*(ncol(preRCAL)-1)+1),ncol=dn)

for(k in 1:dn){
  NMDSpreRCAL[[k]] <- data.frame(NMDS[k], preRCAL)
  names(NMDSpreRCAL[[k]]) <- c("axis", names(preRCAL)) # rename the NMDSpreRCAL data frame columns

# After Rational Checking
if(fname == "Family.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(11, 9, 17, 23, 21)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 24, 31, 21, 17, 12)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(23, 19, 21, 25)]}

if(fname == "Family.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(15, 21, 19)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(24, 23, 15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(17, 21)]}

if(fname == "Genus.Allpredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(14, 11)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(30, 31, 11)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(31, 2, 4, 11)]}

if(fname == "Genus.GISHydropredictors"){
  if(k == 1) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21, 22, 19, 8, 2)]
  if(k == 2) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(21, 22, 17, 3, 2, 15)]
  if(k == 3) NMDSpreRCAL[[k]]<-NMDSpreRCAL[[k]][-c(19, 24, 17, 23, 22)]}

##AIC, BIC
if(Smeth == "AIC" | Smeth == "BIC"){
  findbest5sub<-regsubsets(axis~., data=NMDSpreRCAL[[k]], nbest=5, nvmax=ncol(preRCAL))
  bestsubpar[[k]]<-summary(findbest5sub)$which

```

```

bestsubrssl[1:(5*(ncol(NMDSpreRCAL[[k]])-2)+1),k]<-summary(findbest5sub)$rss}

##StepwiseAIC
if(Smeth == "StepwiseAIC") {
names(NMDSpreRCAL[[k]][1] <-"axis" # rename the NMDSpreRCAL data frame columns
axislm <- lm(axis ~ ., data= NMDSpreRCAL[[k]])
stepaxis[[k]]<- stepAIC(axislm, direction="both")
write.table(coef(stepaxis[[k]]), file=paste(plotname,"final.txt"), append=TRUE)}
}

if(nn == 1 | nn==2){

mo<-vector()
for(i in 1:(ncol(preRCAL)-1)) {
mo<-c(mo, rep(i, 5))}
mo<-c(mo, ncol(preRCAL))
mo<-matrix(data=rep(mo, dn), nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)
parAIC<-nrow(preRCAL)*log(bestsubrssl/nrow(preRCAL))+(2*nrow(preRCAL)*(mo+2))/(nrow(preRCAL)-(mo+2)-1)
parBIC<-nrow(preRCAL)*log(bestsubrssl/nrow(preRCAL))+(mo+2)*log(nrow(preRCAL))

minmo<-matrix(data=0, nrow=2, ncol=dn)
minmorow<-matrix(data=0, nrow=2, ncol=dn)
for(k in 1:dn){
for(i in 1:(5*(ncol(preRCAL)-1)+1)) {
if(parAIC[i, k] == min(parAIC[, k])) {minmo[1, k]<-mo[i, k];minmorow[1, k]<-i}
if(parBIC[i, k] == min(parBIC[, k])) {minmo[2, k]<-mo[i, k];minmorow[2, k]<-i}
}

subpre<-vector("list", dn)
axislm<-vector("list", dn)

i<-minmorow[nn, 1]
ii<-minmorow[nn, 2]
iii<-minmorow[nn, 3]
subpre[[1]]<-vector()
for(j in 2:ncol(bestsubpar[[1]])) {
if (bestsubpar[[1]][i, j] == TRUE) subpre[[1]] <-
c(subpre[[1]], names(NMDSpreRCAL[[1]])[j])}

subpre[[2]]<-vector()
for(j in 2:ncol(bestsubpar[[2]])) {
if (bestsubpar[[2]][ii, j] == TRUE) subpre[[2]] <-
c(subpre[[2]], names(NMDSpreRCAL[[2]])[j])}

subpre[[3]]<-vector()
for(j in 2:ncol(bestsubpar[[3]])) {
if (bestsubpar[[3]][iii, j] == TRUE) subpre[[3]] <-
c(subpre[[3]], names(NMDSpreRCAL[[3]])[j])}

write.table(paste(plotname,"Axis Coefficients"), file=paste(plotname,"final.txt"),
append=TRUE)
for (k in 1:dn){
subsetpre <- NMDSpreRCAL[[k]][c("axis", subpre[[k]])]
axislm[[k]]<- lm(axis ~ ., data=subsetpre)
write.table(coef(axislm[[k]]), file=paste(plotname,"final.txt"), append=TRUE)} }

if(Smeth == "StepwiseAIC") axislm <- stepaxis

```

```

postscript(paste(plotname, "final.ps"))

##### Internal Validation (Calibration
Data)=====
  ## calculate the distance between test sites and reference sites
  sdi j <- d.ij(dn, axislm, preRCAL, preRCAL)
  TL<-quantile(sdi j, p=ThetaL/100)
  TU<-quantile(sdi j, p=ThetaU/100)

  ## calculate the Pc for the ROI null model
  p <- ROI.prob(n, NST, TL, TU, sdi j, RCAL)
  OE.results <- O.E(RCAL, p)
  O <- OE.results$O
  E <- OE.results$E
  OEsu m <- OE.results$results
  write.table("Calibration (Internal)", file=paste(plotname, "final.txt"), append=TRUE)

  write.table(O, file=paste(plotname, "final.txt"), append=TRUE)
  write.table(E, file=paste(plotname, "final.txt"), append=TRUE)
  write.table(OEsu m, file=paste(plotname, "final.txt"), append=TRUE)

  #draw the graphs
  op<-par(mfrow=c(4, 3), mar=c(4, 4, 5, 1))
  plot.main<-paste(plotname, "\nn=", n, "NST=", NST, "ThetaL=", ThetaL, "%
ThetaU=", ThetaU, "%", "\nInternal")
  OE.plot(O, E, OEsu m, plot.main)

##### External Validation (Validation
Data)=====
  sdi j <- d.ij(dn, axislm, preRCAL, preRVAL)
  p <- ROI.prob(n, NST, TL, TU, sdi j, RCAL)
  OE.results <- O.E(RVAL, p)
  O <- OE.results$O
  E <- OE.results$E
  OEsu m <- OE.results$results
  write.table("Validation (External)", file=paste(plotname, "final.txt"), append=TRUE)

  write.table(O, file=paste(plotname, "final.txt"), append=TRUE)
  write.table(E, file=paste(plotname, "final.txt"), append=TRUE)
  write.table(OEsu m, file=paste(plotname, "final.txt"), append=TRUE)

  plot.main<-c("external")
  par(mar=c(4, 4, 2, 1))
  OE.plot(O, E, OEsu m, plot.main)

  dev.off()

```

Sort All Results by validation data RMSE

```

#####
#
# Sort all the results depends on external validation data
# -- 9/5/2007 by Yin-Phan Tsang
#
#
results <- read.delim("gAresults.txt")

ex5.RMSE <- sqrt((results$ex5mean-1)^2+(results$ex5sd)^2)

```

```

results<-cbind(results, ex5. RMSE)
o<-order(results$ex5. RMSE)
results.sort<-data.frame(results$Method[o], results$Smethod[o],

results$NeiNo[o], results$ThetaL[o], results$ThetaU[o], results$TP[o], results$NST[o], results$n[o],

results$in0mean[o], results$in0sd[o], results$in0slope[o], results$in0inter[o], results$in0r2[o],

results$in1mean[o], results$in1sd[o], results$in1slope[o], results$in1inter[o], results$in1r2[o],

results$in5mean[o], results$in5sd[o], results$in5slope[o], results$in5inter[o], results$in5r2[o],
      results$in5RMSE[o],

results$ex0mean[o], results$ex0sd[o], results$ex0slope[o], results$ex0inter[o], results$ex0r2[o],

results$ex1mean[o], results$ex1sd[o], results$ex1slope[o], results$ex1inter[o], results$ex1r2[o],

results$ex5mean[o], results$ex5sd[o], results$ex5slope[o], results$ex5inter[o], results$ex5r2[o],
      results$ex5. RMSE[o])

names(results.sort)<-c("Method", "Smethod",
      "NeiNo", "ThetaL", "ThetaU", "TP", "NST", "n",
      "in0. mean", "in0. sd", "in0. slope", "in0. intercept", "in0. r2",

      "in0. 01. mean", "in0. 01. sd", "in0. 01. slope", "in0. 01. intercept", "in0. 01. r2",
      "in0. 5. mean", "in0. 5. sd", "in0. 5. slope", "in0. 5. intercept", "in0. 5. r2",
      "in0. 5. RMSE",
      "ex0. mean", "ex0. sd", "ex0. slope", "ex0. intercept", "ex0. r2",

      "ex0. 01. mean", "ex0. 01. sd", "ex0. 01. slope", "ex0. 01. intercept", "ex0. 01. r2",
      "ex0. 5. mean", "ex0. 5. sd", "ex0. 5. slope", "ex0. 5. intercept", "ex0. 5. r2",
      "ex0. 5. RMSE")

write.table(results.sort, file="sort.results.all.txt")

```

Functions

```

## Funciton (ANNA.prob)-----
# Calculate the capture prbability (Pc) using ANNA method weighting function
# and with a parameters of neighbor numbers
# The weighting function:
#           WF = 1/sqrt(dij)
#
#       where dij= the distance of reference sites to test site
#
ANNA.prob <- function(neino, sdi,j, Ref) {

  ## 1. As Auther suggested
  ##   remove the identical environmental characteristics reference site (set 1/sqrt(sdi,j)=0
  means no effect)

  for(r in 1: nrow(sdi,j))for(c in 1:ncol(sdi,j)) if(sdi,j[r, c] == 0) sdi,j[r, c]<-NA

  ## 2. count only suggested neighbor numbers : neino
  di,j.rank <- matrix(nrow=nrow(sdi,j), ncol=ncol(sdi,j))
  for(c in 1:ncol(sdi,j)) di,j.rank[, c]<-rank(sdi,j[, c])

  tp <- 1/sqrt(sdi,j)
  for(c in 1:ncol(tp)) for (r in 1:nrow(tp)) if(di,j.rank[r, c] > neino) tp[r, c]<- 0

```

```

    p <- t(tp) %*% as.matrix(Ref)
    for (n in 1:nrow(p)) p[n, ] <- p[n, ]/as.numeric(sum(tp[, n]))

    return(p)
}
#-----

## Function (d.ij)-----
# calculate the physical distance between test sites and reference sites
#

d.ij <- function(dn, axis, preRCAL, preRVAL) {

  ## Calculate qik and qjk
  qi <- list(1:dn) ## test sites
  qj <- list(1:dn) ## reference sites

  for(k in 1:dn){
    preaxis <- names(coef(axis[[k]]))
    coefaxis <- t(as.matrix(coef(axis[[k]])[2:length(preaxis)]))
    preRVAL.axisT <- t(preRVAL[c(preaxis[2:length(preaxis)])])
    preRCAL.axisT <- t(preRCAL[c(preaxis[2:length(preaxis)])])
    qi[[k]] <- coefaxis %*% preRVAL.axisT
    qj[[k]] <- coefaxis %*% preRCAL.axisT

    ## Calculate sdi j
    dij <- list(1:dn)

    for (k in 1:dn){
      dij[[k]] <- matrix(nrow=length(qj[[k]]), ncol=length(qi[[k]]))
      for(n in 1:length(qj[[k]])) dij[[k]][n, ] <- (qi[[k]]-qj[[k]][n])^2
      if(k == 1) sdi j <- dij[[k]] else sdi j <- sdi j+dij[[k]]

      sdi j <- sqrt(sdi j)
      return(sdi j)
    }
  }
#-----

## Funciton (ex.AICBIC)-----
# use stepwiseAIC to select parameter from dn axes
# in NMDSpreRCAL data
#

ex.AICBIC <- function(nn, dn, NMDSpreRCAL) {

  library(leaps)

  axislm <- list(1:dn)
  bestsubpar <- list(1:dn)
  bestsubrss <- matrix(data=100, nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)

  for(k in 1:dn){
    ##AIC, BIC
    findbest5sub <- regsubsets(axis~., data=NMDSpreRCAL[[k]], nbest=5, nvmax=ncol(preRCAL))
    bestsubpar[[k]] <- summary(findbest5sub)$which
    bestsubrss[1:(5*(ncol(NMDSpreRCAL[[k]])-2)+1), k] <- summary(findbest5sub)$rss
  }

  mo <- vector()

```



```

for(i in 1:(ncol(preRCAL)-1)) {
  mo<-c(mo, rep(i, 5))
  mo<-c(mo, ncol(preRCAL))
  mo<-matrix(data=rep(mo, dn), nrow=(5*(ncol(preRCAL)-1)+1), ncol=dn)
  parAIC<-nrow(preRCAL)*log(bestsubrss/nrow(preRCAL))+(2*nrow(preRCAL)*(mo+2))/(nrow(preRCAL)-
(mo+2)-1)
  parBIC<-nrow(preRCAL)*log(bestsubrss/nrow(preRCAL))+(mo+2)*log(nrow(preRCAL))

  minmo<-matrix(data=0, nrow=2, ncol=dn)
  minmorow<-matrix(data=0, nrow=2, ncol=dn)
  for(k in 1:dn) {
    for(i in 1:(5*(ncol(preRCAL)-1)+1)) {
      if(parAIC[i, k] == min(parAIC[, k])) {minmo[1, k]<-mo[i, k];minmorow[1, k]<-i}
      if(parBIC[i, k] == min(parBIC[, k])) {minmo[2, k]<-mo[i, k];minmorow[2, k]<-i}
    }

    subpre<-vector("list", dn)
    axislm<-vector("list", dn)

    i<-minmorow[nn, 1]
    ii<-minmorow[nn, 2]
    iii<-minmorow[nn, 3]
    subpre[[1]]<-vector()
    for(j in 2:ncol(bestsubpar[[1]])) {
      if (bestsubpar[[1]][i, j] == TRUE) subpre[[1]] <-
c(subpre[[1]], names(NMDSpreRCAL[[1]][j]))

      subpre[[2]]<-vector()
      for(j in 2:ncol(bestsubpar[[2]])) {
        if (bestsubpar[[2]][ii, j] == TRUE) subpre[[2]] <-
c(subpre[[2]], names(NMDSpreRCAL[[2]][j]))

        subpre[[3]]<-vector()
        for(j in 2:ncol(bestsubpar[[3]])) {
          if (bestsubpar[[3]][iii, j] == TRUE) subpre[[3]] <-
c(subpre[[3]], names(NMDSpreRCAL[[3]][j]))

          for (k in 1:dn) {
            subsetpre <- NMDSpreRCAL[[k]][c("axis", subpre[[k]])]
            axislm[[k]]<- lm(axis ~ ., data=subsetpre)}

          return(axislm)
        }
      }

}

#-----

## Funciton (Null.prob)-----
# Calculate the capture prbability (Pc) using Null model
# the p is taking all the reference site into accooount and without any
# weighting
#

Null.prob <- function(Ref) {

  WF <- matrix(data=1, nrow=nrow(Ref), ncol=ncol(Ref))

  p <- t(WF) %*% as.matrix(Ref)
  for (k in 1:nrow(p)) p[k, ]<- p[k, ]/as.numeric(sum(WF[, k]))
  return(p)
}

#-----

```

```

## Function (O.E5only)-----
-----
# compared the observation O to the capture probability Pc, either Pc>0, 0.01, 0.5
# and calculate the O, E, O/E, and O/Emean, s.d., O vs. E, slope, intercept, r2
#

O.E5only <- function(Obs,p) {

  ##--For different Pc>0, Pc>0.01, Pc>0.5 -----
  -----
  ## calculate O,E
  O <- matrix(data=0, nrow(Obs), ncol=1)
  E <- matrix(data=0, nrow(Obs), ncol=1)
  for( r in 1:nrow(Obs)) {
    for(c in 1: ncol(Obs)) {

      if (p[r,c] >= 0.5) {
        O[r,1]<- O[r,1]+Obs[r,c]
        E[r,1]<- E[r,1]+p[r,c]}

    }
  }
  return(O,E)
}
#-----

## Function (O.E)-----
-----
# compared the observation O to the capture probability Pc, either Pc>0, 0.01, 0.5
# and calculate the O, E, O/E, and O/Emean, s.d., O vs. E, slope, intercept, r2
#

O.E <- function(Obs,p) {

  ##--For different Pc>0, Pc>0.01, Pc>0.5 -----
  -----
  ## calculate O,E
  O <- matrix(data=0, nrow(Obs), ncol=3)
  E <- matrix(data=0, nrow(Obs), ncol=3)
  for( r in 1:nrow(Obs)) {
    for(c in 1: ncol(Obs)) {
      if (p[r,c] >= 0) {
        O[r,1]<- O[r,1]+Obs[r,c]
        E[r,1]<- E[r,1]+p[r,c]}
      if (p[r,c] >= 0.01) {
        O[r,2]<- O[r,2]+Obs[r,c]
        E[r,2]<- E[r,2]+p[r,c]}
      if (p[r,c] >= 0.5) {
        O[r,3]<- O[r,3]+Obs[r,c]
        E[r,3]<- E[r,3]+p[r,c]}
    }
  }
  OE<-data.frame(O,E)

  # for modify O/E ratio in O=0,E=0 situation
  for(c in 1:ncol(O)) {
    for(r in 1:nrow(O)) {
      if((O[r,c]==0) & (E[r,c]==0)) {
        O[r,c]<-1
        E[r,c]<-1
      }
    }
  }
}

```

```

}

OEratio <- O/E

## calculate O/E mean and s.d.
OEmean<-round(matrix(c(mean(OEratio[, 1]), mean(OEratio[, 2]), mean(OEratio[, 3])), nr=1, nc=3), 3)
OEsd<-round(matrix(sd(OEratio), nr=1, nc=3), 3)

## get regression line's slope, intercept and its r^2
b.m_0<-round(as.numeric(coef(lm(O[, 1]^E[, 1]))), 3)
r2_0<-round(summary(lm(O[, 1]^E[, 1]))$r.squared, 3)
b.m_0.01<-round(as.numeric(coef(lm(O[, 2]^E[, 2]))), 3)
r2_0.01<-round(summary(lm(O[, 2]^E[, 2]))$r.squared, 3)
b.m_0.5<-round(as.numeric(coef(lm(O[, 3]^E[, 3]))), 3)
r2_0.5<-round(summary(lm(O[, 3]^E[, 3]))$r.squared, 3)

results<-
data.frame(OEmean[1], OEsd[1], b.m_0[2], b.m_0[1], r2_0, OEmean[2], OEsd[2], b.m_0.01[2], b.m_0.01[1], r2_0.01, O
Emean[3], OEsd[3], b.m_0.5[2], b.m_0.5[1], r2_0.5)
names(results)<-c(
  "Pc0. mean", "Pc0. sd", "Pc0. slope", "Pc0. intercept", "Pc0. r2",
  "Pc0.01. mean", "Pc0.01. sd", "Pc0.01. slope", "Pc0.01. intercept", "Pc0.01. r2",
  "Pc0.5. mean", "Pc0.5. sd", "Pc0.5. slope", "Pc0.5. intercept", "Pc0.5. r2")

return(O, E, results)
}
#-----

## Function (O.Eonly)-----
---
# compared the observation O to the capture probability Pc, either Pc>0, 0.01, 0.5
# and calculate the O, E, O/E, and O/Emean, s.d., O vs. E, slope, intercept, r2
#

O.Eonly <- function(Obs,p){

  ##--For different Pc>0, Pc>0.01, Pc>0.5 -----
  -----

  ## calculate O, E
  O <- matrix(data=O, nrow(Obs), ncol=3)
  E <- matrix(data=O, nrow(Obs), ncol=3)
  for( r in 1:nrow(Obs)) {
    for(c in 1: ncol(Obs)) {
      if (p[r, c] >= 0) {
        O[r, 1]<- O[r, 1]+Obs[r, c]
        E[r, 1]<- E[r, 1]+p[r, c]}
      if (p[r, c] >= 0.01) {
        O[r, 2]<- O[r, 2]+Obs[r, c]
        E[r, 2]<- E[r, 2]+p[r, c]}
      if (p[r, c] >= 0.5) {
        O[r, 3]<- O[r, 3]+Obs[r, c]
        E[r, 3]<- E[r, 3]+p[r, c]}
    }
  }
  return(O, E)
}
#-----

## Function (OE.plot)-----
-----

```

```

# Need the result from O.E to draw the O/E histograms, 0 vs. E scatter plots
# for Pc >0, 0.01, and 0.5. It will have 3 histogram and 3 scatter plots

OE.plot <- function(O,E,results,plotname) {
  p0.m <- results$Pc0.mean
  p0.sd <- results$Pc0.sd
  p0.s <- results$Pc0.slope
  p0.int <- results$Pc0.intercept
  p0.r2 <- results$Pc0.r2
  p1.m <- results$Pc0.01.mean
  p1.sd <- results$Pc0.01.sd
  p1.s <- results$Pc0.01.slope
  p1.int <- results$Pc0.01.intercept
  p1.r2 <- results$Pc0.01.r2
  p5.m <- results$Pc0.5.mean
  p5.sd <- results$Pc0.5.sd
  p5.s <- results$Pc0.5.slope
  p5.int <- results$Pc0.5.intercept
  p5.r2 <- results$Pc0.5.r2

  OERatio <- O/E

  brk<-
c(0, 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.05, 1.15, 1.25, 1.35, 1.45, 1.55, 1.65, 1.75, 1.85, 1.95
, 2.05, 2.15, 2.25, 2.35, 2.45)
  hist(OERatio[, 1], br=brk, col="blue", border="dark blue", main="\n\nO/E of
Pc>0", xlab="O/E", freq=TRUE, ylim=c(0, 40))
  text(0.3, 30, paste("mean=", p0.m, "\ns.d.=", p0.sd), cex=1, adj=0)
  hist(OERatio[, 2], br=brk, col="blue", border="dark blue", xlab="O/E", freq=TRUE, ylim=c(0, 40),
  main=paste(plotname, "\nO/E of Pc>0.01"))
  text(0.3, 30, paste("mean=", p1.m, "\ns.d.=", p1.sd), cex=1, adj=0)
  hist(OERatio[, 3], br=brk, col="blue", border="dark blue", main="\n\nO/E of
Pc>0.5", xlab="O/E", freq=TRUE, ylim=c(0, 40))
  text(0.3, 30, paste("mean=", p5.m, "\ns.d.=", p5.sd), cex=1, adj=0)

  par(mar=c(4, 4, 1, 1))
  plot(E[, 1], O[, 1], xlab="E", ylab="O", xlim=c(0, 25), ylim=c(0, 25))
  abline(coef(lm(O[, 1]~E[, 1])), col="brown")
  text(3, 20, paste("slope=", p0.s, "\nintercept=", p0.int, "\nr^2=", p0.r2), cex=1, adj=0)

  plot(E[, 2], O[, 2], xlab="E", ylab="O", xlim=c(0, 25), ylim=c(0, 25))
  abline(coef(lm(O[, 2]~E[, 2])), col="brown")
  text(3, 20, paste("slope=", p1.s, "\nintercept=", p1.int, "\nr^2=", p1.r2), cex=1, adj=0)

  plot(E[, 3], O[, 3], xlab="E", ylab="O", xlim=c(0, 25), ylim=c(0, 25))
  abline(coef(lm(O[, 3]~E[, 3])), col="brown")
  text(3, 20, paste("slope=", p5.s, "\nintercept=", p5.int, "\nr^2=", p5.r2), cex=1, adj=0)
}
#-----
## Function (result.5sort2)-----
#
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#
result.5sort2 <- function(results) {
  Pc0.5.results<-data.frame(abs(results$Pc0.5.mean-1), results$Pc0.5.sd, abs(results$Pc0.5.slope-
1), abs(results$Pc0.5.intercept), abs(results$Pc0.5.r2-1))
  Pc0.5.rank <- matrix(nrow=nrow(Pc0.5.results), ncol=ncol(Pc0.5.results))
}

```

```

    for(c in 1:ncol(Pc0.5.results)) Pc0.5.rank[,c]<-rank(Pc0.5.results[,c])
    Pc0.5.geomean <- matrix(nrow=nrow(Pc0.5.rank), ncol=1)
    for(r in 1:nrow(Pc0.5.rank)) Pc0.5.geomean[r]<-
(Pc0.5.rank[r,1]*Pc0.5.rank[r,2]*Pc0.5.rank[r,3]*Pc0.5.rank[r,4]*Pc0.5.rank[r,5])^(1/5)
    results<-cbind(results,Pc0.5.geomean)
    o<-order(results$Pc0.5.geomean)
    results.sort<-data.frame(results$n[o],results$ThetaL[o],results$ThetaU[o],results$TPP[o],
    results$Pc0.mean[o],results$Pc0.sd[o],results$Pc0.slope[o],results$Pc0.intercept[o],results$Pc
0.r2[o],
    results$Pc0.01.mean[o],results$Pc0.01.sd[o],results$Pc0.01.slope[o],results$Pc0.01.intercept[o
],results$Pc0.01.r2[o],
    results$Pc0.5.mean[o],results$Pc0.5.sd[o],results$Pc0.5.slope[o],results$Pc0.5.intercept[o],re
sults$Pc0.5.r2[o],
    results$Pc0.5.geomean[o])
    names(results.sort)<-c("n","ThetaL","ThetaU","TPP",
    "Pc0.mean","Pc0.sd","Pc0.slope","Pc0.intercept","Pc0.r2",
    "Pc0.01.mean","Pc0.01.sd","Pc0.01.slope","Pc0.01.intercept","Pc0.01.r2",
    "Pc0.5.mean","Pc0.5.sd","Pc0.5.slope","Pc0.5.intercept","Pc0.5.r2","Pc0.5.geomean")
    return(results.sort)
}
#-----

## Function (result.5sort3)-----
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#
result.5sort3 <- function(results){
    Pc0.5.results<-data.frame(abs(results$Pc0.5.mean-1),results$Pc0.5.sd,abs(results$Pc0.5.slope-
1),abs(results$Pc0.5.intercept),abs(results$Pc0.5.r2-1))
    Pc0.5.rank <- matrix(nrow=nrow(Pc0.5.results), ncol=ncol(Pc0.5.results))
    for(c in 1:ncol(Pc0.5.results)) Pc0.5.rank[,c]<-rank(Pc0.5.results[,c])
    Pc0.5.geomean <- matrix(nrow=nrow(Pc0.5.rank), ncol=1)
    for(r in 1:nrow(Pc0.5.rank)) Pc0.5.geomean[r]<-
(Pc0.5.rank[r,1]*Pc0.5.rank[r,2]*Pc0.5.rank[r,3]*Pc0.5.rank[r,4]*Pc0.5.rank[r,5])^(1/5)
    results<-cbind(results,Pc0.5.geomean)
    o<-order(results$Pc0.5.geomean)
    results.sort<-data.frame(results$n[o],results$ThetaL[o],results$TPP[o],
    results$Pc0.mean[o],results$Pc0.sd[o],results$Pc0.slope[o],results$Pc0.intercept[o],results$Pc
0.r2[o],
    results$Pc0.01.mean[o],results$Pc0.01.sd[o],results$Pc0.01.slope[o],results$Pc0.01.intercept[o
],results$Pc0.01.r2[o],
    results$Pc0.5.mean[o],results$Pc0.5.sd[o],results$Pc0.5.slope[o],results$Pc0.5.intercept[o],re
sults$Pc0.5.r2[o],
    results$Pc0.5.geomean[o])
    names(results.sort)<-c("n","ThetaL","TPP",
    "Pc0.mean","Pc0.sd","Pc0.slope","Pc0.intercept","Pc0.r2",
    "Pc0.01.mean","Pc0.01.sd","Pc0.01.slope","Pc0.01.intercept","Pc0.01.r2",

```

```

"Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2", "Pc0.5.geomean")

return(results.sort)
}
#-----
-----

## Function (result.5sort)-----
-----
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#

result.5sort <- function(results) {

  Pc0.5.results<-data.frame(abs(results$Pc0.5.mean-1), results$Pc0.5.sd, abs(results$Pc0.5.slope-
1), abs(results$Pc0.5.intercept), abs(results$Pc0.5.r2-1))
  Pc0.5.rank <- matrix(nrow=nrow(Pc0.5.results), ncol=ncol(Pc0.5.results))
  for(c in 1:ncol(Pc0.5.results)) Pc0.5.rank[,c]<-rank(Pc0.5.results[,c])
  Pc0.5.geomean <- matrix(nrow=nrow(Pc0.5.rank), ncol=1)
  for(r in 1:nrow(Pc0.5.rank)) Pc0.5.geomean[r]<-
(Pc0.5.rank[r,1]*Pc0.5.rank[r,2]*Pc0.5.rank[r,3]*Pc0.5.rank[r,4]*Pc0.5.rank[r,5])^(1/5)
  results<-cbind(results, Pc0.5.geomean)
  o<-order(results$Pc0.5.geomean)
  results.sort<-
data.frame(results$ThetaL[o], results$ThetaU[o], results$TP[o], results$NST[o], results$n[o], results$Pc0.me
an[o], results$Pc0.sd[o], results$Pc0.slope[o], results$Pc0.intercept[o], results$Pc0.r2[o], results$Pc0.01.
mean[o], results$Pc0.01.sd[o], results$Pc0.01.slope[o], results$Pc0.01.intercept[o], results$Pc0.01.r2[o], r
esults$Pc0.5.mean[o], results$Pc0.5.sd[o], results$Pc0.5.slope[o], results$Pc0.5.intercept[o], results$Pc0.
5.r2[o], results$Pc0.5.geomean[o])
  names(results.sort)<-c("ThetaL", "ThetaU", "TP", "NST", "n",
                        "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",

"Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",

"Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2", "Pc0.5.geomean")

return(results.sort)
}
#-----
-----

## Function (result.5sortanna)-----
-----
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#

result.5sortanna <- function(results) {

  Pc0.5.results<-data.frame(abs(results$Pc0.5.mean-1), results$Pc0.5.sd, abs(results$Pc0.5.slope-
1), abs(results$Pc0.5.intercept), abs(results$Pc0.5.r2-1))
  Pc0.5.rank <- matrix(nrow=nrow(Pc0.5.results), ncol=ncol(Pc0.5.results))
  for(c in 1:ncol(Pc0.5.results)) Pc0.5.rank[,c]<-rank(Pc0.5.results[,c])
  Pc0.5.geomean <- matrix(nrow=nrow(Pc0.5.rank), ncol=1)
  for(r in 1:nrow(Pc0.5.rank)) Pc0.5.geomean[r]<-
(Pc0.5.rank[r,1]*Pc0.5.rank[r,2]*Pc0.5.rank[r,3]*Pc0.5.rank[r,4]*Pc0.5.rank[r,5])^(1/5)
  results<-cbind(results, Pc0.5.geomean)
  o<-order(results$Pc0.5.geomean)
  results.sort<-data.frame(results$NeiNo[o],

```

```

    results$Pc0.mean[o], results$Pc0.sd[o], results$Pc0.slope[o], results$Pc0.intercept[o], results$Pc
0.r2[o],

    results$Pc0.01.mean[o], results$Pc0.01.sd[o], results$Pc0.01.slope[o], results$Pc0.01.intercept[o
], results$Pc0.01.r2[o],

    results$Pc0.5.mean[o], results$Pc0.5.sd[o], results$Pc0.5.slope[o], results$Pc0.5.intercept[o], re
sults$Pc0.5.r2[o],
        results$Pc0.5.geomean[o])
names(results.sort)<-c("NeiNo",
                    "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",

                    "Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",

                    "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2", "Pc0.5.geomean")

    return(results.sort)
}
#-----
##### Function (result.RMSEsortn)-----
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#
result.RMSEsortn <- function(results){

    Pc0.5.RMSE <- matrix(nrow=nrow(results), ncol=1)
    Pc0.5.RMSE<-sqrt((results$Pc0.5.mean-1)^2+(results$Pc0.5.sd)^2)
    results<-cbind(results, Pc0.5.RMSE)
    o<-order(results$Pc0.5.RMSE)
    results.sort<-data.frame(results$ThetaL[o], results$ThetaU[o], results$NST[o], results$n[o],

    results$Pc0.mean[o], results$Pc0.sd[o], results$Pc0.slope[o], results$Pc0.intercept[o], results$Pc
0.r2[o],

    results$Pc0.01.mean[o], results$Pc0.01.sd[o], results$Pc0.01.slope[o], results$Pc0.01.intercept[o
], results$Pc0.01.r2[o],

    results$Pc0.5.mean[o], results$Pc0.5.sd[o], results$Pc0.5.slope[o], results$Pc0.5.intercept[o], re
sults$Pc0.5.r2[o],
        results$Pc0.5.RMSE[o])
names(results.sort)<-c("ThetaL", "ThetaU", "NST", "n",
                    "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",

                    "Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",
                    "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2",
                    "Pc0.5.RMSE")

    return(results.sort)
}
#-----
##### Function (result.RMSEsort2)-----
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#

```

```

result.RMSEsort2 <- function(results) {

  Pc0.5.RMSE <- matrix(nrow=nrow(results), ncol=1)
  Pc0.5.RMSE<-sqrt((results$Pc0.5.mean-1)^2+(results$Pc0.5.sd)^2)
  results<-cbind(results, Pc0.5.RMSE)
  o<-order(results$Pc0.5.RMSE)
  results.sort<-data.frame(results$n[o], results$ThetaL[o], results$ThetaU[o], results$TPP[o],

  results$Pc0.mean[o], results$Pc0.sd[o], results$Pc0.slope[o], results$Pc0.intercept[o], results$Pc
0.r2[o],

  results$Pc0.01.mean[o], results$Pc0.01.sd[o], results$Pc0.01.slope[o], results$Pc0.01.intercept[o
], results$Pc0.01.r2[o],

  results$Pc0.5.mean[o], results$Pc0.5.sd[o], results$Pc0.5.slope[o], results$Pc0.5.intercept[o], re
sults$Pc0.5.r2[o],
      results$Pc0.5.RMSE[o])
  names(results.sort)<-c("n", "ThetaL", "ThetaU", "TPP",
      "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",

"Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",
      "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2",
      "Pc0.5.RMSE")

  return(results.sort)
}
#-----
#-----

## Function (result.RMSEsort3)-----
#-----
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#

result.RMSEsort3 <- function(results) {

  Pc0.5.RMSE <- matrix(nrow=nrow(results), ncol=1)
  Pc0.5.RMSE<-sqrt((results$Pc0.5.mean-1)^2+(results$Pc0.5.sd)^2)
  results<-cbind(results, Pc0.5.RMSE)
  o<-order(results$Pc0.5.RMSE)
  results.sort<-data.frame(results$n[o], results$ThetaL[o], results$TPP[o],

  results$Pc0.mean[o], results$Pc0.sd[o], results$Pc0.slope[o], results$Pc0.intercept[o], results$Pc
0.r2[o],

  results$Pc0.01.mean[o], results$Pc0.01.sd[o], results$Pc0.01.slope[o], results$Pc0.01.intercept[o
], results$Pc0.01.r2[o],

  results$Pc0.5.mean[o], results$Pc0.5.sd[o], results$Pc0.5.slope[o], results$Pc0.5.intercept[o], re
sults$Pc0.5.r2[o],
      results$Pc0.5.RMSE[o])
  names(results.sort)<-c("n", "ThetaL", "TPP",
      "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",

"Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",
      "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2",
      "Pc0.5.RMSE")

  return(results.sort)
}

```



```

#-----
##### Function (result.RMSEsort)-----
#####
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#

result.RMSEsort <- function(results) {

  Pc0.5.RMSE <- matrix(nrow=nrow(results), ncol=1)
  Pc0.5.RMSE<-sqrt((results$Pc0.5.mean-1)^2+(results$Pc0.5.sd)^2)
  results<-cbind(results, Pc0.5.RMSE)
  o<-order(results$Pc0.5.RMSE)
  results.sort<-
data.frame(results$ThetaL[o], results$ThetaU[o], results$TP[o], results$NST[o], results$n[o],

  results$Pc0.mean[o], results$Pc0.sd[o], results$Pc0.slope[o], results$Pc0.intercept[o], results$Pc
0.r2[o],

  results$Pc0.01.mean[o], results$Pc0.01.sd[o], results$Pc0.01.slope[o], results$Pc0.01.intercept[o
], results$Pc0.01.r2[o],

  results$Pc0.5.mean[o], results$Pc0.5.sd[o], results$Pc0.5.slope[o], results$Pc0.5.intercept[o], re
sults$Pc0.5.r2[o],
  results$Pc0.5.RMSE[o])
  names(results.sort)<-c("ThetaL", "ThetaU", "TP", "NST", "n",
  "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",

  "Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",
  "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2",
  "Pc0.5.RMSE")

  return(results.sort)
}
#-----

##### Function (result.RMSEsortanna)-----
#####
# sort the results depends on Pc>0.5's RMSE
# RMSE=sqrt(B^2+SD^2)
# B=mean(O/E)-1
#

result.RMSEsortanna <- function(results) {

  Pc0.5.RMSE <- matrix(nrow=nrow(results), ncol=1)
  Pc0.5.RMSE<-sqrt((results$Pc0.5.mean-1)^2+(results$Pc0.5.sd)^2)
  results<-cbind(results, Pc0.5.RMSE)
  o<-order(results$Pc0.5.RMSE)
  results.sort<-data.frame(results$NeiNo[o],

  results$Pc0.mean[o], results$Pc0.sd[o], results$Pc0.slope[o], results$Pc0.intercept[o], results$Pc
0.r2[o],

  results$Pc0.01.mean[o], results$Pc0.01.sd[o], results$Pc0.01.slope[o], results$Pc0.01.intercept[o
], results$Pc0.01.r2[o],

  results$Pc0.5.mean[o], results$Pc0.5.sd[o], results$Pc0.5.slope[o], results$Pc0.5.intercept[o], re
sults$Pc0.5.r2[o],

```

```

        results$Pc0.5.RMSE[o])
names(results.sort)<-c("NeiNo",
                      "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",
                      "Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",
                      "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2",
                      "Pc0.5.RMSE")

    return(results.sort)
}
#-----

## Function (result.RMSEsortn)-----
#
# sort the results depends on Pc>0.5's
# mean, sd, slope, intercept, and r^2 geometric mean
#

result.RMSEsortn <- function(results) {

    Pc0.5.RMSE <- matrix(nrow=nrow(results), ncol=1)
    Pc0.5.RMSE<-sqrt((results$Pc0.5.mean-1)^2+(results$Pc0.5.sd)^2)
    results<-cbind(results, Pc0.5.RMSE)
    o<-order(results$Pc0.5.RMSE)
    results.sort<-data.frame(results$ThetaL[o], results$ThetaU[o], results$NST[o], results$n[o],

    results$Pc0.mean[o], results$Pc0.sd[o], results$Pc0.slope[o], results$Pc0.intercept[o], results$Pc
0.r2[o],

    results$Pc0.01.mean[o], results$Pc0.01.sd[o], results$Pc0.01.slope[o], results$Pc0.01.intercept[o
], results$Pc0.01.r2[o],

    results$Pc0.5.mean[o], results$Pc0.5.sd[o], results$Pc0.5.slope[o], results$Pc0.5.intercept[o], re
sults$Pc0.5.r2[o],

        results$Pc0.5.RMSE[o])
names(results.sort)<-c("ThetaL", "ThetaU", "NST", "n",
                      "Pc0.mean", "Pc0.sd", "Pc0.slope", "Pc0.intercept", "Pc0.r2",
                      "Pc0.01.mean", "Pc0.01.sd", "Pc0.01.slope", "Pc0.01.intercept", "Pc0.01.r2",
                      "Pc0.5.mean", "Pc0.5.sd", "Pc0.5.slope", "Pc0.5.intercept", "Pc0.5.r2",
                      "Pc0.5.RMSE")

    return(results.sort)
}
#-----

## Function (RMSE5only)-----
#
# Use 0 and E calculate RMSE of 0/E
#

RMSE5only <- function(O, E) {

    n <- nrow(O)
    RMSE.5 <- sqrt((sum((O/E-1)^2))/n)
}
#-----

## Function (RMSE)-----

```

```

#      Use O and E calculate RMSE of O/E
#

RMSE <- function(O,E) {

  OEratio <-O/E
  n <- nrow(O)
  RMSE0 <- sqrt(sum(((OEratio-1)^2)[,1])/n)
  RMSE.01 <- sqrt(sum(((OEratio-1)^2)[,2])/n)
  RMSE.5 <- sqrt(sum(((OEratio-1)^2)[,3])/n)
  R <- c(RMSE0, RMSE.01, RMSE.5)
  return(R)
}
#-----
##### Funciton (ROI.prob)-----
# Calculate the capture prbability (Pc) useing ROI method Null model
# The weighting function for Null model:
#      WF = 1-(dij/THL) ^n
#
#      THL = ThetaL  for NSi>=NST
#      THL = ThetaL + (ThetaU-ThetaL)*((NST-NSi)/NST)
#
#      where n = parameters of weighting function,
#      THL =
#      dij = the distance from test site i to reference site j
#      ThetaL = the lower threshold value defining a desired proximity
#              for stations to be included in the ROI for the site i
#      NSi = the number of stations in the ROI for site i with the threshold
#           at thetaL.
#      NST = the target number of stations for a region of influence,
#      ThetaU = an upper threshold value for sites with fewer than NST
#              station in the ROI
#
#
ROI.prob <- function(n, NST, ThetaL, ThetaU, sdi,j, Ref) {

  NSi <- matrix(data=0, nrow=1, ncol=ncol(sdi,j))
  THL <- matrix(data=ThetaL, nrow=1, ncol=ncol(sdi,j))
  WF <- matrix(nrow=nrow(sdi,j), ncol=ncol(sdi,j))

  for(c in 1:ncol(sdi,j)) {
    for(r in 1:nrow(sdi,j)) if(sdi,j[r,c]<=ThetaL) NSi[1,c]<-NSi[1,c]+1

    #option 1 for NSi<NST
    if(NSi[1,c]<NST) {
      THL[1,c]<-ThetaL+(ThetaU-ThetaL)*((NST-NSi[1,c])/NST)
      for(r in 1:nrow(sdi,j)) if(sdi,j[r,c] <= THL[1,c]) NSi[1,c] <- NSi[1,c]+1}

    for (r in 1:nrow(sdi,j)) WF[r,c] <- 1.0 - (sdi,j[r,c]/THL[1,c])^n
    for (r in 1:nrow(sdi,j)) if(WF[r,c] < 0.0) WF[r,c]<- 0.0
  }

  p <- t(WF) %*% as.matrix(Ref)
  for (k in 1:nrow(p)) p[k,]<- p[k,]/as.numeric(sum(WF[,k]))
  return(p)
}
#-----

```

```

### Funciton (ROInull.prob)-----
# Calculate the capture prbability (Pc) useing ROI method Null model
# The weighting function for Null model:
#           WF = 1-(dij/THL)^n
#           where n=2,
#           dij= the distance from test site i to reference site j
#           THL=the maximum value of dij
#
ROInull.prob <- function(n, sdi,j,Ref){
  THL <- max(sdi,j)
  WF <- matrix(nrow=nrow(sdi,j), ncol=ncol(sdi,j))

  WF <- 1.0 - (sdi,j/THL)^n
  for (c in 1:ncol(sdi,j)) for (r in 1:nrow(sdi,j)) if(WF[r, c] < 0.0) WF[r, c]<- 0.0

  p <- t(WF) %*% as.matrix(Ref)
  for (k in 1:nrow(p)) p[k, ]<- p[k, ]/as.numeric(sum(WF[, k]))
  return(p)
}
#-----

### Funciton (ROIopl.prob)-----
# Calculate the capture prbability (Pc) useing ROI method Null model
# The weighting function for Null model:
#           WF = 1-(dij/TP)^n
#
#           THL = ThetaL for NSi>=NST
#           THL = ThetaL + (ThetaU-ThetaL)*((NST-NSi)/NST)
#
#           where n = parameters of weighting function,
#           TP = parameter of weighting function, e.g. 85% of distance value
#           dij = the distance from test site i to reference site j
#           ThetaL = the lower threshold value defining a desired proximity
#                   for stations to be included in the ROI for the site i
#           NSi = the number of stations in the ROI for site i with the threshold
#                 at thetaL.
#           NST = the target number of stations for a region of influence,
#           ThetaU = an upper threshold value for sites with fewer than NST
#                   station in the ROI
#
#
ROIopl.prob <- function(n, NST, ThetaL, ThetaU, TP, sdi,j, Ref){

  NSi <- matrix(data=0, nrow=1, ncol=ncol(sdi,j))
  THL <- matrix(data=ThetaL, nrow=1, ncol=ncol(sdi,j))
  WF <- matrix(data=0, nrow=nrow(sdi,j), ncol=ncol(sdi,j))

  for(c in 1:ncol(sdi,j)) {
    for(r in 1:nrow(sdi,j)) if(sdi,j[r, c]<=ThetaL) NSi[1, c]<-NSi[1, c]+1}

  for(c in 1:ncol(sdi,j)){
    #option 1 for NSi<NST
    if(NSi[1, c]<NST) {
      THL[1, c]<-ThetaL+(ThetaU-ThetaL)*(NST-NSi[1, c])/NST}

  for(r in 1:nrow(sdi,j)) {
    if(sdi,j[r, c] <= THL[1, c]){
      WF[r, c] <- 1.0 - (sdi,j[r, c]/TP)^n }}
}

```

```

}

NSi <- matrix(data=0, nrow=1, ncol=ncol(sdi j))
for(c in 1:ncol(sdi j)) {
  for(r in 1:nrow(sdi j)) if(sdi j[r, c]<=THL[1, c]) NSi[1, c]<-NSi[1, c]+1}

p <- t(WF) %*% as.matrix(Ref)
for (k in 1:nrow(p)) p[k, ]<- p[k, ]/as.numeric(sum(WF[, k]))
return(p)
}
#-----

## Funciton (ROIop2. prob)-----
# Calculate the capture prbability (Pc) using ROI method Null model
# The weighting function for Null model:
#      WF = 1-((di j-ThetaL)/(TN-ThetaL))^n   when Di j>ThetaL
#      WF = 1                               otherwise
#
#      TN = Max(TLi, TPP)
#
#      where TLi = Max(Di j)
#
#      where n = parameters of weighting function,
#      TP = parameter of weighting function, e.g. 85% of distance value
#      di j = the distance from test site i to reference site j
#      ThetaL = the lower threshold value defining a desired proximity
#              for stations to be included in the ROI for the site i
#      NSi = the number of stations in the ROI for site i with the threshold
#            at thetaL.
#      NST = the target number of stations for a region of influence,
#      ThetaU = an upper threshold value for sites with fewer than NST
#              station in the ROI
#
#
ROIop2. prob <- function(n, ThetaL, ThetaU, TPP, sdi j, Ref) {

  WF <- matrix(data=0, nrow=nrow(sdi j), ncol=ncol(sdi j))

  for(c in 1:ncol(sdi j)) {
    TLi<-max(sdi j[, c])
    TN<-max(TLi, TPP)
    for(r in 1:nrow(sdi j)) {
      if (sdi j[r, c] <= ThetaU) {
        if(sdi j[r, c] > ThetaL) {
          WF[r, c]<- 1-((sdi j[r, c]-ThetaL)/(TN-ThetaL))^n}
        if(sdi j[r, c] <= ThetaL) {
          WF[r, c]<- 1}
        }
      }
    }

  p <- t(WF) %*% as.matrix(Ref)
  for (k in 1:nrow(p)) p[k, ]<- p[k, ]/as.numeric(sum(WF[, k]))
  return(p)
}
#-----

## Funciton (ROIop2. prob)-----

```

```

# Calculate the capture probability (Pc) using ROI method Null model
# The weighting function for Null model:
#      WF = 1 - ((dij - ThetaL) / (TN - ThetaL))^n   when Dij > ThetaL
#      WF = 1                                       otherwise
#
#      TN = Max(TLi, TPP)
#
#      where TLi = Max(Dij)
#
#      where n = parameters of weighting function,
#      TP = parameter of weighting function, e.g. 85% of distance value
#      dij = the distance from test site i to reference site j
#      ThetaL = the lower threshold value defining a desired proximity
#              for stations to be included in the ROI for the site i
#      NSi = the number of stations in the ROI for site i with the threshold
#            at thetaL.
#      NST = the target number of stations for a region of influence,
#      ThetaU = an upper threshold value for sites with fewer than NST
#              station in the ROI
#
#
ROIop3.prob <- function(n, ThetaL, TPP, sdi, Ref) {

  WF <- matrix(data=0, nrow=nrow(sdi), ncol=ncol(sdi))

  for(c in 1:ncol(sdi)) {
    TLi <- max(sdi[, c])
    TN <- max(TLi, TPP)
    for(r in 1:nrow(sdi)) {
      if (sdi[r, c] <= TLi) {
        if (sdi[r, c] > ThetaL) {
          WF[r, c] <- 1 - ((sdi[r, c] - ThetaL) / (TN - ThetaL))^n
        }
        if (sdi[r, c] <= ThetaL) {
          WF[r, c] <- 1
        }
      }
    }
  }

  p <- t(WF) %*% as.matrix(Ref)
  for (k in 1:nrow(p)) p[k, ] <- p[k, ] / as.numeric(sum(WF[, k]))
  return(p)
}
#-----

### Funciton (stepwise.AIC)-----
# use stepwiseAIC to select parameter from dn axes
# in NMDSpreRCAL data
#

stepwise.AIC <- function(dn, NMDSpreRCAL) {

  stepaxis <- list(1:dn)

  for(k in 1:dn) {
    axislm <- lm(axis ~ ., data= NMDSpreRCAL[[k]])
    stepaxis[[k]] <- stepAIC(axislm, direction="both", trace=FALSE)
  }

  return(stepaxis)
}

```

}

#-----

Appendix C-1. Avenue Code for GISHydro2000

```
'** Create Button (No need to excute again) --
theView = av.GetActiveDoc
HydroExt = Extension.Find("UMDHydro (version 2.0)")
("C:\Documents and Settings\Infant\Desktop\for
  GISHydro\Projects\").AsFileName.SetCWD

IconMgr.Show(0)

'bb = av.getproject.FIndGUI("View").GetButtonbar
,
,
'ROIbutton = Button.Make
'ROIbutton.SetIcon( IconMgr.GetIcons.Get(0))
'ROIbutton.SetClick( "ROI.Show" )
'ROIbutton.SetName("Region-Style BioModels~")
'ROIbutton.SetUpdate( "script.wowUpdate" )
,
,
'' Add the button to the ButtonBar
'bb.Add(ROIbutton, -1 )
,

'** ROI.buttonupdate
' enables the Region-style model button
' only if the view contains outlet point
' and "A Watershed" themes
, -----
theView = av.GetActiveDoc

HydroExt = Extension.Find("UMDHydro (version 2.0)")
xoutletstring = hydroExt.GetPreferences.Get("X outlet Location")

if (xoutletstring <> NIL) then
  test = theView.Findtheme("A Watershed")
  if (test <> NIL) then
    SELF.SetEnabled(TRUE)
    return nil
  end
end
SELF.SetEnabled(FALSE)

'** ROI.CanUnload

' can't unload extension of there are any dialogs in the project,
' or if you can't unload the dialog designer library
, -----
```



```

noDialogs = (av.GetProject = nil) or (av.getProject.GetDialogs.Count = 0)
return (noDialogs and System.CanUnloadLibrary(SELF.GetDependencies.Get(0)))

' ** ROI.CanUnload

' can't unload extension of there are any dialogs in the project,
' or if you can't unload the dialog designer library
' -----
noDialogs = (av.GetProject = nil) or (av.getProject.GetDialogs.Count = 0)
return (noDialogs and System.CanUnloadLibrary(SELF.GetDependencies.Get(0)))

' ** ROI.Close
av.FindDialog("Region-style models dialog").Close

' --- GetHIGHELEV ---

theDEMGrid = self.get(0)
theWatershed = self.get(1)

basintab = theWatershed.getVTab
ctfield = basintab.findfield("Count")
basinarea = basintab.returnvalue(ctfield, 0)

highelev = (theDEMGrid >= 2000.AsGrid) * theWatershed
temptab = highelev.getVTab
tempcnt = temptab.findfield("Count")
tempval = temptab.findfield("Value")
nrec = temptab.getnumrecords
highbasinarea = 0
for each i in 1..nrec
    theval = temptab.returnvalue(tempval, i - 1)
    if (theval = 1) then
        highbasinarea = temptab.returnvalue(tempcnt, i - 1)
    end
end
highbasinarea = highbasinarea / basinarea * 100.0

highelev = highbasinarea

return highelev

' --- GetHUC14HUC8 ---

theView = self.get(0)
thepoint = self.get(1)

hucthemename = SrcName.Make("C:\Documents and Settings\Infant\Desktop\for
GISHydro\Projects\huc14stpm.shp")

```

```

if (hucthemename <> nil) then
  ' Make a Theme from this data source
  theHUCtheme = Theme.Make( hucthemename )
  ' theView.AddTheme( theHUCtheme )
end

huctab = theHUCtheme.GetFtab
hucbitmap = huctab.getselection
hucbitmap.clearall

huctab.SelectByPoint(thepoint, 1, #VTAB_SELTYPE_NEW)
for each huc in huctab.getselection
  huc14field = huctab.findfield("Huc_14")
  huc14 = huctab.returnvalue(huc14field, huc)
  huc8field = huctab.findfield("Huc8")
  huc8 = huctab.returnvalue(huc8field, huc)
end
hucbitmap.clearall

return {huc14, huc8}

' --- GetHUC14HUC8 ---

theView = self.get(0)
thepoint = self.get(1)

hucthemename = SrcName.Make("C:\Documents and Settings\Infant\Desktop\for
GISHydro\Projects\huc14stpm.shp")

if (hucthemename <> nil) then
  ' Make a Theme from this data source
  theHUCtheme = Theme.Make( hucthemename )
  ' theView.AddTheme( theHUCtheme )
end

huctab = theHUCtheme.GetFtab
hucbitmap = huctab.getselection
hucbitmap.clearall

huctab.SelectByPoint(thepoint, 1, #VTAB_SELTYPE_NEW)
for each huc in huctab.getselection
  huc14field = huctab.findfield("Huc_14")
  huc14 = huctab.returnvalue(huc14field, huc)
  huc8field = huctab.findfield("Huc8")
  huc8 = huctab.returnvalue(huc8field, huc)
end
hucbitmap.clearall

```

```

return {huc14, huc8}

' --- GetHYPSO ---

theDEMGrid = self.get(0)
theWatershed = self.get(1)

agrid = theWatershed * theDEMGrid
minelev = agrid.getstatistics.get(0)
avelev = agrid.getstatistics.get(2)
normelev = agrid - minelev.asgrid
maxelev = normelev.getstatistics.get(1)
hyp_value = (normelev.getstatistics.get(2)) / maxelev
hypso = hyp_value

return hypso

' --- GetLENGTH -----

theView = self.get(0)
theFlowDirGrid = self.get(1)
theWatershed = self.get(2)

theWatershedFlowGrid = theWatershed * theFlowDirGrid
upgrid = theWatershedFlowGrid.FlowLength(NIL, TRUE)
maxlength = upgrid.GetStatistics.Get(1)
maxlength = Units.Convert(maxlength, theView.GetUnits, #UNITS_LINEAR_MILES)
length = maxlength

return length

' --- GetLongLat ---

apoint = self.get(0)

' Create a MD Stateplane Projection
r = Rect.Make((-79.75)@37.75, (-75)@40.25)
projection1 = Lambert.Make(r)
projection1.SetDescription("Glenn's MD Projection")
projection1.SetCentralMeridian(-77)
projection1.SetReferenceLatitude(37.6666666667)
projection1.SetLowerStandardParallel(38.3)
projection1.SetUpperStandardParallel(39.45)
projection1.SetFalseEasting(400000)
projection1.SetFalseNorthing(0)
projection1.SetSpheroid(#SPHEROID_GRS80)

tmpDirFN = _optfilestring.AsFileName

```

```

pointFN = FileName.Merge(tmpDirFN.AsString, "temppoint.shp")
newpointFN = FileName.Merge(tmpDirFN.AsString, "newtemppoint.shp")

tempftab = ftab.makenew (pointFN, point)
newfield = field.make("id", #FIELD_SHORT, 5, 0)
tempftab.addfields({newfield})

theshapefield = tempftab.findfield("shape")

therec = tempftab.addrecord
tempftab.setvalue(theshapefield, therec, apoint)
tempftab.setvalue(newfield, therec, 0)
tempftab.seteditable(FALSE)

newftab = tempftab.ExportUnprojected(newpointFN, projection1, FALSE)

newshapefield = newftab.findfield("shape")
newpoint = newftab.returnvalue(newshapefield, 0)
y = newpoint.gety
x = -1.0 * (newpoint.getx)
' msgbox.info(y.asstring, x.asstring)
return {x, y}

' --- GetP100 ----
,

theView = self.get(0)
DEMGrid = self.get(1)
theWatershedGrid = self.get(2)

CellSize = theWatershedGrid.GetCellSize

' Calculate Mean Annual Precipitation

sN = grid.MakeSrcName("c:/umdgism/prec/p100-24m")
meanprec = Grid.Make (sN)
aRect = DEMGrid.getextent

Grid.SetAnalysisCellSize (#GRID_ENVTYPE_VALUE, CellSize)
Grid.SetAnalysisExtent (#GRID_ENVTYPE_VALUE, aRect)

theprec = meanprec * theWatershedGrid
precavg = theprec.getstatistics.get(2)
theavg = precavg / (1000) ' native map values in cm * 1000, this converts to
inches

return theavg

```

```

' --- GetSIN ---

theView = self.get(0)
theDEMGrid = self.get(1)
theFlowDirGrid = self.get(2)
outletpoint = self.get(3)
theWatershed = self.get(4)

thecellsize = theDEMGrid.getcellsize

theWatershedFlowGrid = theWatershed * theFlowDirGrid
upgrid = theWatershedFlowGrid.FlowLength(NIL, TRUE)
downgrid = theWatershedFlowGrid.FlowLength(NIL, FALSE)
mindown = downgrid.getstatistics.get(0)
downgrid = downgrid - mindown.asgrid

sumgrid = upgrid + downgrid
maxlength = upgrid.GetStatistics.Get(1)
maxlengthstat = maxlength
tolerance = 0.1 * (upgrid.GetCellSize)
long_path = (1.AsGrid - (sumgrid > (maxlength -
    tolerance).AsGrid)).SetNull(1.AsGrid)

upvall = 0.9 * maxlengthstat
upval2 = upvall + (thecellsize * (2.sqrt)) 'need to change to cellsize *
    squareroot(2)
aup = downgrid < (upval2.AsGrid)
bup = downgrid > (upvall.AsGrid)

cup = aup * bup * long_path
cup = cup/cup

aFileName = "c:\temp\poly.shp".AsFileName
aPrj = theView.GetProjection
theResult = cup.AsPolygonFTab(aFileName, false, aPrj)

nrecs = theresult.getnumrecords
shapefield = theresult.findfield("shape")
sum = 0

for each i in 1..nrecs
    thepoly = theresult.returnvalue(shapefield, i - 1)
    thecenter = thepoly.returncenter
    thedistance = outletpoint.distance(thecenter)
    sum = sum + thedistance
end
avg = sum / nrecs
sinuosity = upvall / avg

```

```

sin = sinuosity

return sin
' --- GetSoilPct ---
theView = self.get(0)
theWatershedMaskGrid = self.get(1)
soiltheme = self.get(2)

WatershedTab = theWatershedMaskGrid.GetVTab
WatershedCntField = WatershedTab.Findfield("Count")
WatershedArea = WatershedTab.ReturnValue(WatershedCntField,0)

soilgrid = theView.findtheme(soiltheme).getgrid
masksoil = theWatershedMaskGrid*soilgrid
SoilTab = masksoil.getvtab

pctAsoilR = 0
pctBsoilR = 0
pctCsoilR = 0
pctDsoilR = 0

if(SoilTab<>NIL) then
  SoilVal = SoilTab.FindField("Value")
  SoilCnt = SoilTab.Findfield("Count")
  numrecords = SoilTab.GetNumRecords
  for each i in 1..numrecords
    tempval = SoilTab.ReturnValue(SoilVal, i-1)
    if(tempval =1) then
      pctAsoilR = SoilTab.ReturnValue(SoilCnt, i-1)
      pctAsoilR = pctAsoilR/WatershedArea*100
    elseif(tempval =2) then
      pctBsoilR = SoilTab.ReturnValue(SoilCnt, i-1)
      pctBsoilR = pctBsoilR/WatershedArea*100
    elseif(tempval =3) then
      pctCsoilR = SoilTab.ReturnValue(SoilCnt, i-1)
      pctCsoilR = pctCsoilR/WatershedArea*100
    elseif(tempval =4) then
      pctDsoilR = SoilTab.ReturnValue(SoilCnt, i-1)
      pctDsoilR = pctDsoilR/WatershedArea*100
    end
  end
end

return {pctAsoilR, pctBsoilR, pctCsoilR, pctDsoilR}

' -- GetWatershedGridandPoly---
' Obtain Watershed Grid and Polygon

```

```

theView = self.get(0)
theFlowDirGrid = self.get(1)
outletpoint = self.get(2)

' Get Watershed grid
theSrcGrid = theFlowDirGrid.ExtractByPoints(outletpoint.AsMultiPoint,
      Prj.MakeNull, FALSE)
theWatershed = theFlowDirGrid.Watershed(theSrcGrid)
theWatershedMask = theWatershed/theWatershed

' GetWatershed poly
aFileName = (_optfilestring + "tempshed.shp").AsFileName
aPrj = theView.GetProjection
theResult = theWatershedMask.AsPolygonFTab(aFileName, false, aPrj)
theFtheme = ftheme.make(theResult)
shapefield = theResult.findfield("shape")
poly1 = theResult.returnValue(shapefield, 0)
numrecs = theResult.getnumrecords
if (numrecs > 1) then
  for each i in 2..numrecs
    poly2 = theResult.returnValue(shapefield, i - 1)
    poly1 = poly1.returnmerged(poly2)
  end
end
theWatershedVec = poly1

' return watershed grid and polygon
return {theWatershedMask, theWatershedVec}

' =====ROI. Install=====

' Install controls only if there is an active project,
' and if they have never been loaded into the project
' -----
if (av.GetProject = nil) then
  return nil
end
if (SELF.GetPreferences.Get("Loaded") <> NIL) then
  ' upgrade to version 2
  theVer = SELF.GetPreferences.Get("Version")
  if (theVer <> 1.1) then
    SELF.GetPreferences.Set("Version", SELF.GetExtVersion)
  end
  return NIL
end
end

```

```

' install button into a docGUI named "View", if one exists
, -----
viewDocGUI = av.GetProject.FindGUI("View")
if (viewDocGUI <> NIL) then

' Clone button - add cloned button to the project
' and create a space
, -----

theButton = SELF.Get(1).Clone
theSpace = Space.Make
theSpace.SetName("ROI")

' Add button and space to 'View' button bar
' to the right of the ClearSelection button
, -----

viewButtonBar = viewDocGUI.GetButtonBar
location = viewButtonBar.GetControls.Count
viewButtonBar.Add(theSpace, location)
viewButtonBar.Add(theButton, location+1)

' record that view docGUI has been modified
, -----

viewDocGUI.SetModified(TRUE)

' write out "Loaded" and version number
' into the preference dictionary
, -----

SELF.GetPreferences.Add("Loaded", TRUE)
SELF.GetPreferences.Add("Version", SELF.GetExtVersion)

end

' =====ROI.Make=====
theFile = "roi.avx".AsFileName
theName = "Region-Style BioModels~"
theDesc = "Allows you to predict macroinvertebrate biodiversity with Region-Style
          models"
theInstall = av.getproject.findScript("ROI.Install")
theUninstall= av.getproject.findScript("ROI.Uninstall")
theVersion = 1.1
DependsOn = {"$AVBIN/avdlog.dll".AsFileName}

ROIext = Extension.Make(theFile, theName, theInstall, theUninstall, DependsOn)
ROIext.SetAbout(theDesc)
ROIext.SetUnLoadScript(av.FindScript("ROI.Unload"))
ROIext.SetCanUnLoadScript(av.GetProject.FindScript("ROI.CanUnload"))
ROIext.SetExtVersion(1.1)

```



```

' ** Add the Dialog
ROIext.Add(av.FindDialog("Region-style models dialog"))

' ** Add the button that launches the dialog
bbar = av.FindGUI("View").GetButtonBar.GetControls
theButton = bbar.Get(bbar.count-1)
theButton.SetName("ROI")
ROIext.Add(theButton)

' ** Add all scripts needed for extension
for each aDoc in av.GetProject.GetDocs
  if (aDoc.Is(SEd)) then
    if (aDoc.GetName.Contains("ROI.")) then
      if ((aDoc.GetName <> "ROI.Install") and
          (aDoc.GetName <> "ROI.Make") and
          (aDoc.GetName <> "ROI.Uninstall")) then
        if (aDoc.IsCompiled.Not) then
          aDoc.Compile
        end
        roiext.Add(av.FindScript(aDoc.GetName))
      end
    end
  end
end

' commit roi
ROIext.Commit

' ===== ROI Master Scripts for Biodiversity =====

' -----Region-Style BioModels Dialog-----
-----

' Pop up dialog box, to get project name and needed method information
ROIdialog = av.FindDialog("Region-style models dialog")

SiteName=ROIdialog.FindByName("filename").GetText
'SchemeControl = ROIdialog.FindByName("SchemeControl")

ANNA = ROIdialog.FindByName("ANNA").IsSelected
BROI = ROIdialog.FindByName("BROI").IsSelected
NROI = ROIdialog.FindByName("NROI").IsSelected

StepwiseAIC = ROIdialog.FindByName("Stepwise AIC").IsSelected
AIC = ROIdialog.FindByName("Exhaustive AIC").IsSelected
BIC = ROIdialog.FindByName("Exhaustive BIC").IsSelected

Ragan = ROIdialog.FindByName("Ragan").IsSelected
STATSGO = ROIdialog.FindByName("STATSGO").IsSelected

```

```

Family = ROIDialog.FindByName("Family").IsSelected
Genus = ROIDialog.FindByName("Genus").IsSelected

ROIDialog.Close
',
-----

',-----Check Hydro basin statistics-----
area = _areami2
if (area = nil) then
MsgBox.Info( "Please run ""Basin Statistics"" under ""Hydro"" menu first!", "Error
Message!" )
exit
end
',-----

theView = av.GetActiveDoc
HydroExt = Extension.Find("UMDHydro (version 2.0)")
(("$umdgism\roi\").AsFileName).SetCWD

' Get outlet point
xoutlet = HydroExt.getpreferences.Get("X outlet Location").AsNumber
youtlet = HydroExt.getpreferences.Get("Y outlet Location").AsNumber
outletpoint = Point.Make(xoutlet,youtlet)

' Get theme
theFlowDirtheme = theView.findtheme("Flow Direction")
theDEMtheme = theView.findtheme("Original DEM")
theHUCtheme = theView.findtheme("Huc14stpm.shp")

' Get Ftab
' thenhdtab = theView.findtheme("nhd_streamsm.shp").getftab

' Get grid
theDEMGrid = theDEMtheme.
GetGrid
theFlowDirGrid = theFlowDirtheme.GetGrid

' Get Watershed grid(0) and poly(1)
Watershedlist =
    av.run("ROI.GetWatershedGridandPoly", {theView, theFlowDirGrid, outletpoint})

' -- Get predictors --
' AREA
area = _areami2
' or area =HydroExt.getpreferences.Get("Drainage Area") ???

```

```

' WSLOPE (Watershed slope/Landslope =
wslope = _landslope

' CSLOPE (Channel sope = (max_elev-min_elev)/(maxlength*0.75)
cslope = _theslope

' RELIEF
relief =HydroExt.getpreferences.Get("Basin Relief")

' LIME
lime = HydroExt.getpreferences.get("Limestone")

' HUC14
HUC = av.run("ROI.GetHUC14HUC8", {theView, outletpoint})
huc14 = HUC.get(0)

' HUC8
huc8 = HUC.get(1)
huc14 = huc14.AsNumber - (huc8.AsNumber*1000000)

' PERIM
theperim = Watershedlist.Get(1).returnlength
perim = Units.Convert(theperim, theView.GetUnits, #Units_Linear_Miles)

' LENGTH
length = av.run("ROI.GetLENGTH", {theView, theFlowDirGrid, Watershedlist.Get(0)})

' Soil pct
' STATSGO
if (STATSGO) then
' SA
sa = _pctAsoil
' SB
sb = _pctBsoil
' SC
sc = _pctCsoil
' SD
sd = _pctDsoil
end

' Ragan
if (Ragan) then
soiltheme = hydroExt.GetPreferences.Get("Soil Property")
Soilpct = av.run("ROI.GetSoilPct", {theView, Watershedlist.Get(0), soiltheme})
' Ragan data
' SA

```

```

    sa = Soilpct.get(0)
  'SB
    sb = Soilpct.get(1)
  'SC
    sc = Soilpct.get(2)
  'SD
    sd = Soilpct.get(3)
end

' ELEV
elev = theDEMGrid.CellValue(outletpoint, Prj.MakeNull)

' SIN
sin =
    av.run("ROI.GetSIN", {theView, theDEMGrid, theFlowDirGrid, outletpoint, Watershe
dlist.Get(0)})

' TOTLENGTH
' totlength = av.run("ROI.GetTOTLENGTH", {Watershedlist.get(1), thenhdtab, theView})
' totlength = 0

' P2
p2 = _p2yr '????

' P100
p100 = av.run("ROI.GetP100", {theView, theDEMGrid, Watershedlist.get(0)})

' HIGHELEV
highlev = av.run("ROI.GetHIGHELEV", {theDEMGrid, Watershedlist.get(0)})

' HYPSON
hypson = av.run("ROI.GetHYPSON", {theDEMGrid, Watershedlist.get(0)})

' Easting (outlet X coordinate in m MD Stateplane, NAD 1983)
easting = xoutlet

' Northing (outlet Y coordinate in m MD Stateplane, NAD 1983)
northing = youtlet

' LONG
LongLat= av.run("ROI.GetLongLat", {outletpoint})
long = LongLat.get(0)

' Lat
lat = LongLat.get(1)

,
' theString = "YP testing" +NL

```

```

' theString = theString + "Area: " + area.asstring +NL
' theString = theString + "WSLOPE: " +wslope.asstring + NL
' theString = theString + "CSLOPE: " +cslope.asstring + NL
' theString = theString + "RELIEF: " +relief.asstring + NL
' theString = theString + "LIME: " +lime.asstring + NL
' theString = theString + "HUC14: " +huc14.asstring + NL
' theString = theString + "HUC8: " +huc8.asstring + NL
' theString = theString + "PERIM: " + perim.asstring + NL
' theString = theString + "LENGTH: " + length.asstring +NL
' theString = theString + "SA: " + sa.asstring + NL
' theString = theString + "SB: " + sb.asstring + NL
' theString = theString + "SC: " + sc.asstring + NL
' theString = theString + "SD: " + sd.asstring + NL
' theString = theString + "SIN: " + sin.asstring + NL
' theString = theString + "ELEV: " + elev.asstring + NL
' theString = theString + "P2: " + p2.asstring + NL
' theString = theString + "P100: " + p100.asstring + NL
' theString = theString + "HIGHELEV: " + highelev.asstring + NL
' theString = theString + "HYPSO: " + hypso.asstring + NL
' theString = theString + "Outlet Easting: " + xoutlet.asstring + NL
' theString = theString + "Outlet Northing: " + youtlet.asstring + NL
' theString = theString + "Latitude: " + lat.asstring + NL
' theString = theString + "Longitude: " + long.asstring + NL
,
' msgbox.report(thestring,"output")
,

```

```

datastring = area.asstring + TAB +
             wslope.asstring + TAB +
             cslope.asstring + TAB +
             relief.asstring + TAB +
             lime.asstring + TAB +
             huc14.asstring + TAB +
             huc8.asstring + TAB +
             perim.asstring +TAB +
             length.asstring + TAB +
             sa.asstring + TAB +
             sb.asstring + TAB +
             sc.asstring + TAB +
             sd.asstring + TAB +
             elev.asstring + TAB +
             sin.asstring + TAB +
' -----totlength.asstring + TAB +-----
             p2.asstring + TAB +
             p100.asstring + TAB +
             highelev.asstring + TAB +
             hypso.asstring + TAB +
             lat.asstring + TAB +

```

```

        long.asstring + TAB +
        easting.asstring + TAB +
        northing.asstring + NL

' msgbox.report(datastring, "output")

datafilename = FileName.Make("testsitepre.txt")
datafile = TextFile.Make(datafilename, #FILE_PERM_WRITE)
datafile.Write(datastring, 1000)
datafile.Close

' STATSGO
' ANNA
    if((ANNA) and (AIC) and (STATSGO)) then
        System.ExecuteSynchronous("ANNA_A_S.exe")
    elseif((ANNA) and (BIC) and (STATSGO)) then
        System.ExecuteSynchronous("ANNA_B_S.exe")
    elseif((ANNA) and (StepwiseAIC) and (STATSGO)) then
        System.ExecuteSynchronous("ANNA_S_S.exe")
    end
' BROI
    if((BROI) and (AIC) and (STATSGO)) then
        System.ExecuteSynchronous("BROI_A_S.exe")
    elseif((BROI) and (BIC) and (STATSGO)) then
        System.ExecuteSynchronous("BROI_B_S.exe")
    elseif((BROI) and (StepwiseAIC) and (STATSGO)) then
        System.ExecuteSynchronous("BROI_S_S.exe")
    end
' NROI
    if((NROI) and (AIC) and (STATSGO)) then
        System.ExecuteSynchronous("NROI_A_S.exe")
    elseif((NROI) and (BIC) and (STATSGO)) then
        System.ExecuteSynchronous("NROI_B_S.exe")
    elseif((NROI) and (StepwiseAIC) and (STATSGO)) then
        System.ExecuteSynchronous("NROI_S_S.exe")
    end

' Ragan
' ANNA
    if((ANNA) and (AIC) and (Ragan)) then
        System.ExecuteSynchronous("ANNA_A_R.exe")
    elseif((ANNA) and (BIC) and (Ragan)) then
        System.ExecuteSynchronous("ANNA_B_R.exe")
    elseif((ANNA) and (StepwiseAIC) and (Ragan)) then
        System.ExecuteSynchronous("ANNA_S_R.exe")
    end
' BROI

```

```

    if((BROI) and (AIC) and (Ragan)) then
        System.ExecuteSynchronous("BROI_A_R.exe")
    elseif((BROI) and (BIC) and (Ragan)) then
        System.ExecuteSynchronous("BROI_B_R.exe")
    elseif((BROI) and (StepwiseAIC) and (Ragan)) then
        System.ExecuteSynchronous("BROI_S_R.exe")
    end
' NROI
    if((NROI) and (AIC) and (Ragan)) then
        System.ExecuteSynchronous("NROI_A_R.exe")
    elseif((NROI) and (BIC) and (Ragan)) then
        System.ExecuteSynchronous("NROI_B_R.exe")
    elseif((NROI) and (StepwiseAIC) and (Ragan)) then
        System.ExecuteSynchronous("NROI_S_R.exe")
    end

' Load a MD county theme from GISHydro
MDcountytheme = theView.Findtheme("MD Counties")
if ( MDcountytheme = Nil) then
    MDcountythemename = SrcName.Make("C:\umdgism\maryland\mdcountystpm.shp")
    MDcountytheme = Theme.Make( MDcountythemename )
    theView.AddTheme( MDcountytheme )
end
MDcountytheme.SetName ("MD Counties")
MDcountytheme.GetLegend.GetSymbols.Get(0).GetColor.SetTransparent(True)
MDcountytheme.SetVisible(TRUE)

' Make a Reference site Theme from this data source
REFtheme = theView.Findtheme("Reference Sites")
if ( REFtheme = Nil) then
    REFthemename = SrcName.Make("reference sites.shp")
    REFtheme = Theme.Make( REFthemename )
    theView.AddTheme( REFtheme )
end
REFtheme.SetName ("Reference Sites")
REFtheme.GetLegend.GetSymbols.Get(0).SetColor( Color.GetBlue )
REFtheme.SetVisible(TRUE)

' ----- Different taxa levels -----

fbiofilename = FileName.Make("fbio.txt")
gbiofilename = FileName.Make("gbio.txt")
freffilename = FileName.Make("fref.txt")
greffilename = FileName.Make("gref.txt")

if(Family) then
    ' -----Family Level-----

```

```

biofilename = fbiofilename
bionamenumbr = 124
biostring = " Family " + TAB + " Probability:" + NL
biotitle = "Family"
reffilename = freffilename
REFQueryString = "([F_Refid] = 1)"
elseif(Genus) then
' -----Genus Level-----
biofilename = gbiofilename
bionamenumbr = 231
biostring = " Genus " + TAB + " Probability:" + NL
biotitle = "Genus"
reffilename = greffilename
REFQueryString = "([G_Refid] = 1)"
end

biofile = LineFile.Make(biofilename, #FILE_PERM_READ)
biolist = List.Make
biofile.Read(biolist, bionamenumbr)
for each i in 0..(bionamenumbr-1)
biostring = biostring + biolist.get(i).asString + NL
end
msgbox.report(biostring, biotitle ++ "biodiversity")
biofile.Close

reffile = LineFile.Make(reffilename, #FILE_PERM_READ)
reflist = List.Make
reffile.Read(reflist, 130)
refstring = "Reference site " + TAB + "Used:" + NL
for each i in 1..129
refstring = refstring + reflist.get(i).asString + NL
end
msgbox.report(refstring, biotitle ++ "reference site ID")
reffile.Close

' select used reference sites on the theme
REFtab = REFtheme.GetFtab
REFbitmap = REFtab.getselection
REFbitmap.clearall

usedREFIDtablename = reffilename
usedREFIDtable = Vtab.Make(usedREFIDtablename, FALSE, FALSE)
the_to_field = REFtab.findfield("Stationid")
the_from_field = usedREFIDtable.findfield("StationID")
REFtab.Join(the_to_field, usedREFIDtable, the_from_field)

' aVtab.Query (aQueryString, aBitmap, aSelType)

```



```

'REFQueryString = "([F_Refid] = 1)" <--- move to the front
REFTab.Query(REFQueryString, REFbitmap, #VTAB_SELTYPE_NEW)
REFTab.UpdateSelection

REFTab.UnjoinAll

'-----delete text file-----
File.Delete (datafilename)
File.Delete (fbiofilename)
File.Delete (gbiofilename)

'-----output the results to the files-----

if (ANNA) then
  Scheme="ANNA"
elseif (BROI) then
  Scheme="BROI"
elseif (NROI) then
  Scheme="NROI"
end

if (StepwiseAIC) then
  Method="StepwiseAIC"
elseif (AIC) then
  Method="Exhaustive AIC"
elseif (BIC) then
  Method="Exhaustive BIC"
end

if (Ragan) then
  soilsource="Ragan"
elseif (STATSGO) then
  soilsource="STATSGO"
end

if (Family) then
  taxa="Family"
elseif (Genus) then
  taxa="Genus"
end

outputfilename = ( _optfilestring + SiteName + ".txt").AsFileName
if(outputfilename <> NIL) then
  outputfile = LineFile.Make(outputfilename, #FILE_PERM_WRITE )

  thedate=date.now
  themonth =thedata.getmonth
  theday = thedate.getdayofmonth.AsString
  theyear = thedate.getyear.AsString

```

```

outputstring = "Biodiversity for:" ++ SiteName + NL+NL
outputstring = outputstring + "Region-style Model"+NL
outputstring = outputstring + " Analysis Date:          " ++ themonth ++
    theday +"," ++ theyear +NL
outputstring = outputstring + " Select Neighbor Scheme: "++ Scheme +NL
outputstring = outputstring + " Select Parameter Method:"++ Method +NL
outputstring = outputstring + " Soil Data Source:          "++ soilsource +NL

outputfile.WriteElt(outputstring)
outputfile.WriteElt(biostring)
outputfile.WriteElt(refstring)
outputfile.Close
end

' -- zoom to full extent----
r = theView.ReturnExtent

if (r.IsEmpty) then
    return nil
elseif ( r.ReturnSize = (0@0) ) then
    theView.GetDisplay.PanTo(r.ReturnOrigin)
else
    theView.GetDisplay.SetExtent(r.Scale(0.685))
    av.GetProject.SetModified(true)
end

' ** ROI. show
MsgBox.Info( "If you would like to select ""Ragan"" as soil data source,"++
"make Sure it is selected when select the Quads soil data source."++
"The prediction results will be wrong otherwise." , "Warning Message!" )

ROIdialog = av.FindDialog("Region-style models dialog")

' ---initialize dialog-----
ROIdialog.FindByName("filename").SetText("Region-style taxa prediction ")

ROIdialog.FindByName("NROI").Select

ROIdialog.FindByName("Exhaustive AIC").Select

ROIdialog.FindByName("Ragan").Select

ROIdialog.FindByName("Family").Select

ROIdialog.Open

' =====ROI.Uninstall=====

```

```

' Uninstalls controls in the user interface, unless
' there is no project or project is closing.
', -----
if (av.GetProject = nil) then
    return nil
end
if (av.GetProject.IsClosing) then
    return nil
end

' locate button and space, remove from the button bar
', -----
viewDocGUI = av.GetProject.FindGUI("View")
if (viewDocGUI <> NIL) then
    bbar = viewDocGui.GetButtonBar
    aControl = bbar.FindByName("ROI")
    while (aControl <> nil)
        bbar.Remove(aControl)
        aControl = bbar.FindByName("ROI")
    end

    viewDocGUI.SetModified(TRUE)
end

' ** ROI.Unload

' required for the dialog
', -----
Dialog.DetachFromExtension(SELF)

' ** Script 1
' theView = av.GetActiveDoc
' r = theView.ReturnExtent
',
' if (r.IsEmpty) then
'     return nil
' elseif ( r.ReturnSize = (0@0) ) then
'     theView.GetDisplay.PanTo(r.ReturnOrigin)
' else
'     theView.GetDisplay.SetExtent(r.Scale(1.1))
'     av.GetProject.SetModified(true)
' end

((" $umdgism\roi\").AsFileName).SetCWD

```

Appendix C-2. FORTRAN-Code for GISHydro2000

```
!  
! Assessment of Nearest Neighbour Analysis  
! -- Selection - ExhaustiveAIC  
! -- Soil - Ragan  
!  
      program ANNA_A_R  
  
      character(len=15) :: STATIONID(0:129)  
      character(len=1)  :: REF, ctab  
      real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &  
PERIM , LENGTH, SA, SB, SC, SD, ELEV, SINU, P2 ,&  
P100 , HIGHELEV, HYPISO, LAT, LONG, X, Y  
      integer, dimension(0:129) :: HUC14, HUC8  
  
      character(len=17) :: familynames(124)  
      character(len=23) :: genusnames(231)  
      real, dimension(0:129, 124) :: fref, fprob  
      real, dimension(0:129, 231) :: gref, gprob  
  
      real, dimension(0:129) :: axis1, axis2, axis3  
      real, dimension(129)  :: dij  
      integer neino, REFSID(129)  
  
      ctab = CHAR(9)  
11 format(A17, F5.2)  
12 format(A23, F5.2)  
22      format (A15, A1, I5)  
  
! input test site predictors  
      open(10, file='testsitepre.txt', form='formatted', status='old')  
      i=0  
      read(10,*) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &  
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &  
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &  
HIGHELEV (i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)  
      close(10)  
print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &  
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &  
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &  
HIGHELEV (i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)
```

```

! input reference sites predictors
  open(20, file='referencepre_R.txt', form='formatted', status='old')
  read(20, *)

  do i=1, 129
    read(20, *) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
    LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
    SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
    HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
    !print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
! HIGHELEV(i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
  enddo
  close(20)

! input reference site biodiversity composition
!Family level
  open(30, file='referencefamily.txt')
  read(30, *) STATIONID(0), REF, familynames
  do i=1, 129
    read(30, *) STATIONID(i), REF, fref(i, :)
  end do
  !print*, fref

!Genus level
  open(40, file='referencegenus.txt')
  read(40, *) STATIONID(0), REF, genusnames
  do i=1, 129
    read(40, *) STATIONID(i), REF, gref(i, :)
  end do
  !print*, gref

!=== Exhaustive AIC=====
!Family level
  !---select the neighbors-----
  !neighbor number = 17
  !-----
  neino = 17

  do i=0, 129
    axis1(i) = 0.103501402891516 + 0.00142705981016488 * SD(i) + 0.060127473287618 * SINU(i) -
    0.208011303880480 * HYPSO(i) &
    + 4.87448737813407e-07 * X(i) - 1.37533232634292e-06 * Y(i)
    axis2(i) = -4.68452707332487 - 7.94E-07 * HUC14(i) - 2.95E-08 * HUC8(i) - 0.091748923 * SINU(i) &
    + 0.04670853 * P100(i) - 0.466234625 * HYPSO(i) + 0.123474652 * LAT(i)
    axis3(i) = 0.374721271895483 - 2.66836501038278e-08 * HUC8(i) -
    0.00362952990651171 * LENGTH(i) - 0.00127905771301773 * SB(i) &
    - 0.0839121613151153 * SINU(i) - 3.56656830085609e-07 * X(i)

    if (i.NE.0) then
      dij(i) = sqrt(((axis1(i) - axis1(0)) ** 2) + ((axis2(i) - axis2(0)) ** 2) + ((axis3(i) -
      axis3(0)) ** 2))
    end if
  end do

  call ANNAprob(neino, size(fref, 1) - 1, size(fref, 2), dij, fref, fprob, REFSID)
!
  print*, fprob(0, :)
  open(100, file='fbio.txt')
  open(200, file='fref.txt')
  do i=1, 124
    write(100, 11) familynames(i), fprob(0, i)
  end do

```

```

end do
write(200,*) 'StationID', ctab, 'F_RefID'
do i =1, 129
    write(200, 22) STATIONID(i), ctab, REFSID(i)
end do

!Genus level
!---select the neighbors-----
!neighbor number = 16
!-----

neino = 16

do i=0, 129
    axis1(i) = 0. 115378294083608-8. 08205975236042e-05*CSLOPE(i)+0. 08761340169703*SINU(i)-
0. 413037614556601*HYPSO(i)&
    +7. 89510683007598e-07*X(i)-1. 43520199784078e-06*Y(i)
    axis2(i) =0. 601913285-0. 010737439*LIME(i)-0. 002350247*PERIM(i)+0. 001017808*SC(i)-
0. 001563192*HIGHELEV(i)&
    -9. 41E-07*X(i)-1. 31E-06*Y(i)
    axis3(i) =3. 151745421-0. 839735418*WSLOPE(i)+0. 000151981*RELIEF(i)+4. 28E-
07*HUC14(i)+3. 14E-08*HUC8(i)&
    -0. 041017848*P100(i)+0. 238700932*HYPSO(i)-0. 077209157*LAT(i)

    if(i.NE.0) then
        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
    end if
end do

call ANNAprob(neino, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)
!print*, gprob
open(300, file=' gbio. txt')
open(400, file=' gref. txt')
do i=1, 231
    write(300, 12) genusnames(i), gprob(0, i)
end do
write(400, *) 'StationID', ctab, 'G_RefID'
do i =1, 129
    write(400, 22) STATIONID(i), ctab, REFSID(i)
end do

end

```

!!=====

```

===
subroutine ANNAprob(neino, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
integer neino
real, dimension(siteno) :: dij, sortdij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID

do i = 1, siteno
    if(dij(i) .EQ. 0) then
        dij(i) = 1000.0
    end if
end do

sortdij = dij
Call SORT(siteno, sortdij)

```

```

do i = 1, siteno
  WF(i) = 0
  REFSID(i) = 0
  if(dij(i) .LE. sortdij(neino)) then
    WF(i) = 1/sqrt(dij(i))
    REFSID(i) = 1
  end if
  prob(i,:) = WF(i)*ref(i,:)
end do
prob(0,:) = sum(prob,dim=1)/sum(WF)
return
end

```

```

SUBROUTINE SORT(N, RA)
DIMENSION RA(N)
L=N/2+1
IR=N
10  CONTINUE
  IF (L.GT.1) THEN
    L=L-1
    RRA=RA(L)
  ELSE
    RRA=RA(IR)
    RA(IR)=RA(1)
    IR=IR-1
    IF (IR.EQ.1) THEN
      RA(1)=RRA
      RETURN
    END IF
  END IF
  I=L
  J=L+L
20  IF (J.LE.IR) THEN
    IF (J.LT.IR) THEN
      IF (RA(J).LT.RA(J+1)) J=J+1
    END IF
    IF (RRA.LT.RA(J)) THEN
      RA(I)=RA(J)
      I=J
      J=J+J
    ELSE
      J=IR+1
    END IF
    GOTO 20
  END IF
  RA(I)=RRA
  GOTO 10
END

```

```

!
! Assessment of Nearest Neighbour Analysis
! -- Selection - ExhaustiveAIC
! -- Soil - STATSGO
!
  program ANNA_A_S

  character(len=15) :: STATIONID(0:129)
  character(len=1) :: REF, ctab
  real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
  PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &

```

```

P100 , HIGHELEV, HYPSO, LAT, LONG, X, Y
integer, dimension(0:129) :: HUC14,HUC8

character(len=17) :: familynames(124)
character(len=23) :: genusnames(231)
real,dimension(0:129,124) :: fref,fprob
real,dimension(0:129,231) :: gref,gprob

real, dimension(0:129) :: axis1, axis2, axis3
real, dimension(129) :: dij
integer neino,REFSID(129)

ctab = CHAR(9)
11 format(A17,F5.2)
12 format(A23,F5.2)
22 format (A15,A1,I5)

! input test site predictors
open(10,file='testsitepre.txt',form='formatted',status='old')
i=0
read(10,*) AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
close(10)
print*,STATIONID(i), REF,AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
open(20,file='referencepre_S.txt',form='formatted',status='old')
read(20,*)

do i=1,129
read(20,*) STATIONID(i), REF,AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
!print*,STATIONID(i), REF,AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
! LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
! SC(i), SD(i), ELEV(i),SINU(i), TOTLENGTH (i), P2(i), P100(i),&
! HIGHELEV (i), HYPSO(i), X(i), Y (i),LAT(i),LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
open(30,file='referencefamily.txt')
read(30,*) STATIONID(0),REF,familynames
do i=1,129
read(30,*) STATIONID(i),REF,fref(i,:)
end do
!print*,fref

!Genus level
open(40,file='referencegenus.txt')
read(40,*) STATIONID(0),REF,genusnames
do i=1,129
read(40,*) STATIONID(i),REF,gref(i,:)

```



```

end do
!print*, gref

!=== Exhaustive AIC=====
!Family level
!---select the neighbors-----
!neighbor number = 11
!-----
neino = 11

do i=0, 129
axis1(i) = -0.228902566126975+0.00343300716142082*SD(i)+0.0630267442819524*SINU(i)&
+0.0494613662937794*P100(i)-0.19980945923722*HYPSO(i)-
8.37691526171816E-07*Y(i)
axis2(i) = -4.68452707332487-7.94063459434507E-07*HUC14(i)-2.94546884262961E-
08*HUC8(i)&
-0.0917489225780655*SINU(i)+0.0467085299984164*P100(i)-
0.466234624921827*HYPSO(i)&
+0.12347465193802*LAT(i)
axis3(i) = -2.93689970667502-2.84845195384243E-08*HUC8(i)-
0.00369569292494585*LENGTH(i)&
-0.0775318816116617*SINU(i)+0.0403280327951876*LONG(i)

if(i.NE.0)then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call ANNAprob(neino, size(fref, 1)-1, size(fref, 2), dij, fref, fprob, REFSID)
!
print*, fprob(0, :)
open(100, file='fbio.txt')
open(200, file='fref.txt')
do i=1, 124
write(100, 11) familynames(i), fprob(0, i)
end do
write(200, *) 'StationID', ctab, 'F_RefID'
do i =1, 129
write(200, 22) STATIONID(i), ctab, REFSID(i)
end do

!Genus level
!---select the neighbors-----
!neighbor number = 12
!-----

neino = 12

do i=0, 129
axis1(i) = 0.1153782940836-
0.0000808205975236045*CSLOPE(i)+0.0876134016970315*SINU(i)&
-0.413037614556604*HYPSO(i)+7.8951068300761E-07*X(i)-
1.43520199784076E-06*Y(i)
axis2(i) = 5.27530254350311-3.1007633239593E-08*HUC8(i)-0.00252799995614199*PERIM(i)&
+0.00140204995993862*SC(i)-0.0355882192803853*P100(i)-
0.125858522038058*LAT(i)
axis3(i) = -0.0220003274509901-
0.799325726066741*WSLOPE(i)+0.000150496029141555*RELIEF(i)&
+4.83976979040975e-07*HUC14(i)+3.07570302630526e-
08*HUC8(i)+0.00272779276445727*SD(i)&
-0.0329041349558911*P100(i)+0.324484279451702*HYPSO(i)

if(i.NE.0)then

```

```

        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
        end if
    end do

    call ANNAprob(neino, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)
    !print*, gprob
    open(300, file=' gbio. txt')
    open(400, file=' gref. txt')
    do i=1, 231
        write(300, 12) genusnames(i), gprob(0, i)
    end do
    write(400, *) ' StationID', ctab, ' G_RefID'
    do i =1, 129
        write(400, 22) STATIONID(i), ctab, REFSID(i)
    end do

    end

!!=====
===
subroutine ANNAprob(neino, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
integer neino
real, dimension(siteno) :: dij, sortdij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID

do i = 1, siteno
    if(dij(i) .EQ. 0) then
        dij(i) = 1000.0
    end if
end do

sortdij = dij
Call SORT(siteno, sortdij)

do i = 1, siteno
    WF(i) = 0
    REFSID(i) = 0
    if(dij(i) .LE. sortdij(neino)) then
        WF(i) = 1/sqrt(dij(i))
        REFSID(i) = 1
    end if
    prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

SUBROUTINE SORT(N, RA)
DIMENSION RA(N)
L=N/2+1
IR=N
10 CONTINUE
IF (L.GT. 1) THEN
    L=L-1
    RRA=RA(L)
ELSE
    RRA=RA(IR)

```

```

        RA(IR)=RA(1)
        IR=IR-1
        IF (IR.EQ.1) THEN
            RA(1)=RRA
            RETURN
        END IF
    END IF
    I=L
    J=L+L
20    IF (J.LE.IR) THEN
    IF (J.LT.IR) THEN
        IF (RA(J).LT.RA(J+1)) J=J+1
    END IF
    IF (RRA.LT.RA(J)) THEN
        RA(I)=RA(J)
        I=J
        J=J+J
    ELSE
        J=IR+1
    END IF
    GOTO 20
    END IF
    RA(I)=RRA
    GOTO 10
    END

!
! Assessment of Nearest Neighbour Analysis
! -- Selection - ExhaustiveAIC
! -- Soil - Ragan
!
    program ANNA_B_R

    character(len=15) :: STATIONID(0:129)
    character(len=1) :: REF, ctab
    real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
    PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
    P100, HIGHELEV, HYPISO, LAT, LONG, X, Y
    integer, dimension(0:129) :: HUC14, HUC8

    character(len=17) :: familynames(124)
    character(len=23) :: genusnames(231)
    real, dimension(0:129, 124) :: fref, fprob
    real, dimension(0:129, 231) :: gref, gprob

    real, dimension(0:129) :: axis1, axis2, axis3
    real, dimension(129) :: dij
    integer neino, REFSID(129)

    ctab = CHAR(9)
11    format(A17, F5.2)
12    format(A23, F5.2)
22    format(A15, A1, I5)

! input test site predictors
    open(10, file='testsitepre.txt', form='formatted', status='old')
    i=0
    read(10,*) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
    LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
    SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &

```

```

        HIGHELEV (i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)
        close(10)
    print*, STATIONID(i), REF, AREA(i), WLSLOPE(i), CSLOPE(i), RELIEF(i), &
        LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
        SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
        HIGHELEV (i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
    open(20, file='referencepre_R.txt', form='formatted', status='old')
    read(20, *)

    do i=1, 129
        read(20, *) STATIONID(i), REF, AREA(i), WLSLOPE(i), CSLOPE(i), RELIEF(i), &
        LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
        SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
        HIGHELEV (i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)
        !print*, STATIONID(i), REF, AREA(i), WLSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
! HIGHELEV (i), HYPSON(i), X(i), Y(i), LAT(i), LONG(i)
        enddo
    close(20)

! input reference site biodiversity composition
!Family level
    open(30, file='referencefamily.txt')
    read(30, *) STATIONID(0), REF, familynames
    do i=1, 129
        read(30, *) STATIONID(i), REF, fref(i, :)
        end do
    !print*, fref

!Genus level
    open(40, file='referencegenus.txt')
    read(40, *) STATIONID(0), REF, genusnames
    do i=1, 129
        read(40, *) STATIONID(i), REF, gref(i, :)
        end do
    !print*, gref

!=== Exhaustive BIC=====
!Family level
    !---select the neighbors-----
    !neighbor number = 21
    !-----
    neino = 21

    do i=0, 129
        axis1(i) =0. 221544784378139-0. 263036840454740*HYPSON(i)+5. 46892793635458e-07*X(i)&
            -1. 41870482455244e-06*Y(i)
        axis2(i) =-4. 68452707332478-7. 94063459434495e-07*HUC14(i)-2. 94546884262955e-
            08*HUC8(i)&
            -0. 091748922578063*SINU(i)+0. 0467085299984162*P100(i)-
            0. 466234624921829*HYPSON(i)&
            +0. 123474651938017*LAT(i)
        axis3(i) =0. 209020011793382-0. 00161840545458612*SB(i)-0. 102125758799106*SINU(i)

        if (i.NE.0) then
            dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
            axis3(0))**2))
        end if
    end do

```

```

call ANNAprob(neino, size(fref, 1)-1, size(fref, 2), dij, fref, fprob, REFSID)
!
print*, fprob(0, :)
open(100, file='fbio. txt')
open(200, file='fref. txt')
do i=1, 124
    write(100, 11) familynames(i), fprob(0, i)
end do
write(200, *) 'StationID', ctab, 'F_RefID'
do i =1, 129
    write(200, 22) STATIONID(i), ctab, REFSID(i)
end do

!Genus level
!---select the neighbors-----
!neighbor number = 12
!-----

neino = 12

do i=0, 129
    axis1(i) = 0. 104150442366716+0. 098467415429256*SINU(i)-0. 440648301712622*HYPSO(i)&
        +8. 23145543481247e-07*X(i)-1. 51250604759304e-06*Y(i)
    axis2(i) = 0. 810549688-0. 002427075*PERIM(i)-0. 001691071*HIGHELEV(i)-1. 24E-06*X(i)&
        -1. 64E-06*Y(i)
    axis3(i) = 2. 760969421+4. 00E-08*HUC8(i)-0. 07295503*LAT(i)

    if(i.NE.0)then
        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
    end if
end do

call ANNAprob(neino, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)
!print*, gprob
open(300, file='gbio. txt')
open(400, file='gref. txt')
do i=1, 231
    write(300, 12) genusnames(i), gprob(0, i)
end do
write(400, *) 'StationID', ctab, 'G_RefID'
do i =1, 129
    write(400, 22) STATIONID(i), ctab, REFSID(i)
end do

end

!!=====
===
subroutine ANNAprob(neino, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
integer neino
real, dimension(siteno) :: dij, sortdij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID

do i = 1, siteno
    if(dij(i) .EQ. 0) then
        dij(i) = 1000.0
    end if
end do

```

```

end do

sortdij = dij
Call SORT(siteno, sortdij)

do i = 1, siteno
  WF(i) = 0
  REFSID(i) = 0
  if(dij(i) .LE. sortdij(neino)) then
    WF(i) = 1/sqrt(dij(i))
    REFSID(i) = 1
  end if
  prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

```

```

SUBROUTINE SORT(N, RA)
DIMENSION RA(N)
L=N/2+1
IR=N
10  CONTINUE
  IF (L.GT. 1) THEN
    L=L-1
    RRA=RA(L)
  ELSE
    RRA=RA(IR)
    RA(IR)=RA(1)
    IR=IR-1
    IF (IR.EQ. 1) THEN
      RA(1)=RRA
      RETURN
    END IF
  END IF
  I=L
  J=L+L
20  IF (J.LE. IR) THEN
    IF (J.LT. IR) THEN
      IF (RA(J).LT. RA(J+1)) J=J+1
    END IF
    IF (RRA.LT. RA(J)) THEN
      RA(I)=RA(J)
      I=J
      J=J+J
    ELSE
      J=IR+1
    END IF
    GOTO 20
  END IF
  RA(I)=RRA
GOTO 10
END

```

```

!
! Assessment of Nearest Neighbour Analysis
! -- Selection - ExhaustiveAIC
! -- Soil - STATSGO
!

```

```

program ANNA_B_S

```

```

character(len=15) :: STATIONID(0:129)
character(len=1) :: REF, ctab
real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
P100, HIGHELEV, HYPSO, LAT, LONG, X, Y
integer, dimension(0:129) :: HUC14, HUC8

character(len=17) :: familynames(124)
character(len=23) :: genusnames(231)
real, dimension(0:129, 124) :: fref, fprob
real, dimension(0:129, 231) :: gref, gprob

real, dimension(0:129) :: axis1, axis2, axis3
real, dimension(129) :: dij
integer neino, REFSID(129)

ctab = CHAR(9)
11 format (A17, F5.2)
12 format (A23, F5.2)
22 format (A15, A1, I5)

! input test site predictors
open(10, file='testsitepre.txt', form='formatted', status='old')
i=0
read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
close(10)
print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
open(20, file='referencepre_S.txt', form='formatted', status='old')
read(20, *)

do i=1, 129
read(20, *) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
!print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), TOTLENGTH(i), P2(i), P100(i), &
! HIGHELEV(i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
open(30, file='referencefamily.txt')
read(30, *) STATIONID(0), REF, familynames
do i=1, 129
read(30, *) STATIONID(i), REF, fref(i, :)
end do
!print*, fref

```

```

!Genus level
  open(40, file='referencegenus.txt')
  read(40, *) STATIONID(0), REF, genusnames
  do i=1, 129
    read(40, *) STATIONID(i), REF, gref(i, :)
  end do
  !print*, gref

!=== Exhaustive BIC=====
  !Family level
    !---select the neighbors-----
    !neighbor number = 12
    !-----
  neino = 12

  do i=0, 129
    axis1(i) = -0.527670347764048 + 0.00481027605204689 * SD(i) + 0.0650310792558663 * P100(i)
    axis2(i) = -0.49821511574707 - 5.78352960879374e-07 * HUC14(i) - 0.00208005037882994 * SC(i) &
      - 0.0883878663251916 * SINU(i) - 0.465248015085819 * HYPSO(i)
    axis3(i) = -3.22457693563322 - 3.25476880833444E-08 * HUC8(i) + 0.0425161195200811 * LONG(i)

    if(i.NE.0) then
      dij(i) = sqrt(((axis1(i) - axis1(0)) ** 2) + ((axis2(i) - axis2(0)) ** 2) + ((axis3(i) -
axis3(0)) ** 2))
    end if
  end do

  call ANNAprob(neino, size(fref, 1) - 1, size(fref, 2), dij, fref, fprob, REFSID)
!
  print*, fprob(0, :)
  open(100, file='fbio.txt')
  open(200, file='fref.txt')
  do i=1, 124
    write(100, 11) familynames(i), fprob(0, i)
  end do
  write(200, *) 'StationID', ctab, 'F_RefID'
  do i=1, 129
    write(200, 22) STATIONID(i), ctab, REFSID(i)
  end do

!Genus level
  !---select the neighbors-----
  !neighbor number = 30
  !-----

  neino = 30

  do i=0, 129
    axis1(i) = 0.104150442366708 + 0.0984674154292574 * SINU(i) - 0.440648301712624 * HYPSO(i) &
      + 8.23145543481259e-07 * X(i) - 1.51250604759303e-06 * Y(i)
    axis2(i) = 7.244994821464 - 2.80308081073845e-08 * HUC8(i) - 0.00261543850117834 * PERIM(i) &
      - 0.0614198266786265 * P100(i) - 0.169840908042419 * LAT(i)
    axis3(i) = -0.271171913164639 + 3.73799793023689e-08 * HUC8(i) + 0.00263993607998206 * SD(i) &
      + 0.27408320806472 * HYPSO(i)

    if(i.NE.0) then
      dij(i) = sqrt(((axis1(i) - axis1(0)) ** 2) + ((axis2(i) - axis2(0)) ** 2) + ((axis3(i) -
axis3(0)) ** 2))
    end if
  end do

  call ANNAprob(neino, size(gref, 1) - 1, size(gref, 2), dij, gref, gprob, REFSID)
!print*, gprob
  open(300, file='gbio.txt')

```



```

open(400, file=' gref. txt')
do i=1, 231
    write(300, 12) genusnames(i), gprob(0, i)
end do
write(400, *) ' StationID', ctab, ' G_RefID'
do i =1, 129
    write(400, 22) STATIONID(i), ctab, REFSID(i)
end do

end

```

!!=====

```

===
subroutine ANNAprob(neino, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
integer neino
real, dimension(siteno) :: dij, sortdij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID

do i = 1, siteno
    if(dij(i) .EQ. 0) then
        dij(i) = 1000.0
    end if
end do

sortdij = dij
Call SORT(siteno, sortdij)

do i = 1, siteno
    WF(i) = 0
    REFSID(i) = 0
    if(dij(i) .LE. sortdij(neino)) then
        WF(i) = 1/sqrt(dij(i))
        REFSID(i) = 1
    end if
    prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

```

```

SUBROUTINE SORT(N, RA)
DIMENSION RA(N)
L=N/2+1
IR=N
10 CONTINUE
IF (L.GT. 1) THEN
    L=L-1
    RRA=RA(L)
ELSE
    RRA=RA(IR)
    RA(IR)=RA(1)
    IR=IR-1
    IF (IR.EQ. 1) THEN
        RA(1)=RRA
        RETURN
    END IF
END IF
I=L

```

```

20      J=L+L
        IF (J.LE. IR) THEN
        IF (J.LT. IR) THEN
            IF (RA(J).LT. RA(J+1)) J=J+1
        END IF
        IF (RRA.LT. RA(J)) THEN
            RA(I)=RA(J)
            I=J
            J=J+J
        ELSE
            J=IR+1
        END IF
        GOTO 20
    END IF
    RA(I)=RRA
    GOTO 10
END

!
! Assessment of Nearest Neighbour Analysis
! -- Selection - StepwiseAIC
! -- Soil - Ragan
!
    program ANNA_S_R

        character(len=15) :: STATIONID(0:129)
        character(len=1) :: REF, ctab
        real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
        PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
        P100, HIGHELEV, HYPSON, LAT, LONG, X, Y
        integer, dimension(0:129) :: HUC14, HUC8

        character(len=17) :: familynames(124)
        character(len=23) :: genusnames(231)
        real, dimension(0:129, 124) :: fref, fprob
        real, dimension(0:129, 231) :: gref, gprob

        real, dimension(0:129) :: axis1, axis2, axis3
        real, dimension(129) :: dij
        integer neino, REFSID(129)

        ctab = CHAR(9)
11     format(A17, F5.2)
12     format(A23, F5.2)
22     format(A15, A1, I5)

! input test site predictors
    open(10, file='testsitepre.txt', form='formatted', status='old')
    i=0
    read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
    LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
    SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
    HIGHELEV(i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)
    close(10)
    print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
    LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
    SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
    HIGHELEV(i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors

```

```

open(20,file='referencepre_R.txt',form='formatted',status='old')
read(20,*)

do i=1,129
  read(20,*) STATIONID(i), REF, AREA(i), WLSLOPE(i), CSLOPE(i), RELIEF(i), &
  LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
  SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
  HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
  !print*, STATIONID(i), REF, AREA(i), WLSLOPE(i), CSLOPE(i), RELIEF(i), &
  ! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
  ! SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
  ! HIGHELEV(i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
  enddo
close(20)

! input reference site biodiversity composition
!Family level
  open(30,file='referencefamily.txt')
  read(30,*) STATIONID(0), REF, familynames
do i=1,129
  read(30,*) STATIONID(i), REF, fref(i,:)
end do
!print*, fref

!Genus level
  open(40,file='referencegenus.txt')
  read(40,*) STATIONID(0), REF, genusnames
do i=1,129
  read(40,*) STATIONID(i), REF, gref(i,:)
end do
!print*, gref

!=== Stepwise AIC=====
!Family level
  !---select the neighbors-----
  !neighbor number = 13
  !-----
neino = 13

do i=0,129
  axis1(i) = 0.103501402891516+0.060127473287618*SINU(i)-0.208011303880481*HYPSO(i)&
  +4.87448737813407e-07*X(i)-1.37533232634292e-
06*Y(i)+0.00142705981016488*SD(i)
  axis2(i) = -4.38155411892806+0.0119996882743646*LIME(i)-9.1459667141927e-07*HUC14(i)&
  -2.92298312690414e-08*HUC8(i)-
0.090724612776598*SINU(i)+0.0457777467969937*P100(i)&
  -0.42772973937687*HYPSO(i)+0.115512655619171*LAT(i)
  axis3(i) = 0.374721271895483-2.66836501038278e-08*HUC8(i)-0.00127905771301773*SB(i)&
  -0.0839121613151153*SINU(i)-3.56656830085609e-07*X(i)-
0.00362952990651171*LENGTH(i)

  if(i.NE.0)then
    dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
  end if
end do

call ANNAprob(neino, size(fref,1)-1, size(fref,2), dij, fref, fprob, REFSID)
!
print*, fprob(0,:)
open(100,file='fbio.txt')
open(200,file='fref.txt')
do i=1,124

```

```

        write(100,11) familynames(i), fprob(0, i)
    end do
    write(200,*) 'StationID', ctab, 'F_RefID'
    do i =1, 129
        write(200, 22) STATIONID(i), ctab, REFSID(i)
    end do

    !Genus level
    !---select the neighbors-----
    !neighbor number = 18
    !-----

    neino = 18

    do i=0, 129
        axis1(i) = 0.1153782940836-
0.0000808205975236045*CSLOPE(i)+0.0876134016970315*SINU(i)&
-0.413037614556604*HYPSO(i)+7.8951068300761E-07*X(i)-
1.43520199784076E-06*Y(i)
        axis2(i) = 0.601372207265127-0.014604153625166*LIME(i)+2.87390072789817E-07*HUC14(i)-
0.00218450404189344*PERIM(i)&
+0.000940112535214826*SC(i)-
0.00147646824871111*HIGHELEV(i)-9.95268916914963E-07*X(i)-0.0000013297097386447*Y(i)

        axis3(i) = 3.15174542140211-
0.83973541839516*WSLOPE(i)+0.000151981430864665*RELIEF(i)+4.28059610275076E-07*HUC14(i)&
+3.14352669639371E-08*HUC8(i)-
0.041017848451035*P100(i)+0.23870093176202*HYPSO(i)-0.0772091569236982*LAT(i)

        if(i.NE.0)then
            dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
        end if
    end do

    call ANNAprob(neino, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)
    !print*, gprob
    open(300, file='gbio.txt')
    open(400, file='gref.txt')
    do i=1, 231
        write(300, 12) genusnames(i), gprob(0, i)
    end do
    write(400,*) 'StationID', ctab, 'G_RefID'
    do i =1, 129
        write(400, 22) STATIONID(i), ctab, REFSID(i)
    end do

    end

```

```

!!=====
===
subroutine ANNAprob(neino, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
integer neino
real, dimension(siteno) :: dij, sortdij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID

do i = 1, siteno
    if(dij(i) .EQ. 0) then
        dij(i) = 1000.0
    end if
end do

```

```

        end if
    end do

    sortdij = dij
    Call SORT(siteno, sortdij)

    do i = 1, siteno
        WF(i) = 0
        REFSID(i) = 0
        if(dij(i) .LE. sortdij(neino)) then
            WF(i) = 1/sqrt(dij(i))
            REFSID(i) = 1
        end if
        prob(i, :) = WF(i)*ref(i, :)
    end do
    prob(0, :) = sum(prob, dim=1)/sum(WF)
    return
end

```

```

SUBROUTINE SORT(N, RA)
DIMENSION RA(N)
L=N/2+1
IR=N
10  CONTINUE
    IF (L.GT. 1) THEN
        L=L-1
        RRA=RA(L)
    ELSE
        RRA=RA(IR)
        RA(IR)=RA(1)
        IR=IR-1
        IF (IR.EQ. 1) THEN
            RA(1)=RRA
            RETURN
        END IF
    END IF
    I=L
    J=L+L
20  IF (J.LE. IR) THEN
    IF (J.LT. IR) THEN
        IF (RA(J).LT. RA(J+1)) J=J+1
    END IF
    IF (RRA.LT. RA(J)) THEN
        RA(I)=RA(J)
        I=J
        J=J+J
    ELSE
        J=IR+1
    END IF
    GOTO 20
END IF
RA(I)=RRA
GOTO 10
END

```

```

!
! Assessment of Nearest Neighbour Analysis
! -- Selection - StepwiseAIC
! -- Soil - STATSGO
!

```

```

program ANNA_S_S

character(len=15) :: STATIONID(0:129)
character(len=1) :: REF, ctab
real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM , LENGTH , SA, SB, SC, SD, ELEV, SINU, P2, &
P100 , HIGHELEV, HYPISO, LAT, LONG, X, Y
integer, dimension(0:129) :: HUC14, HUC8

character(len=17) :: familynames(124)
character(len=23) :: genusnames(231)
real, dimension(0:129, 124) :: fref, fprob
real, dimension(0:129, 231) :: gref, gprob

real, dimension(0:129) :: axis1, axis2, axis3
real, dimension(129) :: dij
integer neino, REFSID(129)

ctab = CHAR(9)
11 format(A17, F5.2)
12 format(A23, F5.2)
22 format (A15, A1, I5)

! input test site predictors
open(10, file='testsitepre.txt', form='formatted', status='old')
i=0
read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)
close(10)
print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
open(20, file='referencepre_S.txt', form='formatted', status='old')
read(20, *)

do i=1, 129
read(20, *) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)
!print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), TOTLENGTH (i), P2(i), P100(i), &
! HIGHELEV (i), HYPISO(i), X(i), Y(i), LAT(i), LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
open(30, file='referencefamily.txt')
read(30, *) STATIONID(0), REF, familynames
do i=1, 129
read(30, *) STATIONID(i), REF, fref(i, :)
end do
!print*, fref

```

```

!Genus level
open(40, file='referencegenus.txt')
read(40, *) STATIONID(0), REF, genusnames
do i=1, 129
read(40, *) STATIONID(i), REF, gref(i, :)
end do
!print*, gref

!=== Stepwise AIC=====
!Family level
!---select the neighbors-----
!neighbor number = 15
!-----
neino = 15

do i=0, 129
axis1(i) = -0.228902566126975+0.00343300716142082*SD(i)+0.0630267442819524*SINU(i)&
+0.0494613662937794*P100(i)-0.19980945923722*HYPSO(i)-
8.37691526171816e-07*Y(i)
axis2(i) = -4.38155411892814+0.0119996882743647*LIME(i)-9.14596671419284e-
07*HUC14(i)&
-2.92298312690420e-08*HUC8(i)-
0.0907246127766004*SINU(i)+0.045777467969939*P100(i)&
-0.427729739376867*HYPSO(i)+0.115512655619173*LAT(i)
axis3(i) = 0.357058326218296-2.80197266756115E-08*HUC8(i)-
0.00122024560082751*PERIM(i)&
-0.0802674660348683*SINU(i)-4.60917004462554E-07*X(i)

if(i.NE.0) then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call ANNAprob(neino, size(fref, 1)-1, size(fref, 2), dij, fref, fprob, REFSID)
print*, fprob(0, :)
!
open(100, file='fbio.txt')
open(200, file='fref.txt')
do i=1, 124
write(100, 11) familynames(i), fprob(0, i)
end do
write(200, *) 'StationID', ctab, 'F_RefID'
do i =1, 129
write(200, 22) STATIONID(i), ctab, REFSID(i)
end do

!Genus level
!---select the neighbors-----
!neighbor number = 27
!-----

neino = 27

do i=0, 129
axis1(i) = 0.336733741557961-0.000100794776112316*CSLOPE(i)-
0.0000833100341409037*RELIEF(i)&
+0.00200161092985827*SB(i)+0.00342214718303411*SD(i)+0.092631888341564*SINU(i)&
-0.348486383817856*HYPSO(i)-1.83374685029326E-06*Y(i)
axis2(i) = 5.83302279672671-2.5368932042422E-08*HUC8(i)-0.00244299761143094*PERIM(i)&
-0.055034290300751*P100(i)-
0.000614906296416811*HIGHELEV(i)-0.136111136082763*LAT(i)&

```

```

+0.0013364310140715*SC(i)
axis3(i) = -0.0220003274509901-
0.799325726066741*WSLOPE(i)+0.000150496029141555*RELIEF(i)&
+4.83976979040975e-07*HUC14(i)+3.07570302630526e-
08*HUC8(i)+0.00272779276445727*SD(i)&
-0.0329041349558911*P100(i)+0.324484279451702*HYPSO(i)
if(i.NE.0) then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call ANNAprob(neino, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)
!print*, gprob
open(300, file='gbio.txt')
open(400, file='gref.txt')
do i=1, 231
write(300, 12) genusnames(i), gprob(0, i)
end do
write(400, *) 'StationID', ctab, 'G_RefID'
do i =1, 129
write(400, 22) STATIONID(i), ctab, REFSID(i)
end do

end
end

```

!!=====

```

===
subroutine ANNAprob(neino, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
integer neino
real, dimension(siteno) :: dij, sortdij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID

do i = 1, siteno
if(dij(i) .EQ. 0) then
dij(i) = 1000.0
end if
end do

sortdij = dij
Call SORT(siteno, sortdij)

do i = 1, siteno
WF(i) = 0
REFSID(i) = 0
if(dij(i) .LE. sortdij(neino)) then
WF(i) = 1/sqrt(dij(i))
REFSID(i) = 1
end if
prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

```

```

SUBROUTINE SORT(N, RA)
DIMENSION RA(N)
L=N/2+1

```



```

IR=N
10 CONTINUE
   IF (L.GT.1) THEN
       L=L-1
       RRA=RA(L)
   ELSE
       RRA=RA(IR)
       RA(IR)=RA(1)
       IR=IR-1
       IF (IR.EQ.1) THEN
           RA(1)=RRA
           RETURN
       END IF
   END IF
   I=L
   J=L+L
20 IF (J.LE.IR) THEN
   IF (J.LT.IR) THEN
       IF (RA(J).LT.RA(J+1)) J=J+1
   END IF
   IF (RRA.LT.RA(J)) THEN
       RA(I)=RA(J)
       I=J
       J=J+J
   ELSE
       J=IR+1
   END IF
   GOTO 20
END IF
RA(I)=RRA
GOTO 10
END

!
! Burn's Region of Influence Method
! -- Selection - Exhaustive AIC
! -- Soil - Ragan
!
program BROI_A_R

character(len=15) :: STATIONID(0:129)
character(len=1)  :: REF, ctab
real, dimension(0:129) :: AREA,WSLOPE,CSLOPE, RELIEF, LIME, &
PERIM , LENGTH ,SA,SB,SC, SD, ELEV, SINU, P2 ,&
P100 , HIGHELEV, HYPSON, LAT, LONG, X,Y
integer, dimension(0:129) :: HUC14,HUC8

character(len=17) :: familynames(124)
character(len=23) :: genusnames(231)
real,dimension(0:129,124) :: fref, fprob
real,dimension(0:129,231) :: gref, gprob

real, dimension(0:129) :: axis1, axis2, axis3
real, dimension(129) :: dij
integer NST,REFSID(129)
real n,ThetaL, ThetaU,TP

ctab = CHAR(9)
11 format (A17,F5.2)
12 format (A23,F5.2)
22 format (A15,A1,I5)

```

```

! input test site predictors
  open(10,file='testsitepre.txt',form='formatted',status='old')
  i=0
  read(10,*) AREA(i),WSLOPE(i),CSLOPE(i),RELIEF(i),&
  LIME(i),HUC14(i),HUC8(i),PERIM(i),LENGTH(i),SA(i),SB(i),&
  SC(i),SD(i),ELEV(i),SINU(i),P2(i),P100(i),&
  HIGHELEV(i),HYPSO(i),LAT(i),LONG(i),X(i),Y(i)
  close(10)
print*,STATIONID(i),REF,AREA(i),WSLOPE(i),CSLOPE(i),RELIEF(i),&
LIME(i),HUC14(i),HUC8(i),PERIM(i),LENGTH(i),SA(i),SB(i),&
SC(i),SD(i),ELEV(i),SINU(i),P2(i),P100(i),&
HIGHELEV(i),HYPSO(i),LAT(i),LONG(i),X(i),Y(i)

! input reference sites predictors
  open(20,file='referencepre_R.txt',form='formatted',status='old')
  read(20,*)

  do i=1,129
  read(20,*) STATIONID(i),REF,AREA(i),WSLOPE(i),CSLOPE(i),RELIEF(i),&
  LIME(i),HUC14(i),HUC8(i),PERIM(i),LENGTH(i),SA(i),SB(i),&
  SC(i),SD(i),ELEV(i),SINU(i),P2(i),P100(i),&
  HIGHELEV(i),HYPSO(i),LAT(i),LONG(i),X(i),Y(i)
  !print*,STATIONID(i),REF,AREA(i),WSLOPE(i),CSLOPE(i),RELIEF(i),&
! LIME(i),HUC14(i),HUC8(i),PERIM(i),LENGTH(i),SA(i),SB(i),&
! SC(i),SD(i),ELEV(i),SINU(i),P2(i),P100(i),&
! HIGHELEV(i),HYPSO(i),X(i),Y(i),LAT(i),LONG(i)
  enddo
  close(20)

! input reference site biodiversity composition
!Family level
  open(30,file='referencefamily.txt')
  read(30,*) STATIONID(0),REF,familynames
do i=1,129
  read(30,*) STATIONID(i),REF,fref(i,:)
  end do
  !print*,fref

!Genus level
  open(40,file='referencegenus.txt')
  read(40,*) STATIONID(0),REF,genusnames
  do i=1,129
  read(40,*) STATIONID(i),REF,gref(i,:)
  end do
  !print*,gref

!=== Exhaustive AIC =====
!Family level 10 25 75 10 2.5

!---select the neighbors-----
!ThetaL = 10% of total dij data = 0.07397749
!ThetaU = 25% of total dij data = 0.1298829
!TP = 75% of total dij data = 0.2741530
!NST = 10
!n = 2.5
!-----
ThetaL = 0.07397749
ThetaU = 0.1298829
TP = 0.2741530
NST = 10
n = 2.5

```

```

do i=0, 129
    axis1(i) =0. 103501402891516+0. 00142705981016488*SD(i)+0. 060127473287618*SINU(i)-
0. 208011303880480*HYPSO(i)&
                                +4. 87448737813407e-07*X(i)-1. 37533232634292e-06*Y(i)
    axis2(i) =-4. 68452707332487-7. 94E-07*HUC14(i)-2. 95E-08*HUC8(i)-0. 091748923*SINU(i)&
                                +0. 04670853*P100(i)-0. 466234625*HYPSO(i)+0. 123474652*LAT(i)
    axis3(i) = 0. 374721271895483-2. 66836501038278e-08*HUC8(i)-
0. 00362952990651171*LENGTH(i)-0. 00127905771301773*SB(i)&
                                -0. 0839121613151153*SINU(i)-3. 56656830085609e-07*X(i)

    if(i.NE.0)then
        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
    end if
end do

call BROIprob(n, NST, ThetaL, ThetaU, TP, size(fref, 1)-1, size(fref, 2), dij, fref, fprob, REFSID)
!print*, fprob(0, :)
open(100, file='fbio. txt')
open(200, file='fref. txt')
do i=1, 124
    write(100, 11) familynames(i), fprob(0, i)
end do
write(200, *) 'StationID', ctab, 'F_RefID'
do i =1, 129
    write(200, 22) STATIONID(i), ctab, REFSID(i)
end do

!Genus level
!--select the neighbors-----
!ThetaL = 5% of tatal dij data = 0. 06376093          5          30          75
11          2
!ThetaU = 30% of taltal dij data = 0. 1668679
!TP = 75% of tatal dij data = 0. 3225224
!NST = 11
!n = 2
!-----
ThetaL = 0. 06376093
ThetaU = 0. 1668679
TP = 0. 3225224
NST = 11
n = 2

do i=0, 129
    axis1(i) = 0. 115378294083608-8. 08205975236042e-05*CSLOPE(i)+0. 08761340169703*SINU(i)-
0. 413037614556601*HYPSO(i)&
                                +7. 89510683007598e-07*X(i)-1. 43520199784078e-06*Y(i)
    axis2(i) =0. 601913285-0. 010737439*LIME(i)-0. 002350247*PERIM(i)+0. 001017808*SC(i)-
0. 001563192*HIGHELEV(i)&
                                -9. 41E-07*X(i)-1. 31E-06*Y(i)
    axis3(i) =3. 151745421-0. 839735418*WSLOPE(i)+0. 000151981*RELIEF(i)+4. 28E-
07*HUC14(i)+3. 14E-08*HUC8(i)&
                                -0. 041017848*P100(i)+0. 238700932*HYPSO(i)-0. 077209157*LAT(i)

    if(i.NE.0)then
        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
    end if
end do

call BROIprob(n, NST, ThetaL, ThetaU, TP, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)

```

```

!print*, gprob
open(300, file=' gbio. txt')
open(400, file=' gref. txt')
do i=1, 231
    write(300, 12) genusnames(i), gprob(0, i)
end do
write(400, *) ' StationID', ctab, ' G_RefID'
do i =1, 129
    write(400, 22) STATIONID(i), ctab, REFSID(i)
end do

end

!!=====
===
subroutine BROIprob(n, NST, ThetaL, ThetaU, TP, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
real n, ThetaL, ThetaU, TP, TH
real, dimension(siteno) :: dij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID
integer NS, NST

NS = 0
do i = 1, siteno
    if(dij(i) .LE. ThetaL) then
        NS = NS+1
    end if
end do
TH = ThetaL
if(NS .LT. NST) then
    TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1, siteno
    WF(i) = 0
    REFSID(i) = 0
    if(dij(i) .LE. TH) then
        WF(i) = 1.0- (dij(i)/TP)**n
        REFSID(i) = 1
    end if
    prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

!
! Burn's Region of Influence Method
! -- Selection - Exhaustive AIC
! -- Soil - STATSGO
!

program BROI_A_S

character(len=15) :: STATIONID(0:129)
character(len=1) :: REF, ctab
real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM , LENGTH , SA, SB, SC, SD, ELEV, SINU, P2 ,&
P100 , HIGHELEV, HYP50, LAT, LONG, X, Y
integer, dimension(0:129) :: HUC14, HUC8

```

```

character(len=17) :: familynames(124)
character(len=23) :: genusnames(231)
real,dimension(0:129,124) :: fref,fprob
real,dimension(0:129,231) :: gref,gprob

real, dimension(0:129) :: axis1, axis2, axis3
real, dimension(129) :: dij
integer NST,REFSID(129)
real n,ThetaL, ThetaU,TP

ctab = CHAR(9)
11 format (A17,F5.2)
12 format (A23,F5.2)
22 format (A15,A1,I5)

! input test site predictors
open(10,file='testsitepre.txt',form='formatted',status='old')
i=0
read(10,*) AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
HIGHELEV (i), HYPSON(i), LAT(i),LONG(i), X(i), Y(i)
close(10)
print*,STATIONID(i), REF, AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
HIGHELEV (i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
open(20,file='referencepre_S.txt',form='formatted',status='old')
read(20,*)

do i=1,129
read(20,*) STATIONID(i), REF, AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
HIGHELEV (i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)
!print*,STATIONID(i), REF, AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
! LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
! SC(i), SD(i), ELEV(i),SINU(i), TOTLENGTH (i), P2(i), P100(i),&
! HIGHELEV (i), HYPSON(i), X(i), Y(i),LAT(i),LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
open(30,file='referencefamily.txt')
read(30,*) STATIONID(0),REF,familynames
do i=1,129
read(30,*) STATIONID(i),REF,fref(i,:)
end do
!print*,fref

!Genus level
open(40,file='referencegenus.txt')
read(40,*) STATIONID(0),REF,genusnames
do i=1,129
read(40,*) STATIONID(i),REF,gref(i,:)
end do
!print*,gref

```

```

!=== Exhaustive AIC =====
!Family level                                10    30    50    11
  2
  !---select the neighbors-----
  !ThetaL = 10% of tatal dij data = 0.0692568
  !ThetaU = 30% of taltal dij data = 0.1297865
  !TP = 50% of total dij data = 0.1842593
  !NST = 11
  !n = 2
  !-----
ThetaL = 0.0692568
ThetaU = 0.1297865
TP = 0.1842593
NST = 11
n = 2

do i=0,129
  axis1(i) = -0.228902566126975+0.00343300716142082*SD(i)+0.0630267442819524*SINU(i)&
    +0.0494613662937794*P100(i)-0.19980945923722*HYPSO(i)-
8.37691526171816E-07*Y(i)
  axis2(i) = -4.68452707332487-7.94063459434507E-07*HUC14(i)-2.94546884262961E-
08*HUC8(i)&
    -0.0917489225780655*SINU(i)+0.0467085299984164*P100(i)-
0.466234624921827*HYPSO(i)&
    +0.12347465193802*LAT(i)
  axis3(i) = -2.93689970667502-2.84845195384243E-08*HUC8(i)-
0.00369569292494585*LENGTH(i)&
    -0.0775318816116617*SINU(i)+0.0403280327951876*LONG(i)

  if(i.NE.0)then
    dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
  end if
end do

call BROIprob(n,NST,ThetaL,ThetaU,TP,size(fref,1)-1,size(fref,2),dij,fref,fprob,REFSID)
!print*,fprob(0,:)
open(100,file='fbio.txt')
open(200,file='fref.txt')
do i=1,124
  write(100,11) familynames(i),fprob(0,i)
end do
write(200,*) 'StationID',ctab,'F_RefID'
do i =1,129
  write(200,22) STATIONID(i),ctab,REFSID(i)
end do

!Genus level                                10    30    75    8
  3
  !---select the neighbors-----
  !ThetaL = 10% of tatal dij data = 0.09150788
  !ThetaU = 30% of taltal dij data = 0.1666041
  !TP = 75% of tatal dij data = 0.3148751
  !NST = 8
  !n = 3
  !-----
ThetaL = 0.09150788
ThetaU = 0.1666041
TP = 0.3148751
NST = 8
n = 3

```

```

do i=0, 129
    axis1(i) = 0. 1153782940836-
0. 0000808205975236045*CSLOPE(i)+0. 0876134016970315*SINU(i)&
-0. 413037614556604*HYPSO(i)+7. 8951068300761E-07*X(i)-
1. 43520199784076E-06*Y(i)
    axis2(i) = 5. 27530254350311-3. 1007633239593E-08*HUC8(i)-0. 00252799995614199*PERIM(i)&
+0. 00140204995993862*SC(i)-0. 0355882192803853*P100(i)-
0. 125858522038058*LAT(i)
    axis3(i) = -0. 0220003274509901-
0. 799325726066741*WSLOPE(i)+0. 000150496029141555*RELIEF(i)&
+4. 83976979040975e-07*HUC14(i)+3. 07570302630526e-
08*HUC8(i)+0. 00272779276445727*SD(i)&
-0. 0329041349558911*P100(i)+0. 324484279451702*HYPSO(i)
    if(i.NE.0) then
        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
    end if
end do

call BROIprob(n, NST, ThetaL, ThetaU, TP, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)
!print*, gprob
open(300, file=' gbio. txt')
open(400, file=' gref. txt')
do i=1, 231
    write(300, 12) genusnames(i), gprob(0, i)
end do
write(400, *) ' StationID', ctab, ' G_RefID'
do i =1, 129
    write(400, 22) STATIONID(i), ctab, REFSID(i)
end do

end

```

!!=====

```

===
subroutine BROIprob(n, NST, ThetaL, ThetaU, TP, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
real n, ThetaL, ThetaU, TP, TH
real, dimension(siteno) :: dij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID
integer NS, NST

NS = 0
do i = 1, siteno
    if(dij(i) .LE. ThetaL) then
        NS = NS+1
    end if
end do
TH = ThetaL
if(NS .LT. NST) then
    TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1, siteno
    WF(i) = 0
    REFSID(i) = 0
    if(dij(i) .LE. TH) then
        WF(i) = 1.0- (dij(i)/TP)**n
        REFSID(i) = 1
    end if
end do

```

```

        prob(i,:) = WF(i)*ref(i,:)
    end do
    prob(0,:) = sum(prob, dim=1)/sum(WF)
    return
end

!
! Burn's Region of Influence Method
! -- Selection - Exhaustive BIC
! -- Soil - Ragan
!
    program BROI_B_R

        character(len=15) :: STATIONID(0:129)
        character(len=1) :: REF, ctab
        real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
        PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
        P100, HIGHELEV, HYPSO, LAT, LONG, X, Y
        integer, dimension(0:129) :: HUC14, HUC8

        character(len=17) :: familynames(124)
        character(len=23) :: genusnames(231)
        real, dimension(0:129, 124) :: fref, fprob
        real, dimension(0:129, 231) :: gref, gprob

        real, dimension(0:129) :: axis1, axis2, axis3
        real, dimension(129) :: dij
        integer NST, REFSID(129)
        real n, ThetaL, ThetaU, TP

        ctab = CHAR(9)
11    format(A17, F5.2)
12    format(A23, F5.2)
22    format(A15, A1, I5)

! input test site predictors
    open(10, file='testsitepre.txt', form='formatted', status='old')
    i=0
    read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
    LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
    SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
    HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
    close(10)
    print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
    LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
    SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
    HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
    open(20, file='referencepre_R.txt', form='formatted', status='old')
    read(20, *)

    do i=1, 129
        read(20, *) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
        LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
        SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
        HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
        !print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
        ! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
        ! SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
        ! HIGHELEV(i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
    end do

```



```

        enddo
        close(20)

! input reference site biodiversity composition
!Family level
    open(30,file='referencefamily.txt')
    read(30,*) STATIONID(0),REF,familynames
do i=1,129
    read(30,*) STATIONID(i),REF,fref(i,:)
end do
!print*,fref

!Genus level
    open(40,file='referencegenus.txt')
    read(40,*) STATIONID(0),REF,genusnames
do i=1,129
    read(40,*) STATIONID(i),REF,gref(i,:)
end do
!print*,gref

!=== Exhaustive BIC =====
!Family level
!---select the neighbors-----
!ThetaL = 5% of total dij data = 0.0434659
!ThetaU = 25% of total dij data = 0.1128817
!TP = 50% of total dij data = 0.1890502
!NST = 8
!n = 2
!-----
ThetaL = 0.0434659
ThetaU = 0.1128817
TP = 0.1890502
NST = 8
n = 2

do i=0,129
    axis1(i) =0.221544784378139-0.263036840454740*HYPSO(i)+5.46892793635458e-07*X(i)-
1.41870482455244e-06*Y(i)
    axis2(i) =-4.68452707332478-7.94063459434495e-07*HUC14(i)-2.94546884262955e-
08*HUC8(i)-0.091748922578063*SINU(i)&
+0.0467085299984162*P100(i)-
0.466234624921829*HYPSO(i)+0.123474651938017*LAT(i)
    axis3(i) =0.209020011793382-0.00161840545458612*SB(i)-0.102125758799106*SINU(i)

    if(i.NE.0)then
        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
    end if
end do

call BROIprob(n,NST,ThetaL,ThetaU,TP,size(fref,1)-1,size(fref,2),dij,fref,fprob,REFSID)
!print*,fprob(0,:)
open(100,file='fbio.txt')
open(200,file='fref.txt')
do i=1,124
    write(100,11) familynames(i),fprob(0,i)
end do
write(200,*) 'StationID',ctab,'F_RefID'
do i =1,129
    write(200,22) STATIONID(i),ctab,REFSID(i)
end do

```

```

!Genus level
!---select the neighbors-----
!ThetaL = 10% of total dij data =0.0814493      10 25      50      8      2.5
!ThetaU = 25% of taltal dij data =0.1344693
!TP = 50% of tatal dij data = 0.2120401
!NST = 8
!n = 2.5
!-----
ThetaL = 0.0814493
ThetaU = 0.1344693
TP = 0.2120401
NST = 8
n = 2.5

do i=0,129
  axis1(i) = 0.104150442366716+0.098467415429256*SINU(i)-0.440648301712622*HYPSO(i)&
    +8.23145543481247e-07*X(i)-1.51250604759304e-06*Y(i)
  axis2(i) = 0.810549688-0.002427075*PERIM(i)-0.001691071*HIGHELEV(i)-1.24E-06*X(i)&
    -1.64E-06*Y(i)
  axis3(i) = 2.760969421+4.00E-08*HUC8(i)-0.07295503*LAT(i)

  if(i.NE.0)then
    dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
  end if
end do

call BROIprob(n,NST,ThetaL,ThetaU,TP,size(gref,1)-1,size(gref,2),dij,gref,gprob,REFSID)
!print*,gprob
open(300,file='gbio.txt')
open(400,file='gref.txt')
do i=1,231
  write(300,12) genusnames(i),gprob(0,i)
end do
write(400,*) 'StationID',ctab,'G_RefID'
do i =1,129
  write(400,22) STATIONID(i),ctab,REFSID(i)
end do

end

!!=====
===
subroutine BROIprob(n,NST,ThetaL,ThetaU,TP,siteno,taxano,dij,ref,prob,REFSID)
integer siteno,taxano
real n,ThetaL,ThetaU,TP,TH
real,dimension(siteno)::dij
real,dimension(0:siteno,taxano)::ref,prob
real,dimension(siteno)::WF
integer,dimension(siteno)::REFSID
integer NS,NST

NS = 0
do i = 1,siteno
  if(dij(i).LE.ThetaL) then
    NS = NS+1
  end if
end do
TH = ThetaL
if(NS.LT.NST) then

```

```

TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1, siteno
    WF(i) = 0
    REFSID(i) = 0
    if(dij(i) .LE. TH) then
        WF(i) = 1.0- (dij(i)/TP)**n
        REFSID(i) = 1
    end if
    prob(i,:) = WF(i)*ref(i,:)
end do
prob(0,:) = sum(prob, dim=1)/sum(WF)
return
end

!
! Burn's Region of Influence Method
! -- Selection - Exhaustive BIC
! -- Soil - STATSGO
!
    program BROI_B_S

    character(len=15) :: STATIONID(0:129)
    character(len=1) :: REF, ctab
    real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
P100, HIGHELEV, HYPSON, LAT, LONG, X, Y
    integer, dimension(0:129) :: HUC14, HUC8

    character(len=17) :: familynames(124)
    character(len=23) :: genusnames(231)
    real, dimension(0:129, 124) :: fref, fprob
    real, dimension(0:129, 231) :: gref, gprob

    real, dimension(0:129) :: axis1, axis2, axis3
    real, dimension(129) :: dij
    integer NST, REFSID(129)
    real n, ThetaL, ThetaU, TP

    ctab = CHAR(9)
11 format (A17, F5.2)
12 format (A23, F5.2)
22 format (A15, A1, I5)

! input test site predictors
    open(10, file='testsitepre.txt', form='formatted', status='old')
    i=0
    read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)
    close(10)
    print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSON(i), LAT(i), LONG(i), X(i)

! input reference sites predictors
    open(20, file='referencepre_S.txt', form='formatted', status='old')
    read(20, *)

```

```

do i=1, 129
  read(20,*) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
  LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
  SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
  HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
  !print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), TOTLENGTH (i), P2(i), P100(i), &
! HIGHELEV (i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
  open(30, file='referencefamily.txt')
  read(30,*) STATIONID(0), REF, familynames
do i=1, 129
  read(30,*) STATIONID(i), REF, fref(i, :)
end do
!print*, fref

!Genus level
  open(40, file='referencegenus.txt')
  read(40,*) STATIONID(0), REF, genusnames
do i=1, 129
  read(40,*) STATIONID(i), REF, gref(i, :)
end do
!print*, gref

!=== Exhaustive BIC =====
!Family level
2.5
!---select the neighbors-----
!ThetaL = 10% of total dij data = 0.05472064
!ThetaU = 50% of taltal dij data = 0.170987
!TP = 50% of total dij data = 0.170987
!NST = 11
!n = 2.5
!-----
ThetaL = 0.05472064
ThetaU = 0.170987
TP = 0.170987
NST = 11
n = 2.5

do i=0, 129
  axis1(i) = -0.527670347764048 + 0.00481027605204689 * SD(i) + 0.0650310792558663 * P100(i)
  axis2(i) = 0.49821511574707 - 5.78352960879374e-07 * HUC14(i) - 0.00208005037882994 * SC(i) &
  -0.0883878663251916 * SINU(i) - 0.465248015085819 * HYPSO(i)
  axis3(i) = -3.22457693563322 - 3.25476880833444E-08 * HUC8(i) + 0.0425161195200811 * LONG(i)

  if (i.NE.0) then
    dij(i) = sqrt(((axis1(i) - axis1(0))**2) + ((axis2(i) - axis2(0))**2) + ((axis3(i) -
axis3(0))**2))
  end if
end do

call BROIprob(n, NST, ThetaL, ThetaU, TP, size(fref, 1) - 1, size(fref, 2), dij, fref, fprob, REFSID)
!print*, fprob(0, :)
open(100, file='fbio.txt')
open(200, file='fref.txt')
do i=1, 124

```

```

        write(100,11) familynames(i), fprob(0, i)
    end do
    write(200,*) 'StationID', ctab, 'F_RefID'
    do i =1, 129
        write(200, 22) STATIONID(i), ctab, REFSID(i)
    end do

        !Genus level
        3.5
        !---select the neighbors-----
        !ThetaL = 5% of tatal dij data = 0.04906203
        !ThetaU = 30% of taltal dij data = 0.1526518
        !TP = 50% of tatal dij data = 0.2112865
        !NST = 8
        !n = 3.5
        !-----
        ThetaL = 0.04906203
        ThetaU = 0.1526518
        TP = 0.2112865
        NST = 8
        n = 3.5

do i=0, 129
    axis1(i) = 0.104150442366708+0.0984674154292574*SINU(i)-0.440648301712624*HYPSO(i)&
        +8.23145543481259e-07*X(i)-1.51250604759303e-06*Y(i)
    axis2(i) = 7.244994821464-2.80308081073845e-08*HUC8(i)-0.00261543850117834*PERIM(i)&
        -0.0614198266786265*P100(i)-0.169840908042419*LAT(i)
    axis3(i) = -0.271171913164639+3.73799793023689e-08*HUC8(i)+0.00263993607998206*SD(i)-
        +0.27408320806472*HYPSO(i)
    if(i.NE.0) then
        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
    end if
end do

call BROIprob(n, NST, ThetaL, ThetaU, TP, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)
!print*, gprob
open(300, file='gbio.txt')
open(400, file='gref.txt')
do i=1, 231
    write(300, 12) genusnames(i), gprob(0, i)
end do
write(400,*) 'StationID', ctab, 'G_RefID'
do i =1, 129
    write(400, 22) STATIONID(i), ctab, REFSID(i)
end do

end

!!=====
===
subroutine BROIprob(n, NST, ThetaL, ThetaU, TP, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
real n, ThetaL, ThetaU, TP, TH
real, dimension(siteno) :: dij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID
integer NS, NST

```

```

NS = 0
do i = 1, siteno
    if(dij(i) .LE. ThetaL) then
        NS = NS+1
    end if
end do
TH = ThetaL
if(NS .LT. NST) then
    TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1, siteno
    WF(i) = 0
    REFSID(i) = 0
    if(dij(i) .LE. TH) then
        WF(i) = 1.0- (dij(i)/TP)**n
        REFSID(i) = 1
    end if
    prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

!
! Burn's Region of Influence Method
! -- Selection - StepwiseAIC
! -- Soil - Ragan
!
    program BROI_S_R

    character(len=15) :: STATIONID(0:129)
    character(len=1) :: REF, ctab
    real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM , LENGTH , SA, SB, SC, SD, ELEV, SINU, P2 ,&
P100 , HIGHELEV, HYPSO, LAT, LONG, X, Y
    integer, dimension(0:129) :: HUC14, HUC8

    character(len=17) :: familynames(124)
    character(len=23) :: genusnames(231)
    real, dimension(0:129, 124) :: fref, fprob
    real, dimension(0:129, 231) :: gref, gprob

    real, dimension(0:129) :: axis1, axis2, axis3
    real, dimension(129) :: dij
    integer NST, REFSID(129)
    real n, ThetaL, ThetaU, TP

    ctab = CHAR(9)
11 format (A17, F5.2)
12 format (A23, F5.2)
22 format (A15, A1, I5)

! input test site predictors
    open(10, file=' testsitepre.txt', form=' formatted', status=' old')
    i=0
    read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
    close(10)
    print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &

```

```

LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
open(20,file='referencepre_R.txt',form='formatted',status='old')
read(20,*)

do i=1,129
read(20,*) STATIONID(i), REF, AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
!print*,STATIONID(i), REF, AREA(i),WSLOPE(i),CSLOPE(i), RELIEF(i),&
! LIME(i),HUC14(i),HUC8(i), PERIM(i),LENGTH(i), SA(i), SB(i),&
! SC(i), SD(i), ELEV(i),SINU(i), P2(i), P100(i),&
! HIGHELEV (i), HYPSO(i), X(i), Y(i),LAT(i),LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
open(30,file='referencefamily.txt')
read(30,*) STATIONID(0),REF, familynames
do i=1,129
read(30,*) STATIONID(i),REF, fref(i,:)
end do
!print*,fref

!Genus level
open(40,file='referencegenus.txt')
read(40,*) STATIONID(0),REF, genusnames
do i=1,129
read(40,*) STATIONID(i),REF, gref(i,:)
end do
!print*,gref

!=== Stepwise AIC=====
!Family level 10 30 50 8 2.5
!---select the neighbors-----
!ThetaL = 10% of total dij data = 0.0738206
!ThetaU = 30% of total dij data = 0.1431516
!TP = 50% of total dij data = 0.2022362
!NST = 8
!n = 2.5
!-----
ThetaL = 0.0738206
ThetaU = 0.1431516
TP = 0.2022362
NST = 8
n = 2.5

do i=0,129
axis1(i) = 0.103501402891516+0.060127473287618*SINU(i)-0.208011303880481*HYPSO(i)&
+4.87448737813407e-07*X(i)-1.37533232634292e-
06*Y(i)+0.00142705981016488*SD(i)
axis2(i) = -4.38155411892806+0.0119996882743646*LIME(i)-9.1459667141927e-07*HUC14(i)&
-2.92298312690414e-08*HUC8(i)-
0.090724612776598*SINU(i)+0.045777467969937*P100(i)&
-0.42772973937687*HYPSO(i)+0.115512655619171*LAT(i)
axis3(i) = 0.374721271895483-2.66836501038278e-08*HUC8(i)-0.00127905771301773*SB(i)&

```

```

                                -0.0839121613151153*SINU(i)-3.56656830085609e-07*X(i)-
0.00362952990651171*LENGTH(i)

        if(i.NE.0)then
            dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
        end if
    end do

    call BR0Iprob(n,NST,ThetaL,ThetaU,TP,size(fref,1)-1,size(fref,2),dij,fref,fprob,REFSID)
!
    print*,fprob(0,:)
    open(100,file='fbio.txt')
    open(200,file='fref.txt')
    do i=1,124
        write(100,11) familynames(i),fprob(0,i)
    end do
    write(200,*) 'StationID',ctab,'F_RefID'
    do i=1,129
        write(200,22) STATIONID(i),ctab,REFSID(i)
    end do

        !Genus level
        !---select the neighbors-----
        !ThetaL = 5% of total dij data = 0.062747                5    25    50    10
        2
        !ThetaU = 25% of total dij data = 0.1487477
        !TP = 50% of total dij data = 0.2307785
        !NST = 10
        !n = 2
        !-----
        ThetaL = 0.062747
        ThetaU = 0.1487477
        TP = 0.2357665
        NST = 10
        n = 2

    do i=0,129
        axis1(i) = 0.115378294083608-8.08205975236042e-05*CSLOPe(i)+0.08761340169703*SINU(i)&
-0.413037614556601*HYPSO(i)+7.89510683007598e-07*X(i)-
1.43520199784078e-06*Y(i)
        axis2(i) = 0.601372207-0.014604154*LIME(i)+2.87E-07*HUC14(i)-0.002184504*PERIM(i)&
+0.000940113*SC(i)-0.001476468*HIGHELEV(i)-9.95E-07*X(i)-
1.33E-06*Y(i)
        axis3(i) = 3.151745421-0.839735418*WSLOPE(i)+0.000151981*RELIEF(i)+4.28E-07*HUC14(i)&
+3.14E-08*HUC8(i)-
0.041017848*P100(i)+0.238700932*HYPSO(i)-0.077209157*LAT(i)

        if(i.NE.0)then
            dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
        end if
    end do

    call BR0Iprob(n,NST,ThetaL,ThetaU,TP,size(gref,1)-1,size(gref,2),dij,gref,gprob,REFSID)
!print*,gprob
    open(300,file='gbio.txt')
    open(400,file='gref.txt')
    do i=1,231
        write(300,12) genusnames(i),gprob(0,i)
    end do
    write(400,*) 'StationID',ctab,'G_RefID'

```



```

do i =1, 129
    write(400, 22) STATIONID(i), ctab, REFSID(i)
end do

end

!!=====
===
subroutine BROIprob(n, NST, ThetaL, ThetaU, TP, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
real n, ThetaL, ThetaU, TP, TH
real, dimension(siteno) :: dij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID
integer NS, NST

NS = 0
do i = 1, siteno
    if(dij(i) .LE. ThetaL) then
        NS = NS+1
    end if
end do
TH = ThetaL
if(NS .LT. NST) then
    TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1, siteno
    WF(i) = 0
    REFSID(i) = 0
    if(dij(i) .LE. TH) then
        WF(i) = 1.0- (dij(i)/TP)**n
        REFSID(i) = 1
    end if
    prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

!
! Burn's Region of Influence Method
! -- Selection - StepwiseAIC
! -- Soil - STATSGO
!
program BROI_S_S

character(len=15) :: STATIONID(0:129)
character(len=1) :: REF, ctab
real, dimension(0:129) :: AREA, W Slope, CSLOPE, RELIEF, LIME, &
PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
P100, HIGHELEV, HYP SO, LAT, LONG, X, Y
integer, dimension(0:129) :: HUC14, HUC8

character(len=17) :: familynames(124)
character(len=23) :: genusnames(231)
real, dimension(0:129, 124) :: fref, fprob
real, dimension(0:129, 231) :: gref, gprob

real, dimension(0:129) :: axis1, axis2, axis3

```

```

real, dimension(129) :: dij
integer NST, REFSID(129)
real n, ThetaL, ThetaU, TP

ctab = CHAR(9)
11 format (A17, F5.2)
12 format (A23, F5.2)
22 format (A15, A1, I5)

! input test site predictors
open(10, file='testsitepre.txt', form='formatted', status='old')
i=0
read(10,*) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
close(10)
print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
open(20, file='referencepre_S.txt', form='formatted', status='old')
read(20,*)

do i=1, 129
read(20,*) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
!print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), TOTLENGTH(i), P2(i), P100(i), &
! HIGHELEV(i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
open(30, file='referencefamily.txt')
read(30,*) STATIONID(0), REF, familynames
do i=1, 129
read(30,*) STATIONID(i), REF, fref(i, :)
end do
!print*, fref

!Genus level
open(40, file='referencegenus.txt')
read(40,*) STATIONID(0), REF, genusnames
do i=1, 129
read(40,*) STATIONID(i), REF, gref(i, :)
end do
!print*, gref

!=== Stepwise AIC=====
!Family level 10 50 50 8 2.5
!---select the neighbors-----
!ThetaL = 10% of total dij data = 0.06941501
!ThetaU = 50% of total dij data = 0.1827150
!TP = 50% of total dij data = 0.1827150
!NST = 8

```

```

!n = 2.5
!-----
ThetaL = 0.06941501
ThetaU = 0.1827150
TP = 0.1827150
NST = 8
n = 2.5

do i=0, 129
  axis1(i) = -0.228902566126975+0.00343300716142082*SD(i)+0.0630267442819524*SINU(i)&
    +0.0494613662937794*P100(i)-0.19980945923722*HYPSO(i)-
8.37691526171816e-07*Y(i)
  axis2(i) = -4.38155411892814+0.0119996882743647*LIME(i)-9.14596671419284e-
07*HUC14(i)&
    -2.92298312690420e-08*HUC8(i)-
0.0907246127766004*SINU(i)+0.045777467969939*P100(i)&
    -0.427729739376867*HYPSO(i)+0.115512655619173*LAT(i)
  axis3(i) = 0.357058326218296-2.80197266756115E-08*HUC8(i)-
0.00122024560082751*PERIM(i)&
    -0.0802674660348683*SINU(i)-4.60917004462554E-07*X(i)

  if(i.NE.0) then
    dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
  end if
end do

call BROIprob(n, NST, ThetaL, ThetaU, TP, size(fref, 1)-1, size(fref, 2), dij, fref, fprob, REFSID)
print*, fprob(0, :)
!
open(100, file='fbio.txt')
open(200, file='fref.txt')
do i=1, 124
  write(100, 11) familynames(i), fprob(0, i)
end do
write(200, *) 'StationID', ctab, 'F_RefID'
do i =1, 129
  write(200, 22) STATIONID(i), ctab, REFSID(i)
end do

!Genus level
!---select the neighbors-----
!ThetaL = 10% of tatal dij data = 0.09482395
!ThetaU = 25% of taltal dij data = 0.1513344
!TP = 50% of tatal dij data = 0.2225801
!NST = 8
!n = 2.5
!-----
ThetaL = 0.09482395
ThetaU = 0.1513344
TP = 0.2225801
NST = 8
n = 2.5

do i=0, 129
  axis1(i) = 0.336733741557961-0.000100794776112316*CSLOPE(i)-
0.0000833100341409037*RELIEF(i)&
    +0.00200161092985827*SB(i)+0.00342214718303411*SD(i)+0.092631888341564*SINU(i)&
    -0.348486383817856*HYPSO(i)-1.83374685029326E-06*Y(i)
  axis2(i) = 5.83302279672671-2.5368932042422E-08*HUC8(i)-0.00244299761143094*PERIM(i)&

```

```

-0.055034290300751*P100(i)-
0.000614906296416811*HIGHELEV(i)-0.136111136082763*LAT(i)&
+0.0013364310140715*SC(i)
axis3(i) = -0.0220003274509901-
0.799325726066741*WSLOPE(i)+0.000150496029141555*RELIEF(i)&
+4.83976979040975e-07*HUC14(i)+3.07570302630526e-
08*HUC8(i)+0.00272779276445727*SD(i)&
-0.0329041349558911*P100(i)+0.324484279451702*HYPSO(i)
if(i.NE.0)then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call BROIprob(n,NST,ThetaL,ThetaU,TP,size(gref,1)-1,size(gref,2),dij,gref,gprob,REFSID)
!print*,gprob
open(300,file='gbio.txt')
open(400,file='gref.txt')
do i=1,231
write(300,12) genusnames(i),gprob(0,i)
end do
write(400,*) 'StationID',ctab,'G_RefID'
do i =1,129
write(400,22) STATIONID(i),ctab,REFSID(i)
end do

end

```

!!=====

```

===
subroutine BROIprob(n,NST,ThetaL,ThetaU,TP,siteno,taxano,dij,ref,prob,REFSID)
integer siteno,taxano
real n,ThetaL,ThetaU,TP,TH
real,dimension(siteno)::dij
real,dimension(0:siteno,taxano)::ref,prob
real,dimension(siteno)::WF
integer,dimension(siteno)::REFSID
integer NS, NST

NS = 0
do i = 1,siteno
if(dij(i) .LE. ThetaL) then
NS = NS+1
end if
end do
TH = ThetaL
if(NS .LT. NST) then
TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1,siteno
WF(i) = 0
REFSID(i) = 0
if(dij(i) .LE. TH) then
WF(i) = 1.0- (dij(i)/TP)**n
REFSID(i) = 1
end if
prob(i,:) = WF(i)*ref(i,:)
end do
prob(0,:) = sum(prob,dim=1)/sum(WF)
return
end

```

```

!
! New Datum Region of Influence Method
! -- Selection - Exhaustive AIC
! -- Soil - Ragan
!
      program NROI_A_R

      character(len=15) :: STATIONID(0:129)
      character(len=1)  :: REF, ctab
      real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM , LENGTH , SA, SB, SC, SD, ELEV, SINU, P2 ,&
P100 , HIGHELEV, HYPISO, LAT, LONG, X, Y
      integer, dimension(0:129) :: HUC14, HUC8

      character(len=17) :: familynames(124)
      character(len=23) :: genusnames(231)
      real, dimension(0:129, 124) :: fref, fprob
      real, dimension(0:129, 231) :: gref, gprob

      real, dimension(0:129) :: axis1, axis2, axis3
      real, dimension(129) :: dij
      integer NST, REFSID(129)
      real n, ThetaL, ThetaU

      ctab = CHAR(9)
11  format(A17, F5.2)
12  format(A23, F5.2)
22  format(A15, A1, I5)

! input test site predictors
      open(10, file='testsitepre.txt', form='formatted', status='old')
      i=0
      read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)
      close(10)
      print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
      open(20, file='referencepre_R.txt', form='formatted', status='old')
      read(20, *)

      do i=1, 129
      read(20, *) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV(i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)
      !print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
! HIGHELEV(i), HYPISO(i), X(i), Y(i), LAT(i), LONG(i)
      enddo
      close(20)

! input reference site biodiversity composition
!Family level

```

```

open(30,file='referencefamily.txt')
read(30,*) STATIONID(0),REF,familynames
do i=1,129
read(30,*) STATIONID(i),REF,fref(i,:)
end do
!print*,fref

!Genus level
open(40,file='referencegenus.txt')
read(40,*) STATIONID(0),REF,genusnames
do i=1,129
read(40,*) STATIONID(i),REF,gref(i,:)
end do
!print*,gref

!=== Exhaustive AIC =====
!Family level 10 40 - 5
2.5
!---select the neighbors-----
!ThetaL = 10% of tatal dij data = 0.07397749
!ThetaU = 40% of taltal dij data = 0.1737273
!NST = 5
!n = 2.5
!-----
ThetaL = 0.07397749
ThetaU = 0.1737273
NST = 8
n = 2.5

do i=0,129
axis1(i) = 0.103501402891516+0.00142705981016488*SD(i)+0.060127473287618*SINU(i)-
0.208011303880480*HYPSO(i)&
+4.87448737813407e-07*X(i)-1.37533232634292e-06*Y(i)
axis2(i) = -4.68452707332487-7.94E-07*HUC14(i)-2.95E-08*HUC8(i)-0.091748923*SINU(i)&
+0.04670853*P100(i)-0.466234625*HYPSO(i)+0.123474652*LAT(i)
axis3(i) = 0.374721271895483-2.66836501038278e-08*HUC8(i)-
0.00362952990651171*LENGTH(i)-0.00127905771301773*SB(i)&
-0.0839121613151153*SINU(i)-3.56656830085609e-07*X(i)

if(i.NE.0)then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call TROIprob(n,NST,ThetaL,ThetaU,size(fref,1)-1,size(fref,2),dij,fref,fprob,REFSID)
!print*,fprob(0,:)
open(100,file='fbio.txt')
open(200,file='fref.txt')
do i=1,124
write(100,11) familynames(i),fprob(0,i)
end do
write(200,*) 'StationID',ctab,'F_RefID'
do i =1,129
write(200,22) STATIONID(i),ctab,REFSID(i)
end do

!Genus level 6 40 - 4 4
!---select the neighbors-----
!ThetaL = 6% of tatal dij data = 0.06953577
!ThetaU = 40% of taltal dij data = 0.1995810

```

```

!NST =4
!n = 4
!-----
ThetaL = 0.06953577
ThetaU = 0.1995810
NST = 4
n = 4

do i=0,129
  axis1(i) = 0.115378294083608-8.08205975236042e-05*CSLOPE(i)+0.08761340169703*SINU(i)-
0.413037614556601*HYPSO(i)&
  axis2(i) =0.601913285-0.010737439*LIME(i)-0.002350247*PERIM(i)+0.001017808*SC(i)-
0.001563192*HIGHELEV(i)&
  axis3(i) =3.151745421-0.839735418*WSLOPE(i)+0.000151981*RELIEF(i)+4.28E-
07*HUC14(i)+3.14E-08*HUC8(i)&
  dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call TROIprob(n,NST,ThetaL,ThetaU,size(gref,1)-1,size(gref,2),dij,gref,gprob,REFSID)
!print*,gprob
open(300,file='gbio.txt')
open(400,file='gref.txt')
do i=1,231
  write(300,12) genusnames(i),gprob(0,i)
end do
write(400,*) 'StationID',ctab,'G_RefID'
do i =1,129
  write(400,22) STATIONID(i),ctab,REFSID(i)
end do

end

```

!!=====

```

===
subroutine TROIprob(n,NST,ThetaL,ThetaU,siteno,taxano,dij,ref,prob,REFSID)
integer siteno,taxano
real n,ThetaL,ThetaU,TH
real,dimension(siteno)::dij
real,dimension(0:siteno,taxano)::ref,prob
real,dimension(siteno)::WF
integer,dimension(siteno)::REFSID
integer NS,NST

NS = 0
do i = 1,siteno
  if(dij(i) .LE. ThetaL) then
    NS = NS+1
  end if
end do
TH = ThetaL
if(NS .LT.NST) then
  TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1,siteno

```

```

        WF(i) = 1.0 - (di j(i)/TH)**n
        REFSID(i) = 1
        if(WF(i) .LT. 0) then
            WF(i) = 0
            REFSID(i) = 0
        end if
        prob(i,:) = WF(i)*ref(i,:)
    end do
    prob(0,:) = sum(prob, dim=1)/sum(WF)
    return
end

!
! New Datum Region of Influence Method
! -- Selection - Exhaustive AIC
! -- Soil - STATSGO
!
    program NROI_A_S

    character(len=15) :: STATIONID(0:129)
    character(len=1) :: REF, ctab
    real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
    PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
    P100, HIGHELEV, HYPSO, LAT, LONG, X, Y
    integer, dimension(0:129) :: HUC14, HUC8

    character(len=17) :: familynames(124)
    character(len=23) :: genusnames(231)
    real, dimension(0:129, 124) :: fref, fprob
    real, dimension(0:129, 231) :: gref, gprob

    real, dimension(0:129) :: axis1, axis2, axis3
    real, dimension(129) :: di j
    integer NST, REFSID(129)
    real n, ThetaL, ThetaU

    ctab = CHAR(9)
11 format(A17, F5.2)
12 format(A23, F5.2)
22 format(A15, A1, I5)

! input test site predictors
    open(10, file='testsitepre.txt', form='formatted', status='old')
    i=0
    read(10,*) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
    LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
    SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
    HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
    close(10)
    print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
    LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
    SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
    HIGHELEV(i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
    open(20, file='referencepre_S.txt', form='formatted', status='old')
    read(20,*)

    do i=1, 129
        read(20,*) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
        LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
        SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &

```



```

HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
!print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), TOTLENGTH (i), P2(i), P100(i), &
! HIGHELEV (i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
open(30, file='referencefamily.txt')
read(30, *) STATIONID(0), REF, familynames
do i=1, 129
read(30, *) STATIONID(i), REF, fref(i, :)
end do
!print*, fref

!Genus level
open(40, file='referencegenus.txt')
read(40, *) STATIONID(0), REF, genusnames
do i=1, 129
read(40, *) STATIONID(i), REF, gref(i, :)
end do
!print*, gref

!=== Exhaustive AIC =====
!Family level
!-----select the neighbors-----
!ThetaL = 8% of tatal dij data = 0.06204577
!ThetaU = 40% of taltal dij data = 0.1571108
!NST = 8
!n = 3.5
!-----
ThetaL = 0.06204577
ThetaU = 0.1571108
NST = 8
n = 3.5

do i=0, 129
axis1(i) = -0.228902566126975+0.00343300716142082*SD(i)+0.0630267442819524*SINU(i)&
+0.0494613662937794*P100(i)-0.19980945923722*HYPSO(i)-
8.37691526171816E-07*Y(i)
axis2(i) = -4.68452707332487-7.94063459434507E-07*HUC14(i)-2.94546884262961E-
08*HUC8(i)&
-0.0917489225780655*SINU(i)+0.0467085299984164*P100(i)-
0.466234624921827*HYPSO(i)&
+0.12347465193802*LAT(i)
axis3(i) = -2.93689970667502-2.84845195384243E-08*HUC8(i)-
0.00369569292494585*LENGTH(i)&
-0.0775318816116617*SINU(i)+0.0403280327951876*LONG(i)

if(i.NE.0) then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call TR0Iprob(n, NST, ThetaL, ThetaU, size(fref, 1)-1, size(fref, 2), dij, fref, fprob, REFSID)
!print*, fprob(0, :)
open(100, file='fbio.txt')
open(200, file='fref.txt')
do i=1, 124

```

```

        write(100,11) familynames(i), fprob(0, i)
    end do
    write(200,*) 'StationID', ctab, 'F_RefID'
    do i =1, 129
        write(200, 22) STATIONID(i), ctab, REFSID(i)
    end do

        !Genus level
        !---select the neighbors-----
        !ThetaL = 8% of tatal dij data = 0.0813133
        !ThetaU = 40% of taltal dij data = 0.1985606
        !NST = 6
        !n = 3
        !-----
        ThetaL = 0.0813133
        ThetaU = 0.1985606
        NST = 6
        n = 3

    do i=0, 129
        axis1(i) = 0.1153782940836-
0.0000808205975236045*CSLOPE(i)+0.0876134016970315*SINU(i)&
-0.413037614556604*HYPSO(i)+7.8951068300761E-07*X(i)-
1.43520199784076E-06*Y(i)
        axis2(i) = 5.27530254350311-3.1007633239593E-08*HUC8(i)-0.00252799995614199*PERIM(i)&
+0.00140204995993862*SC(i)-0.0355882192803853*P100(i)-
0.125858522038058*LAT(i)
        axis3(i) = -0.0220003274509901-
0.799325726066741*WSLOPE(i)+0.000150496029141555*RELIEF(i)&
+4.83976979040975e-07*HUC14(i)+3.07570302630526e-
08*HUC8(i)+0.00272779276445727*SD(i)&
-0.0329041349558911*P100(i)+0.324484279451702*HYPSO(i)

        if(i.NE.0) then
            dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
        end if
    end do

    call TROIprob(n, NST, ThetaL, ThetaU, size(gref, 1)-1, size(gref, 2), dij, gref, gprob, REFSID)
    !print*, gprob
    open(300, file='gbio.txt')
    open(400, file='gref.txt')
    do i=1, 231
        write(300, 12) genusnames(i), gprob(0, i)
    end do
    write(400,*) 'StationID', ctab, 'G_RefID'
    do i =1, 129
        write(400, 22) STATIONID(i), ctab, REFSID(i)
    end do

    end

```

```

!!=====
===
subroutine TROIprob(n, NST, ThetaL, ThetaU, siteno, taxano, dij, ref, prob, REFSID)
integer siteno, taxano
real n, ThetaL, ThetaU, TH
real, dimension(siteno) :: dij
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID

```

```

integer NS, NST

NS = 0
do i = 1, siteno
    if(dij(i) .LE. ThetaL) then
        NS = NS+1
    end if
end do
TH = ThetaL
if(NS .LT. NST) then
    TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1, siteno
    WF(i) = 1.0- (dij(i)/TH)**n
    REFSID(i) = 1
    if(WF(i) .LT. 0) then
        WF(i) = 0
        REFSID(i) = 0
    end if
    prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

!
! New Datum Region of Influence Method
! -- Selection - StepwiseAIC
! -- Soil - Ragan
!
program NROI_S_R

character(len=15) :: STATIONID(0:129)
character(len=1) :: REF, ctab
real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
P100, HIGHELEV, HYPSON, LAT, LONG, X, Y
integer, dimension(0:129) :: HUC14, HUC8

character(len=17) :: familynames(124)
character(len=23) :: genusnames(231)
real, dimension(0:129, 124) :: fref, fprob
real, dimension(0:129, 231) :: gref, gprob

real, dimension(0:129) :: axis1, axis2, axis3
real, dimension(129) :: dij
integer NST, REFSID(129)
real n, ThetaL, ThetaU

ctab = CHAR(9)
11 format(A17, F5.2)
12 format(A23, F5.2)
22 format (A15, A1, I5)

! input test site predictors
open(10, file='testsitepre.txt', form='formatted', status='old')
i=0
read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPSON(i), LAT(i), LONG(i), X(i), Y(i)

```

```

close(10)
print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
open(20, file='referencepre_R.txt', form='formatted', status='old')
read(20, *)

do i=1, 129
read(20, *) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
!print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
! HIGHELEV (i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
enddo
close(20)

! input reference site biodiversity composition
!Family level
open(30, file='referencefamily.txt')
read(30, *) STATIONID(0), REF, familynames
do i=1, 129
read(30, *) STATIONID(i), REF, fref(i, :)
end do
!print*, fref

!Genus level
open(40, file='referencegenus.txt')
read(40, *) STATIONID(0), REF, genusnames
do i=1, 129
read(40, *) STATIONID(i), REF, gref(i, :)
end do
!print*, gref

!=== Stepwise AIC=====
!Family level 10 30 - 8
2.5
!---select the neighbors-----
!ThetaL = 10% of tatal dij data =0.0738206
!ThetaU = 30% of taltal dij data =0.1431516
!NST = 8
!n = 2.5
!-----
ThetaL =0.0738206
ThetaU =0.1431516
NST = 8
n = 2.5

do i=0, 129
axis1(i) = 0.103501402891516+0.060127473287618*SINU(i)-0.208011303880481*HYPSO(i)&
+4.87448737813407e-07*X(i)-1.37533232634292e-
06*Y(i)+0.00142705981016488*SD(i)
axis2(i) = -4.38155411892806+0.0119996882743646*LIME(i)-9.1459667141927e-07*HUC14(i)&
-2.92298312690414e-08*HUC8(i)-
0.090724612776598*SINU(i)+0.045777467969937*P100(i)&
-0.42772973937687*HYPSO(i)+0.115512655619171*LAT(i)
axis3(i) = 0.374721271895483-2.66836501038278e-08*HUC8(i)-0.00127905771301773*SB(i)&

```

```

-0.0839121613151153*SINU(i)-3.56656830085609e-07*X(i)-
0.00362952990651171*LENGTH(i)

      if(i.NE.0) then
          dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
      end if
end do

call TR0Iprob(n,NST,ThetaL,ThetaU,size(fref,1)-1,size(fref,2),dij,fref,fprob,REFSID)
! print*,fprob(0,:)
open(100,file='fbio.txt')
open(200,file='fref.txt')
do i=1,124
    write(100,11) familynames(i),fprob(0,i)
end do
write(200,*) 'StationID',ctab,'F_RefID'
do i=1,129
    write(200,22) STATIONID(i),ctab,REFSID(i)
end do

!Genus level                                4      30      -      8
3.5
!---select the neighbors-----
!ThetaL = 4% of tatal dij data = 0.05528094
!ThetaU = 30% of taltal dij data = 0.1655731
!NST = 8
!n = 3.5
!-----
ThetaL = 0.05528094
ThetaU = 0.1655731
NST = 8
n = 3.5

do i=0,129
    axis1(i) = 0.115378294083608-8.08205975236042e-05*CSLOPe(i)+0.08761340169703*SINU(i)&
-0.413037614556601*HYPSO(i)+7.89510683007598e-07*X(i)-
1.43520199784078e-06*Y(i)
    axis2(i) = 0.601372207-0.014604154*LIME(i)+2.87E-07*HUC14(i)-0.002184504*PERIM(i)&
+0.000940113*SC(i)-0.001476468*HIGHELEV(i)-9.95E-07*X(i)-
1.33E-06*Y(i)
    axis3(i) = 3.151745421-0.839735418*WSLOPE(i)+0.000151981*RELIEF(i)+4.28E-07*HUC14(i)&
+3.14E-08*HUC8(i)-
0.041017848*P100(i)+0.238700932*HYPSO(i)-0.077209157*LAT(i)

      if(i.NE.0) then
          dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
      end if
end do

call TR0Iprob(n,NST,ThetaL,ThetaU,size(gref,1)-1,size(gref,2),dij,gref,gprob,REFSID)
!print*,gprob
open(300,file='gbio.txt')
open(400,file='gref.txt')
do i=1,231
    write(300,12) genusnames(i),gprob(0,i)
end do
write(400,*) 'StationID',ctab,'G_RefID'
do i=1,129
    write(400,22) STATIONID(i),ctab,REFSID(i)
end do

```

```

end

!!=====
===
subroutine TROIprob(n, NST, ThetaL, ThetaU, siteno, taxano, di j, ref, prob, REFSID)
integer siteno, taxano
real n, ThetaL, ThetaU, TH
real, dimension(siteno) :: di j
real, dimension(0:siteno, taxano) :: ref, prob
real, dimension(siteno) :: WF
integer, dimension(siteno) :: REFSID
integer NS, NST

NS = 0
do i = 1, siteno
    if(di j(i) .LE. ThetaL) then
        NS = NS+1
    end if
end do
TH = ThetaL
if(NS .LT. NST) then
    TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1, siteno
    WF(i) = 1.0- (di j(i)/TH)**n
    REFSID(i) = 1
    if(WF(i) .LT. 0) then
        WF(i) = 0
        REFSID(i) = 0
    end if
    prob(i, :) = WF(i)*ref(i, :)
end do
prob(0, :) = sum(prob, dim=1)/sum(WF)
return
end

!
! New Datum Region of Influence Method
! -- Selection - StepwiseAIC
! -- Soil - STATSGO
!
program NROI_S_S

character(len=15) :: STATIONID(0:129)
character(len=1) :: REF, ctab
real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM, LENGTH, SA, SB, SC, SD, ELEV, SINU, P2, &
P100, HIGHELEV, HYPISO, LAT, LONG, X, Y
integer, dimension(0:129) :: HUC14, HUC8

character(len=17) :: familynames(124)
character(len=23) :: genusnames(231)
real, dimension(0:129, 124) :: fref, fprob
real, dimension(0:129, 231) :: gref, gprob

real, dimension(0:129) :: axis1, axis2, axis3
real, dimension(129) :: di j
integer NST, REFSID(129)
real n, ThetaL, ThetaU

```

```

        ctab = CHAR(9)
11  format (A17,F5.2)
12  format (A23,F5.2)
22  format (A15,A1,I5)

! input test site predictors
  open(10,file='testsitepre.txt',form='formatted',status='old')
  i=0
  read(10,*) AREA(i),WSLOPE(i),CSLOPE(i),RELIEF(i),&
  LIME(i),HUC14(i),HUC8(i),PERIM(i),LENGTH(i),SA(i),SB(i),&
  SC(i),SD(i),ELEV(i),SINU(i),P2(i),P100(i),&
  HIGHELEV(i),HYPSO(i),LAT(i),LONG(i),X(i),Y(i)
  close(10)
!print*,STATIONID(i),REF,AREA(i),WSLOPE(i),CSLOPE(i),RELIEF(i),&
LIME(i),HUC14(i),HUC8(i),PERIM(i),LENGTH(i),SA(i),SB(i),&
SC(i),SD(i),ELEV(i),SINU(i),P2(i),P100(i),&
HIGHELEV(i),HYPSO(i),LAT(i),LONG(i),X(i),Y(i)

! input reference sites predictors
  open(20,file='referencepre_S.txt',form='formatted',status='old')
  read(20,*)

  do i=1,129
    read(20,*) STATIONID(i),REF,AREA(i),WSLOPE(i),CSLOPE(i),RELIEF(i),&
    LIME(i),HUC14(i),HUC8(i),PERIM(i),LENGTH(i),SA(i),SB(i),&
    SC(i),SD(i),ELEV(i),SINU(i),P2(i),P100(i),&
    HIGHELEV(i),HYPSO(i),LAT(i),LONG(i),X(i),Y(i)
    !print*,STATIONID(i),REF,AREA(i),WSLOPE(i),CSLOPE(i),RELIEF(i),&
    ! LIME(i),HUC14(i),HUC8(i),PERIM(i),LENGTH(i),SA(i),SB(i),&
    ! SC(i),SD(i),ELEV(i),SINU(i),TOTLENGTH(i),P2(i),P100(i),&
    ! HIGHELEV(i),HYPSO(i),X(i),Y(i),LAT(i),LONG(i)
  enddo
  close(20)

! input reference site biodiversity composition
!Family level
  open(30,file='referencefamily.txt')
  read(30,*) STATIONID(0),REF,familynames
  do i=1,129
    read(30,*) STATIONID(i),REF,fref(i,:)
  end do
  !print*,fref

!Genus level
  open(40,file='referencegenus.txt')
  read(40,*) STATIONID(0),REF,genusnames
  do i=1,129
    read(40,*) STATIONID(i),REF,gref(i,:)
  end do
  !print*,gref

!=== Stepwise AIC=====
!Family level
!-----select the neighbors-----
!ThetaL = 8% of tatal dij data = 0.06235811
!ThetaU = 70% of taltal dij data = 0.2460933
!NST = 4
!n = 4
!-----
ThetaL = 0.06235811
ThetaU = 0.2460933
NST = 4

```

```

n = 4

do i=0, 129
    axis1(i) = -0.228902566126975+0.00343300716142082*SD(i)+0.0630267442819524*SINU(i)&
    +0.0494613662937794*P100(i)-0.19980945923722*HYPSO(i)-
8.37691526171816e-07*Y(i)
    axis2(i) = -4.38155411892814+0.0119996882743647*LIME(i)-9.14596671419284e-
07*HUC14(i)&
    -2.92298312690420e-08*HUC8(i)-
0.0907246127766004*SINU(i)+0.045777467969939*P100(i)&
    -0.427729739376867*HYPSO(i)+0.115512655619173*LAT(i)
    axis3(i) = 0.357058326218296-2.80197266756115E-08*HUC8(i)-
0.00122024560082751*PERIM(i)&
    -0.0802674660348683*SINU(i)-4.60917004462554E-07*X(i)

    if(i.NE.0) then
        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
    end if
end do

call TROIprob(n,NST,ThetaL,ThetaU,size(fref,1)-1,size(fref,2),dij,fref,fprob,REFSID)
print*,fprob(0,:)
!
open(100,file='fbio.txt')
open(200,file='fref.txt')
do i=1,124
    write(100,11) familynames(i),fprob(0,i)
end do
write(200,*) 'StationID',ctab,'F_RefID'
do i =1,129
    write(200,22) STATIONID(i),ctab,REFSID(i)
end do

!Genus level
!-----select the neighbors-----
!ThetaL = 8% of tatal dij data = 0.08591814
!ThetaU = 60% of taltal dij data = 0.2541870
!NST = 4
!n = 3.5
!-----
ThetaL = 0.08591814
ThetaU = 0.2541870
NST = 4
n = 3.5

do i=0, 129
    axis1(i) = 0.336733741557961-0.000100794776112316*CSLOPE(i)-
0.0000833100341409037*RELIEF(i)&
    +0.00200161092985827*SB(i)+0.00342214718303411*SD(i)+0.092631888341564*SINU(i)&
    -0.348486383817856*HYPSO(i)-1.83374685029326E-06*Y(i)
    axis2(i) = 5.83302279672671-2.5368932042422E-08*HUC8(i)-0.00244299761143094*PERIM(i)&
    -0.055034290300751*P100(i)-
0.000614906296416811*HIGHELEV(i)-0.136111136082763*LAT(i)&
    +0.0013364310140715*SC(i)
    axis3(i) = -0.0220003274509901-
0.799325726066741*WSLOPE(i)+0.000150496029141555*RELIEF(i)&
    +4.83976979040975e-07*HUC14(i)+3.07570302630526e-
08*HUC8(i)+0.00272779276445727*SD(i)&
    -0.0329041349558911*P100(i)+0.324484279451702*HYPSO(i)

    if(i.NE.0) then

```



```

        dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
        end if
    end do

    call TROIprob(n,NST,ThetaL,ThetaU,size(gref,1)-1,size(gref,2),dij,gref,gprob,REFSID)
    !print*,gprob
    open(300,file='gbio.txt')
    open(400,file='gref.txt')
    do i=1,231
        write(300,12) genusnames(i),gprob(0,i)
    end do
    write(400,*) 'StationID',ctab,'G_RefID'
    do i=1,129
        write(400,22) STATIONID(i),ctab,REFSID(i)
    end do

end

end

!!=====
===
subroutine TROIprob(n,NST,ThetaL,ThetaU,siteno,taxano,dij,ref,prob,REFSID)
integer siteno,taxano
real n,ThetaL,ThetaU,TH
real,dimension(siteno)::dij
real,dimension(0:siteno,taxano)::ref,prob
real,dimension(siteno)::WF
integer,dimension(siteno)::REFSID
integer NS,NST

NS=0
do i=1,siteno
    if(dij(i).LE.ThetaL) then
        NS=NS+1
    end if
end do
TH=ThetaL
if(NS.LT.NST) then
    TH=ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i=1,siteno
    WF(i)=1.0-(dij(i)/TH)**n
    REFSID(i)=1
    if(WF(i).LT.0) then
        WF(i)=0
        REFSID(i)=0
    end if
    prob(i,:)=WF(i)*ref(i,:)
end do
prob(0,:)=sum(prob,dim=1)/sum(WF)
return
end

!
! New Datum Region of Influence Method
! -- Selection - Exhaustive BIC
! -- Soil - Ragan
!
program NROI_B_R

character(len=15)::STATIONID(0:129)

```

```

        character(len=1) :: REF, ctab
        real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM , LENGTH , SA, SB, SC, SD, ELEV, SINU, P2 , &
P100 , HIGHELEV, HYPISO, LAT, LONG, X, Y
        integer, dimension(0:129) :: HUC14, HUC8

        character(len=17) :: familynames(124)
        character(len=23) :: genusnames(231)
        real, dimension(0:129, 124) :: fref, fprob
        real, dimension(0:129, 231) :: gref, gprob

        real, dimension(0:129) :: axis1, axis2, axis3
        real, dimension(129) :: dij
        integer NST, REFSID(129)
        real n, ThetaL, ThetaU

        ctab = CHAR(9)
11 format (A17, F5.2)
12 format (A23, F5.2)
22 format (A15, A1, I5)

! input test site predictors
        open(10, file='testsitepre.txt', form='formatted', status='old')
        i=0
        read(10,*) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)
        close(10)
print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
        open(20, file='referencepre_R.txt', form='formatted', status='old')
        read(20,*)

        do i=1, 129
        read(20,*) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPISO(i), LAT(i), LONG(i), X(i), Y(i)
        !print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
! HIGHELEV (i), HYPISO(i), X(i), Y(i), LAT(i), LONG(i)
        enddo
        close(20)

! input reference site biodiversity composition
!Family level
        open(30, file='referencefamily.txt')
        read(30,*) STATIONID(0), REF, familynames
        do i=1, 129
        read(30,*) STATIONID(i), REF, fref(i, :)
        end do
        !print*, fref

!Genus level
        open(40, file='referencegenus.txt')
        read(40,*) STATIONID(0), REF, genusnames

```

```

do i=1,129
read(40,*) STATIONID(i),REF, gref(i,:)
end do
!print*, gref

!=== Exhaustive BIC =====
!Family level 6 30 - 5
3
!---select the neighbors-----
!ThetaL = 6% of tatal dij data = 0.04776626
!ThetaU = 30% of taltal dij data = 0.1296230
!NST = 5
!n = 3
!-----
ThetaL = 0.04776626
ThetaU = 0.1296230
NST = 5
n = 3

do i=0,129
axis1(i) = 0.221544784378139-0.263036840454740*HYPSO(i)+5.46892793635458e-07*X(i)&
-1.41870482455244e-06*Y(i)
axis2(i) = -4.68452707332478-7.94063459434495e-07*HUC14(i)-2.94546884262955e-
08*HUC8(i)&
-0.091748922578063*SINU(i)+0.0467085299984162*P100(i)-
0.466234624921829*HYPSO(i)&
+0.123474651938017*LAT(i)
axis3(i) = 0.209020011793382-0.00161840545458612*SB(i)-0.102125758799106*SINU(i)

if(i.NE.0)then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call TR0Iprob(n,NST,ThetaL,ThetaU,size(fref,1)-1,size(fref,2),dij,fref,fprob,REFSID)
!print*,fprob(0,:)
open(100,file='fbio.txt')
open(200,file='fref.txt')
do i=1,124
write(100,11) familynames(i),fprob(0,i)
end do
write(200,*) 'StationID',ctab,'F_RefID'
do i=1,129
write(200,22) STATIONID(i),ctab,REFSID(i)
end do

!Genus level 8 30 - 5
4
!---select the neighbors-----
!ThetaL = 8% of tatal dij data = 0.06860956
!ThetaU = 30% of taltal dij data = 0.1494163
!NST = 5
!n = 4
!-----
ThetaL = 0.06860956
ThetaU = 0.1494163
NST = 5
n = 4

do i=0,129
axis1(i) = 0.104150442366716+0.098467415429256*SINU(i)-0.440648301712622*HYPSO(i)&

```

```

+8.23145543481247e-07*X(i)-1.51250604759304e-06*Y(i)
axis2(i) = 0.810549688-0.002427075*PERIM(i)-0.001691071*HIGHELEV(i)-1.24E-06*X(i)&
-1.64E-06*Y(i)
axis3(i) = 2.760969421+4.00E-08*HUC8(i)-0.07295503*LAT(i)

if(i.NE.0)then
    dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call TROIprob(n,NST,ThetaL,ThetaU,size(gref,1)-1,size(gref,2),dij,gref,gprob,REFSID)
!print*,gprob
open(300,file='gbio.txt')
open(400,file='gref.txt')
do i=1,231
    write(300,12) genusnames(i),gprob(0,i)
end do
write(400,*) 'StationID',ctab,'G_RefID'
do i =1,129
    write(400,22) STATIONID(i),ctab,REFSID(i)
end do

end

!!=====
===
subroutine TROIprob(n,NST,ThetaL,ThetaU,siteno,taxano,dij,ref,prob,REFSID)
integer siteno,taxano
real n,ThetaL,ThetaU,TH
real,dimension(siteno)::dij
real,dimension(0:siteno,taxano)::ref,prob
real,dimension(siteno)::WF
integer,dimension(siteno)::REFSID
integer NS,NST

NS = 0
do i = 1,siteno
    if(dij(i).LE.ThetaL)then
        NS = NS+1
    end if
end do
TH = ThetaL
if(NS.LT.NST)then
    TH = ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i = 1,siteno
    WF(i) = 1.0-(dij(i)/TH)**n
    REFSID(i) = 1
    if(WF(i).LT.0)then
        WF(i) = 0
        REFSID(i) = 0
    end if
    prob(i,:) = WF(i)*ref(i,:)
end do
prob(0,:) = sum(prob,dim=1)/sum(WF)
return
end

!
! New Datum Region of Influence Method

```

```

! -- Selection - Exhaustive BIC
! -- Soil - STATSGO
!
      program NROI_B_S

      character(len=15) :: STATIONID(0:129)
      character(len=1) :: REF, ctab
      real, dimension(0:129) :: AREA, WSLOPE, CSLOPE, RELIEF, LIME, &
PERIM , LENGTH , SA, SB, SC, SD, ELEV, SINU, P2 ,&
P100 , HIGHELEV, HYPSO, LAT, LONG, X, Y
      integer, dimension(0:129) :: HUC14, HUC8

      character(len=17) :: familynames(124)
      character(len=23) :: genusnames(231)
      real, dimension(0:129, 124) :: fref, fprob
      real, dimension(0:129, 231) :: gref, gprob

      real, dimension(0:129) :: axis1, axis2, axis3
      real, dimension(129) :: dij
      integer NST, REFSID(129)
      real n, ThetaL, ThetaU

      ctab = CHAR(9)
11 format(A17, F5.2)
12 format(A23, F5.2)
22 format (A15, A1, I5)

! input test site predictors
      open(10, file='testsitepre.txt', form='formatted', status='old')
      i=0
      read(10, *) AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
      close(10)
      print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)

! input reference sites predictors
      open(20, file='referencepre_S.txt', form='formatted', status='old')
      read(20, *)

      do i=1, 129
      read(20, *) STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
SC(i), SD(i), ELEV(i), SINU(i), P2(i), P100(i), &
HIGHELEV (i), HYPSO(i), LAT(i), LONG(i), X(i), Y(i)
      !print*, STATIONID(i), REF, AREA(i), WSLOPE(i), CSLOPE(i), RELIEF(i), &
! LIME(i), HUC14(i), HUC8(i), PERIM(i), LENGTH(i), SA(i), SB(i), &
! SC(i), SD(i), ELEV(i), SINU(i), TOTLENGTH (i), P2(i), P100(i), &
! HIGHELEV (i), HYPSO(i), X(i), Y(i), LAT(i), LONG(i)
      enddo
      close(20)

! input reference site biodiversity composition
!Family level
      open(30, file='referencefamily.txt')
      read(30, *) STATIONID(0), REF, familynames
      do i=1, 129
      read(30, *) STATIONID(i), REF, fref(i, :)

```

```

end do
!print*, fref

!Genus level
open(40, file='referencegenus.txt')
read(40, *) STATIONID(0), REF, genusnames
do i=1, 129
read(40, *) STATIONID(i), REF, gref(i, :)
end do
!print*, gref

!=== Exhaustive BIC =====
!Family level          6      30      -      7      2
!---select the neighbors-----
!ThetaL = 6% of tatal dij data = 0.03705849
!ThetaU = 30% of taltal dij data = 0.1152541
!NST = 7
!n = 2
!-----
ThetaL = 0.03705849
ThetaU = 0.1152541
NST = 7
n = 2

do i=0, 129
axis1(i) = -0.527670347764048+0.00481027605204689*SD(i)+0.0650310792558663*P100(i)
axis2(i) = 0.49821511574707-5.78352960879374e-07*HUC14(i)-0.00208005037882994*SC(i)&
-0.0883878663251916*SINU(i)-0.465248015085819*HYPSO(i)
axis3(i) = -3.22457693563322-3.25476880833444E-08*HUC8(i)+0.0425161195200811*LONG(i)

if(i.NE.0) then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call TROIprob(n, NST, ThetaL, ThetaU, size(fref, 1)-1, size(fref, 2), dij, fref, fprob, REFSID)
!print*, fprob(0, :)
open(100, file='fbio.txt')
open(200, file='fref.txt')
do i=1, 124
write(100, 11) familynames(i), fprob(0, i)
end do
write(200, *) 'StationID', ctab, 'F_RefID'
do i =1, 129
write(200, 22) STATIONID(i), ctab, REFSID(i)
end do

!Genus level          8      50      -      5      2.5
!---select the neighbors-----
!ThetaL = 8% of tatal dij data = 0.06795401
!ThetaU = 50% of taltal dij data = 0.2112865
!NST = 5
!n = 2.5
!-----
ThetaL = 0.06795401
ThetaU = 0.2112865
NST = 5
n = 2.5

do i=0, 129
axis1(i) = 0.104150442366708+0.0984674154292574*SINU(i)-0.440648301712624*HYPSO(i)&

```

```

+8.23145543481259e-07*X(i)-1.51250604759303e-06*Y(i)
axis2(i) = 7.244994821464-2.80308081073845e-08*HUC8(i)-0.00261543850117834*PERIM(i)&
-0.0614198266786265*P100(i)-0.169840908042419*LAT(i)
axis3(i) = -0.271171913164639+3.73799793023689e-08*HUC8(i)+0.00263993607998206*SD(i)&
+0.27408320806472*HYPSO(i)
if(i.NE.0)then
dij(i)=sqrt(((axis1(i)-axis1(0))**2)+((axis2(i)-axis2(0))**2)+((axis3(i)-
axis3(0))**2))
end if
end do

call TROIprob(n,NST,ThetaL,ThetaU,size(gref,1)-1,size(gref,2),dij,gref,gprob,REFSID)
!print*,gprob
open(300,file='gbio.txt')
open(400,file='gref.txt')
do i=1,231
write(300,12)genusnames(i),gprob(0,i)
end do
write(400,*)'StationID',ctab,'G_RefID'
do i=1,129
write(400,22)STATIONID(i),ctab,REFSID(i)
end do

end

```

!!=====

```

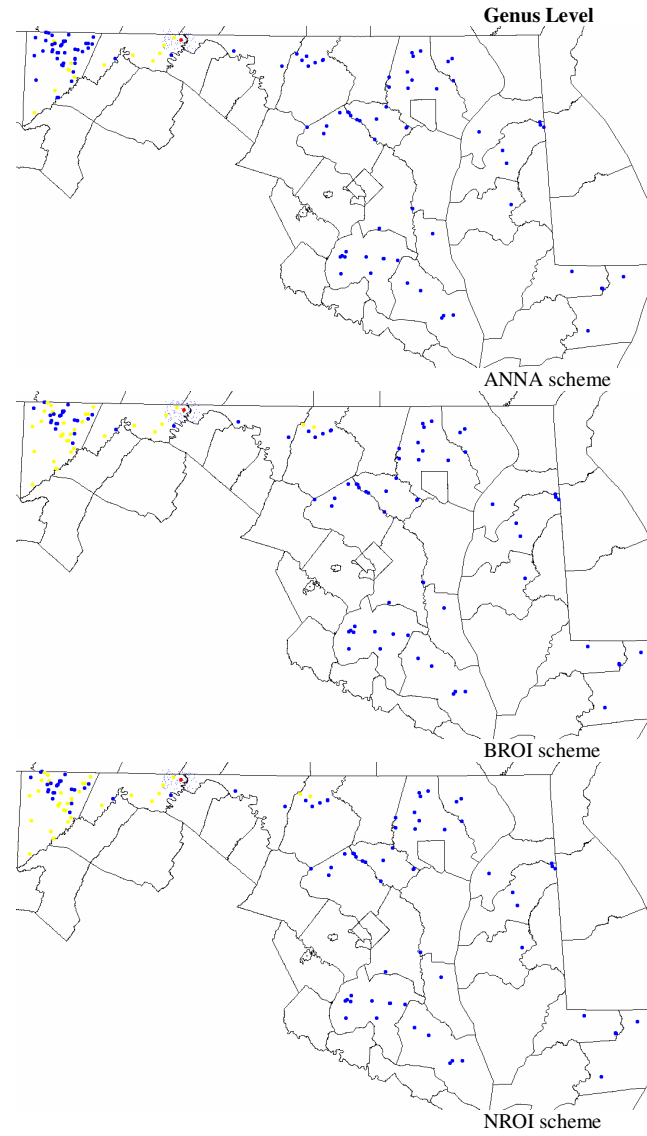
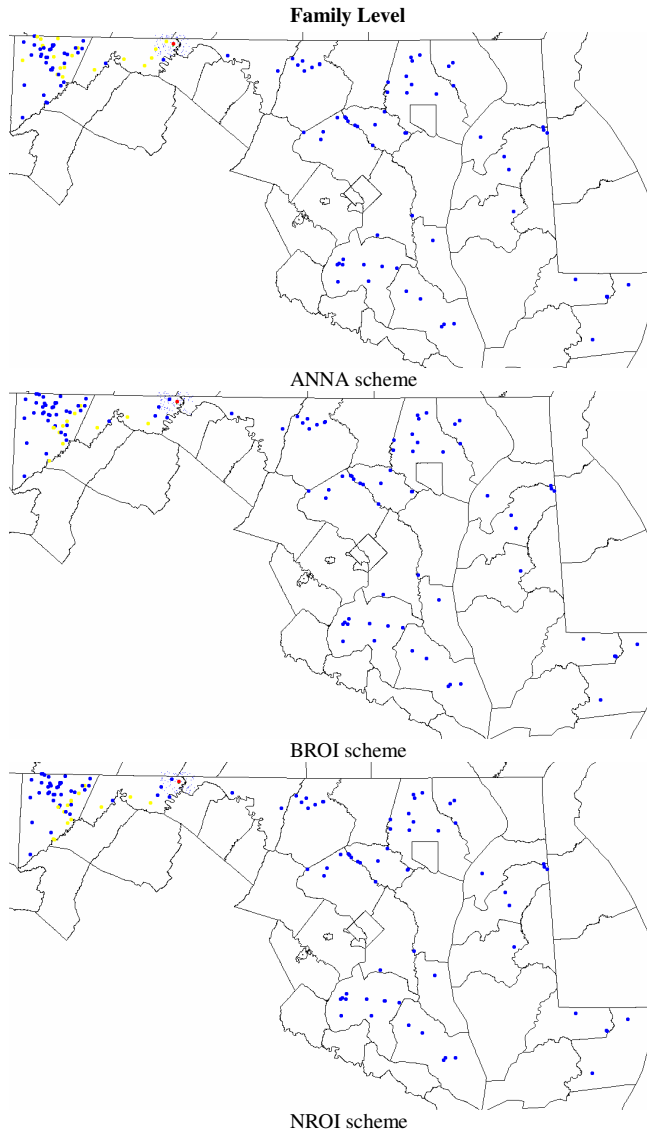
===
subroutine TROIprob(n,NST,ThetaL,ThetaU,siteno,taxano,dij,ref,prob,REFSID)
integer siteno,taxano
real n,ThetaL,ThetaU,TH
real,dimension(siteno)::dij
real,dimension(0:siteno,taxano)::ref,prob
real,dimension(siteno)::WF
integer,dimension(siteno)::REFSID
integer NS,NST

NS=0
do i=1,siteno
if(dij(i).LE.ThetaL)then
NS=NS+1
end if
end do
TH=ThetaL
if(NS.LT.NST)then
TH=ThetaL+(ThetaU-ThetaL)*(NST-NS)/NST
endif
do i=1,siteno
WF(i)=1.0-(dij(i)/TH)**n
REFSID(i)=1
if(WF(i).LT.0)then
WF(i)=0
REFSID(i)=0
end if
prob(i,:)=WF(i)*ref(i,:)
end do
prob(0,:)=sum(prob,dim=1)/sum(WF)
return
end

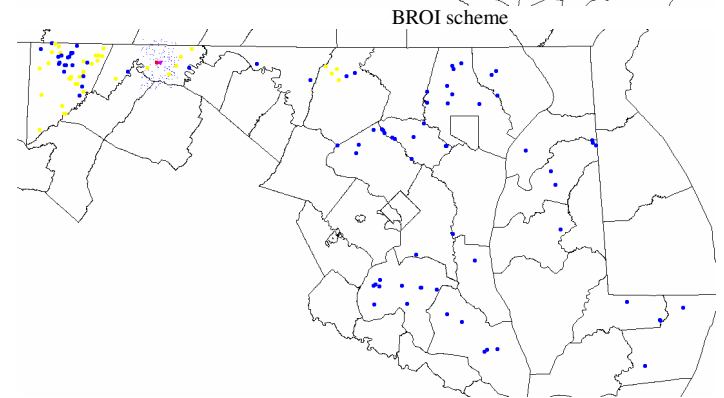
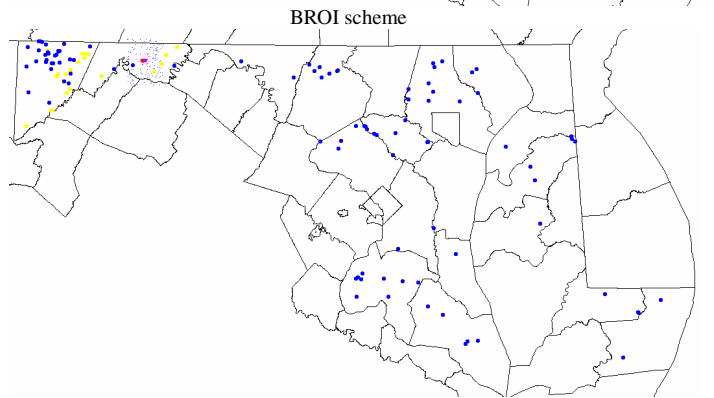
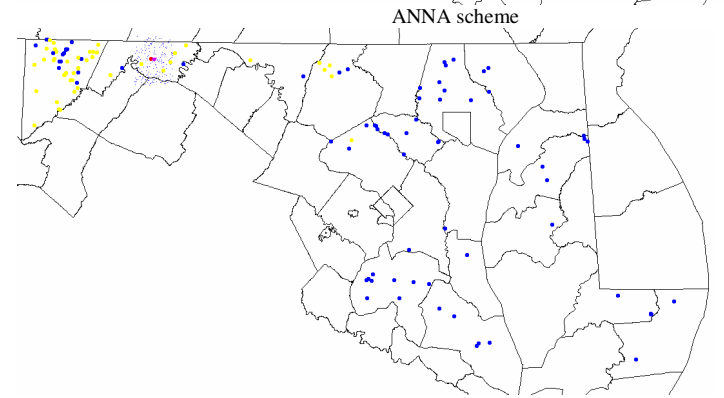
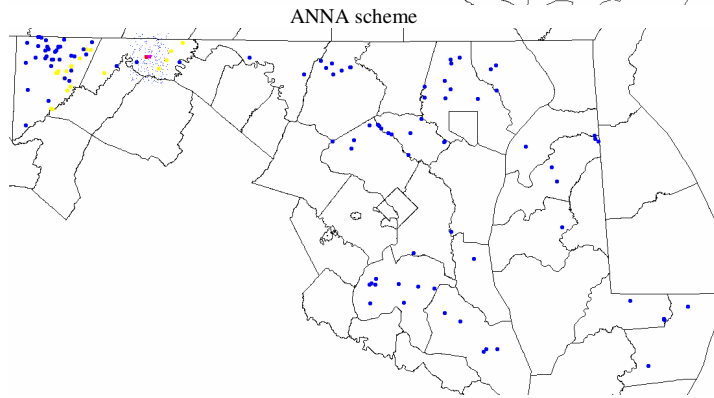
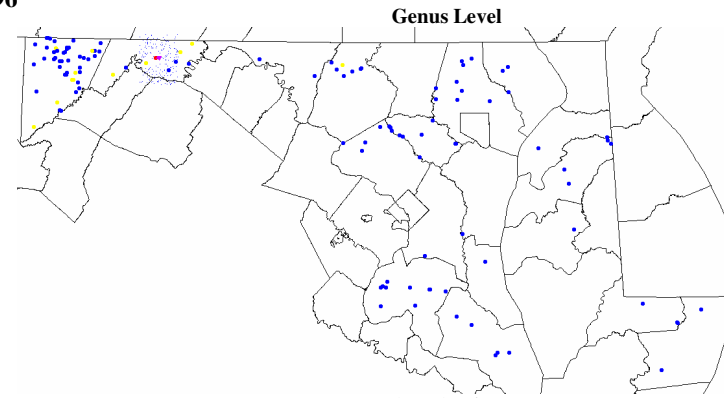
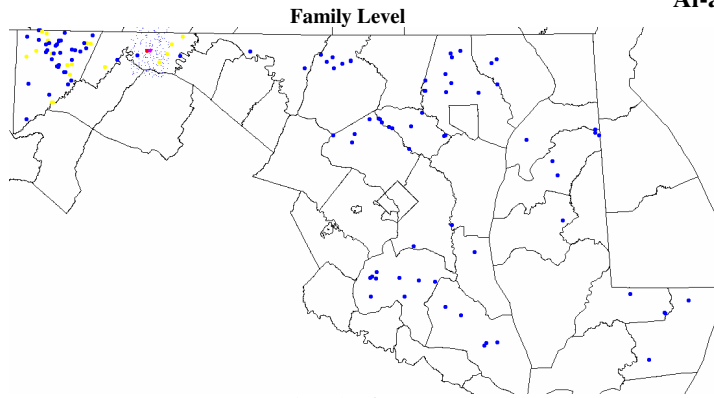
```

Appendix D. 27 Validation Reference Sites

AI-a-318-126-95

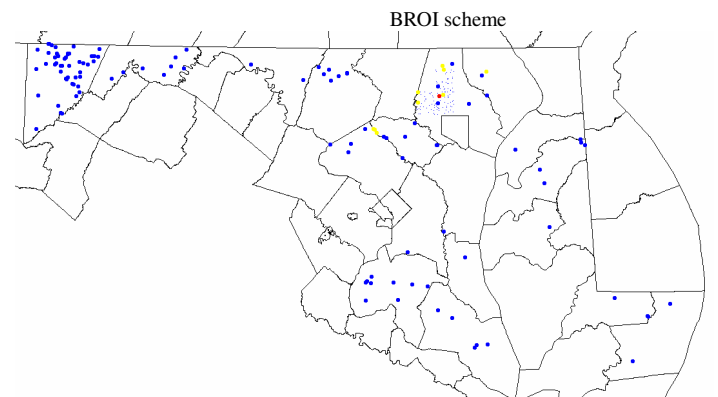
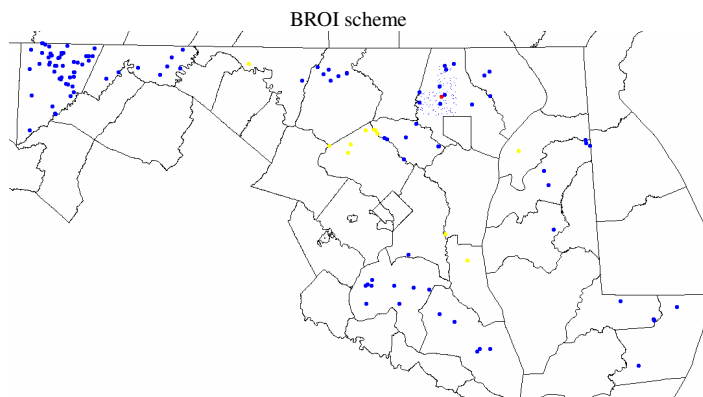
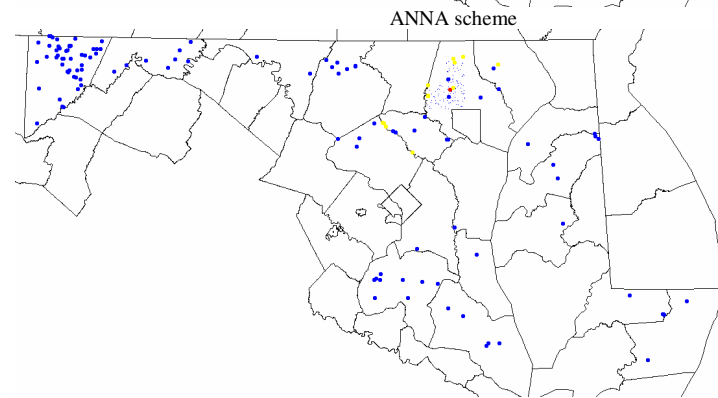
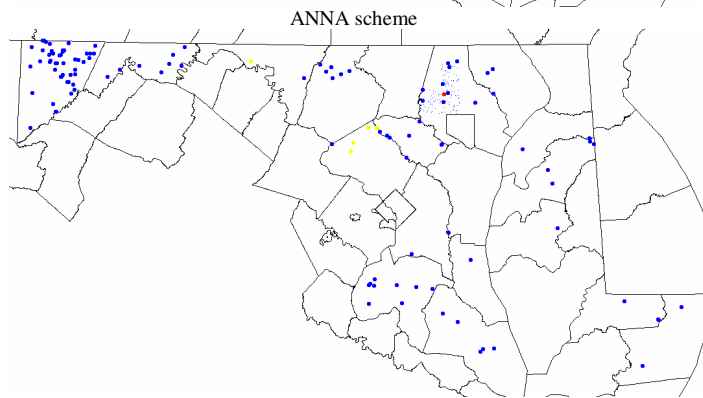
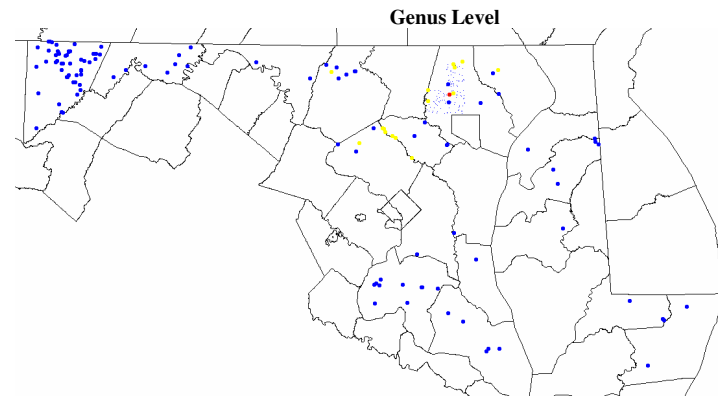
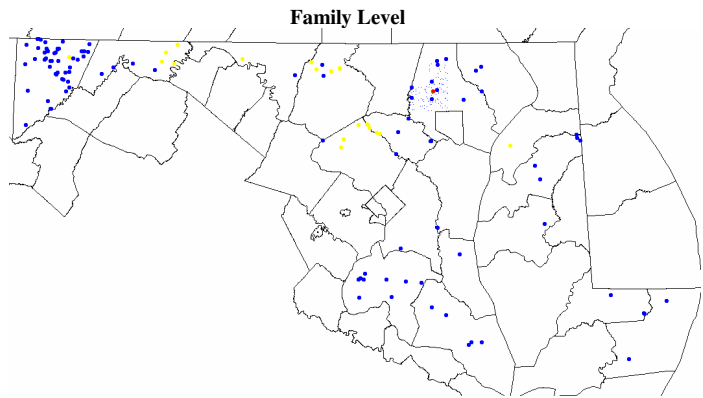


AI-a-480-205-96



NROI scheme

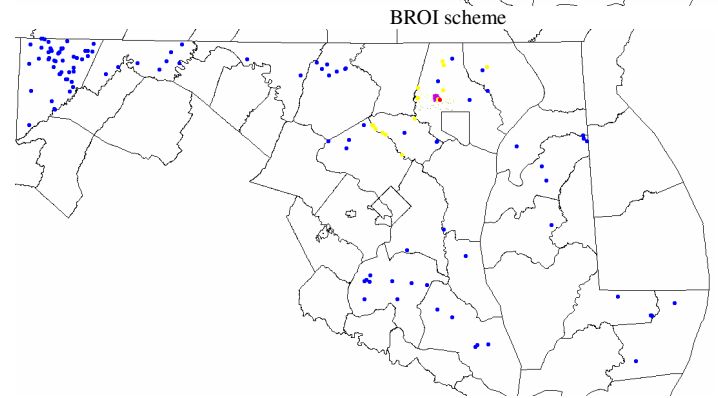
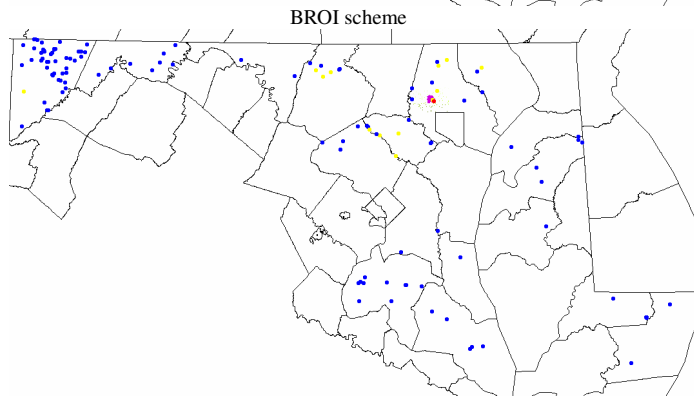
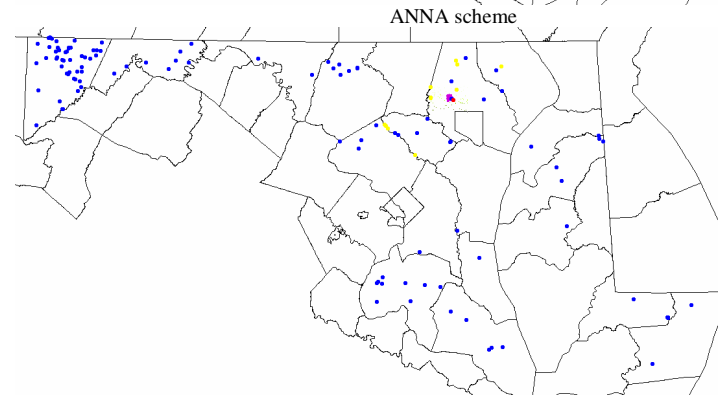
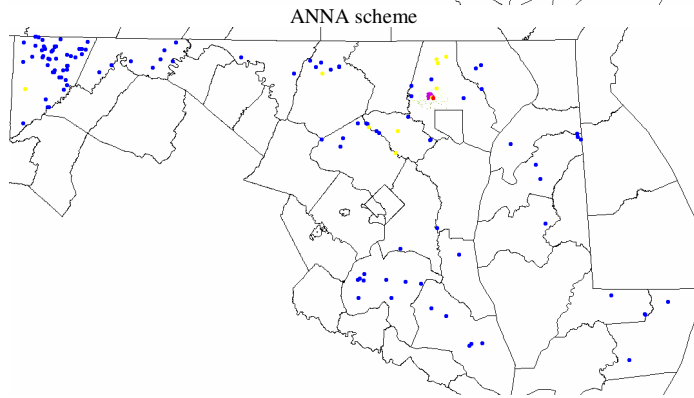
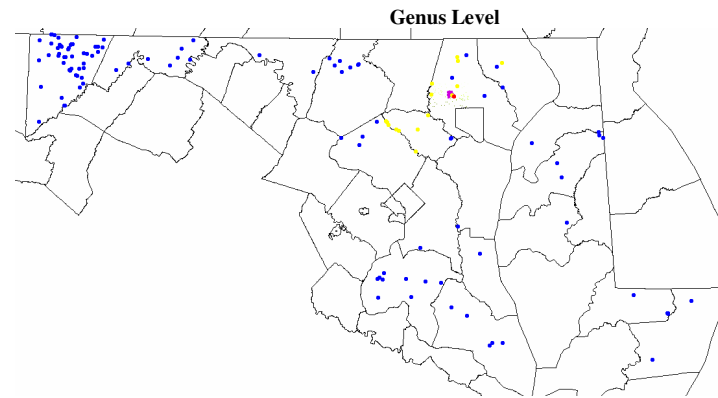
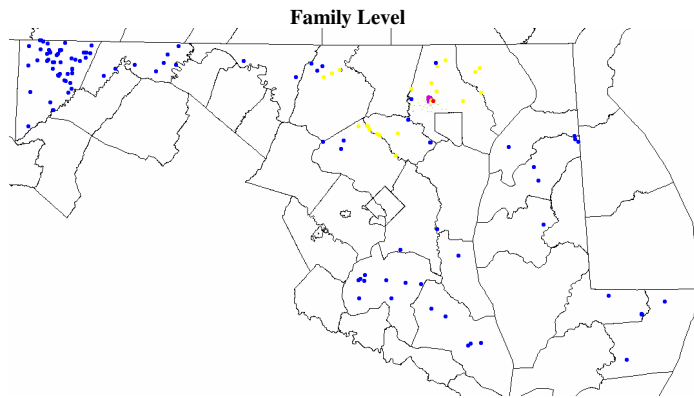
NROI scheme



NROI scheme

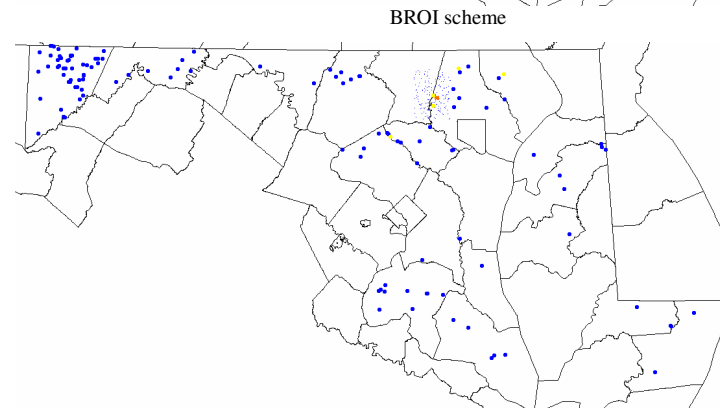
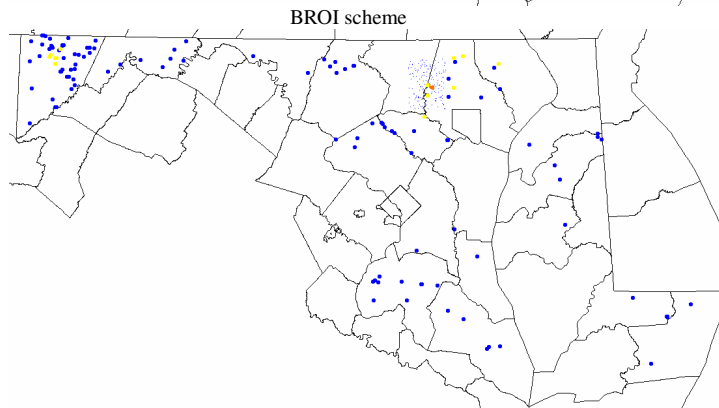
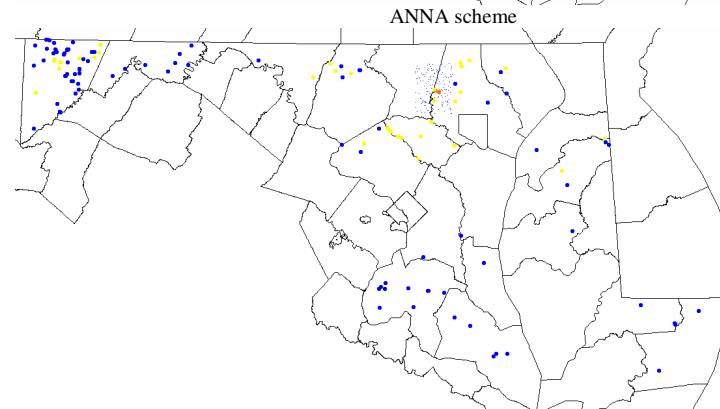
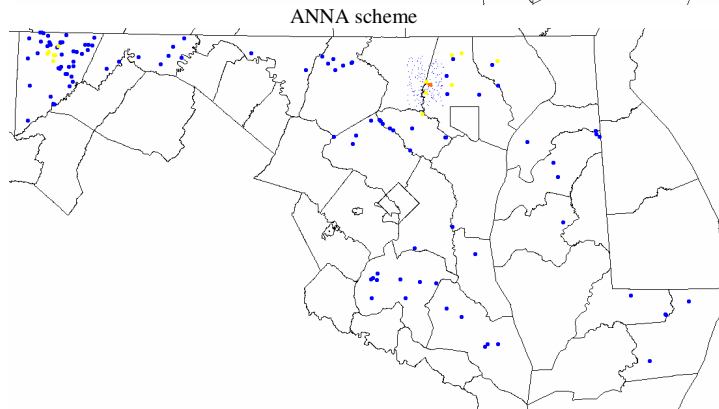
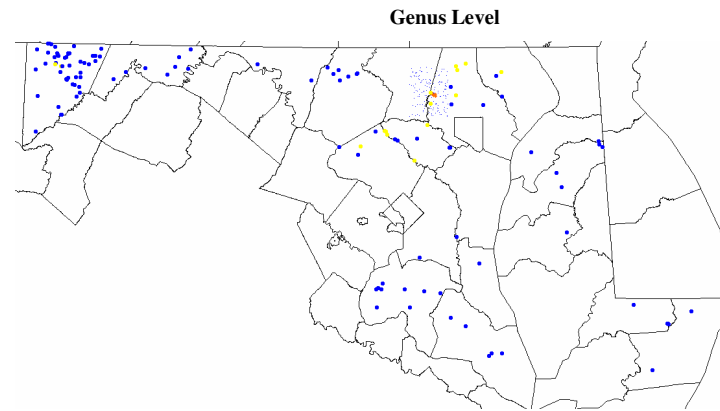
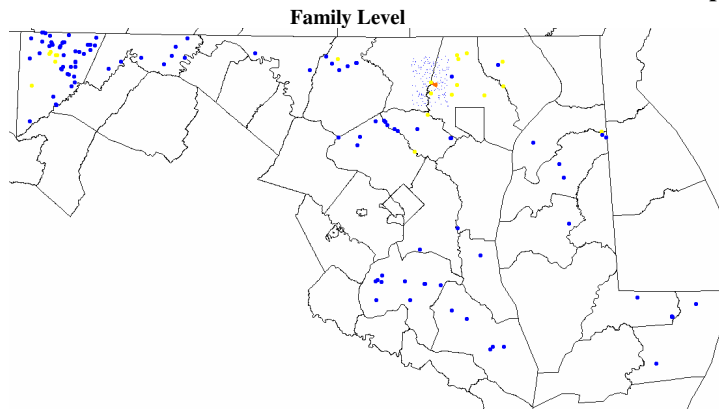
NROI scheme

Ba-p-077-315-96



NROI scheme

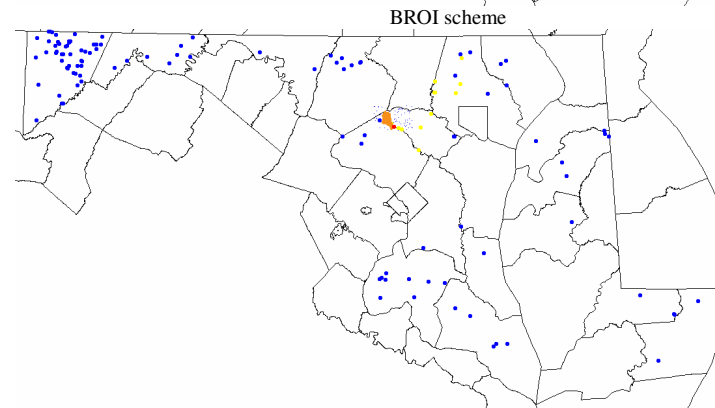
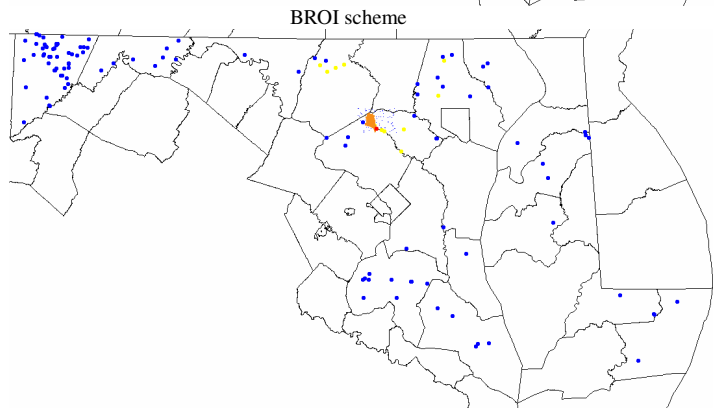
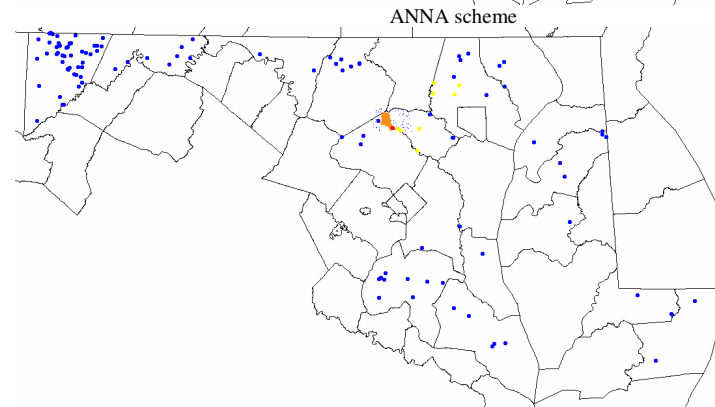
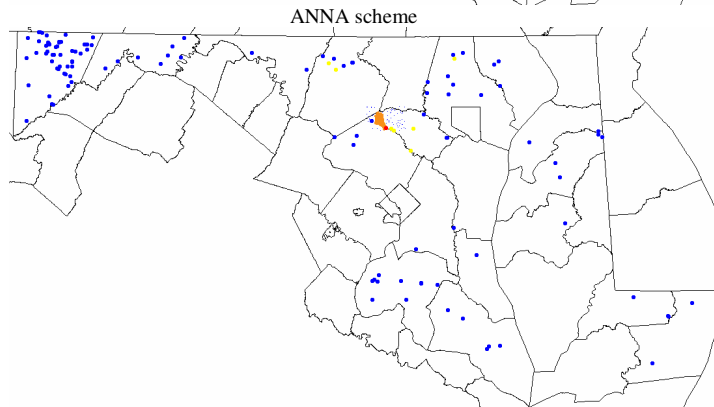
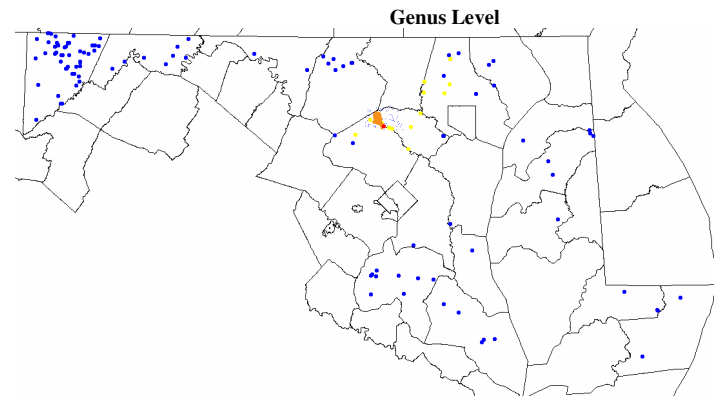
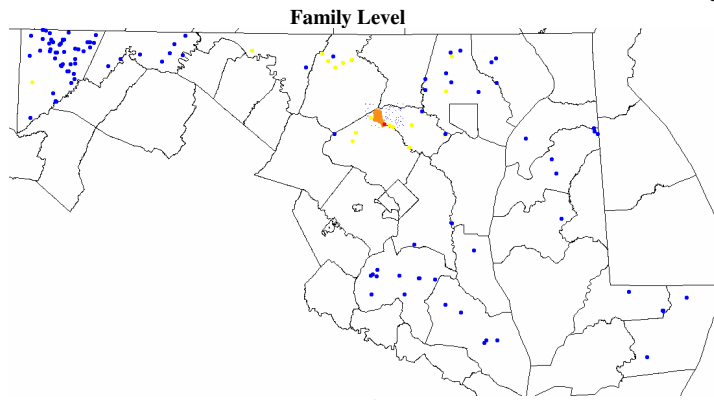
NROI scheme



NROI scheme

NROI scheme

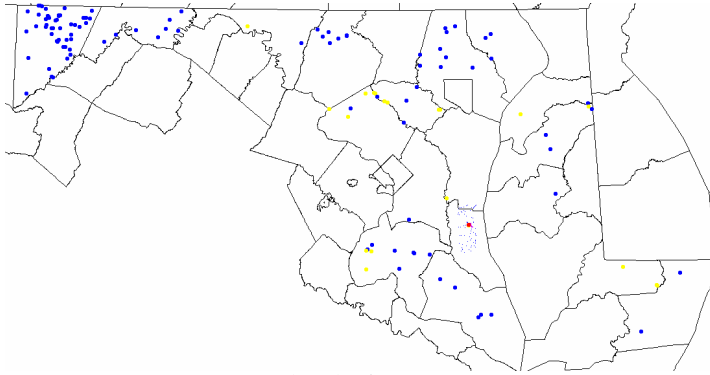
Brig-307-r-2000



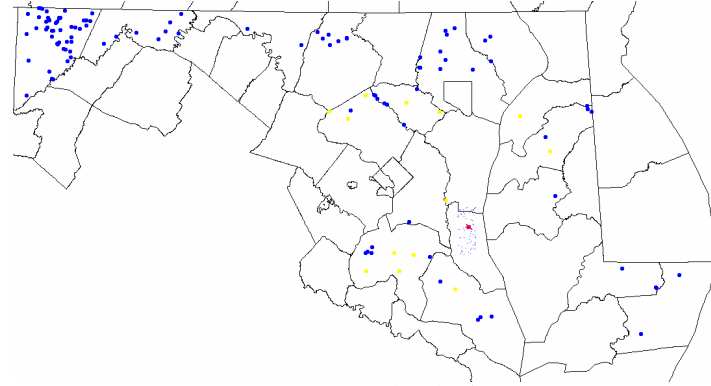
NROI scheme

NROI scheme

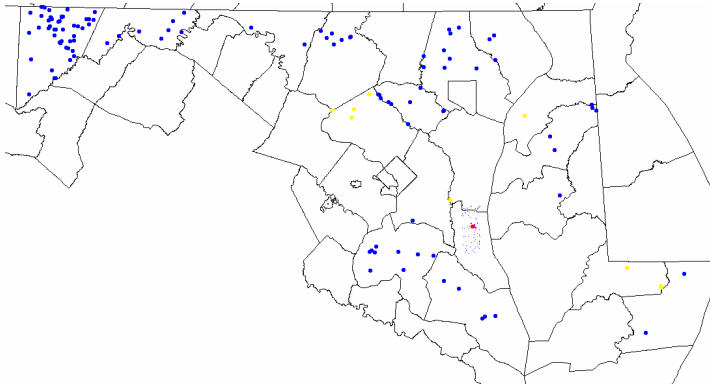
Family Level



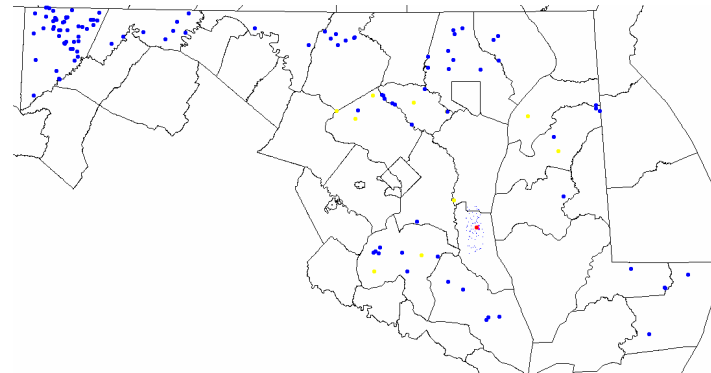
Genus Level



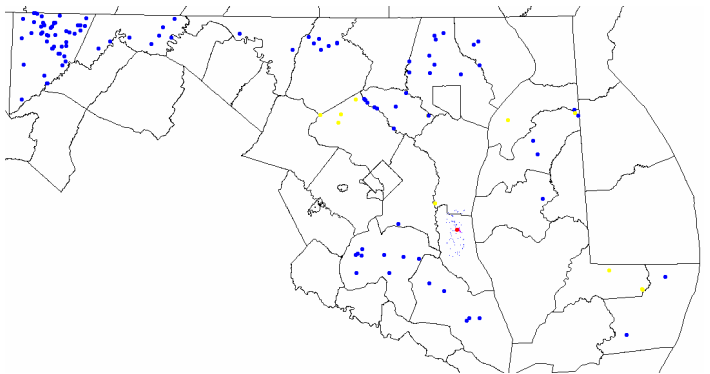
ANNA scheme



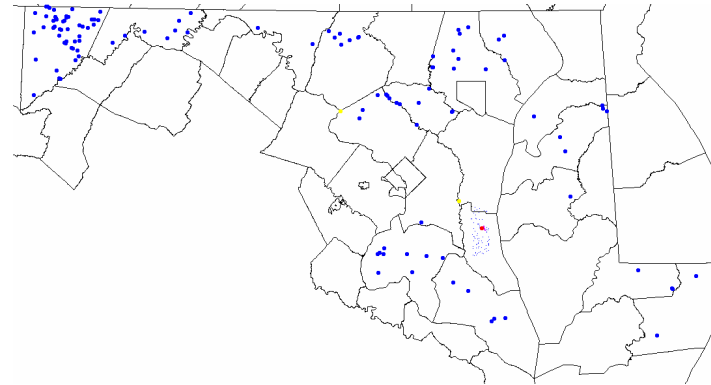
ANNA scheme



BROI scheme



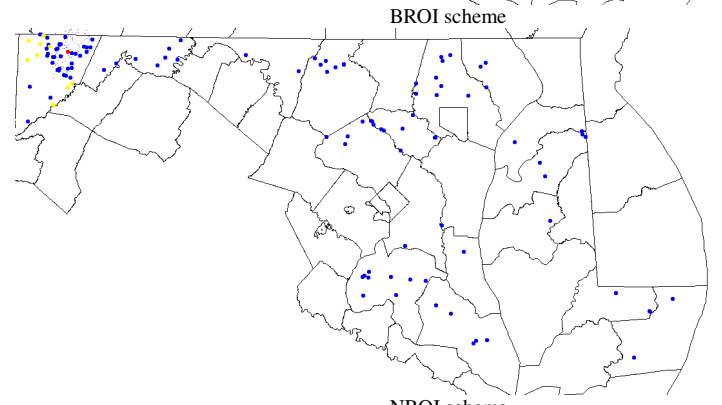
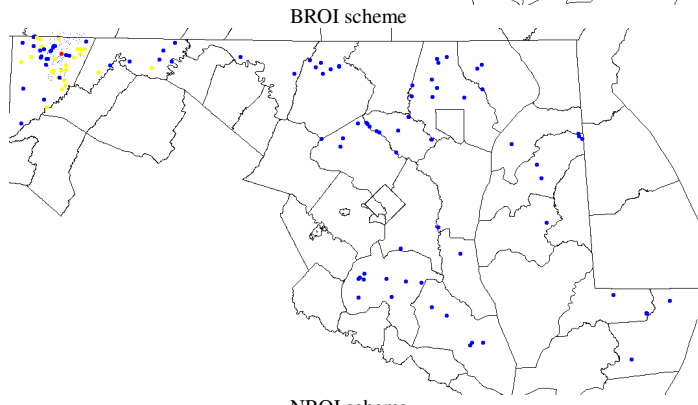
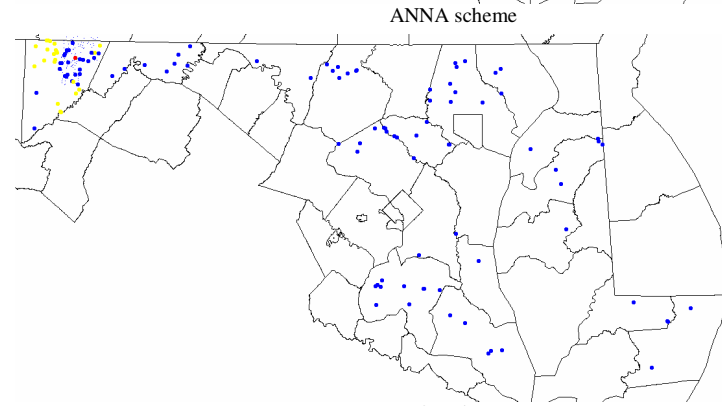
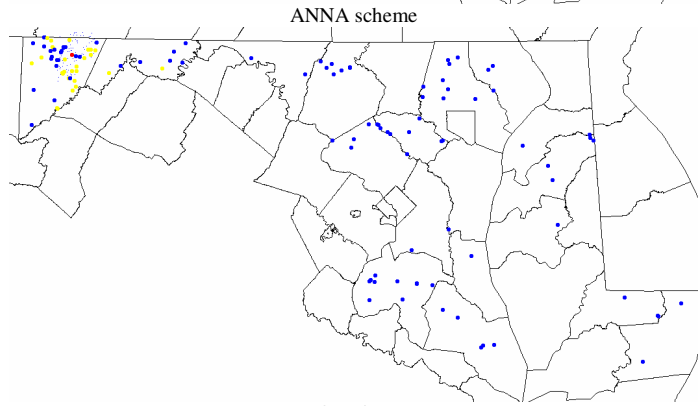
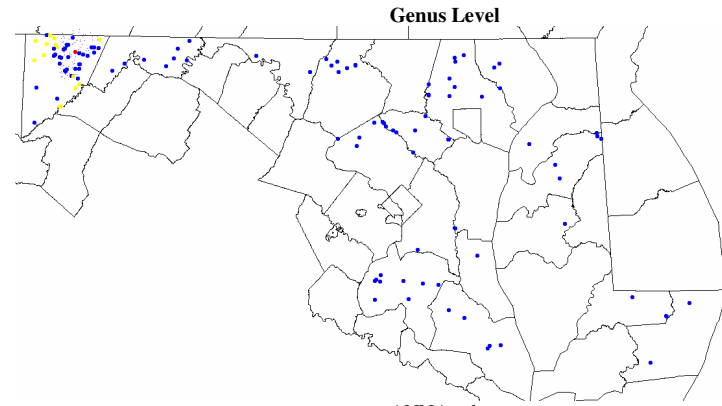
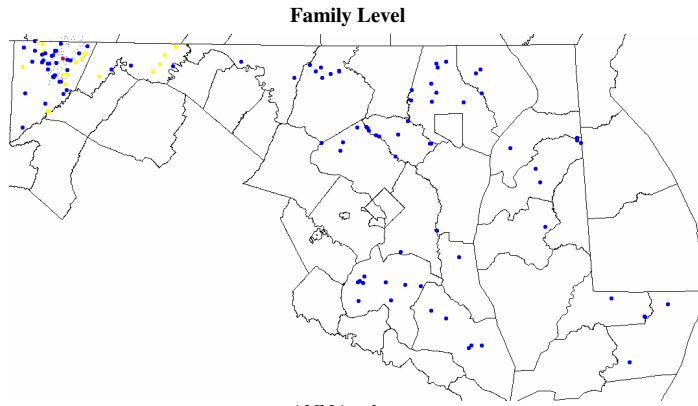
BROI scheme



NROI scheme

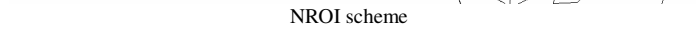
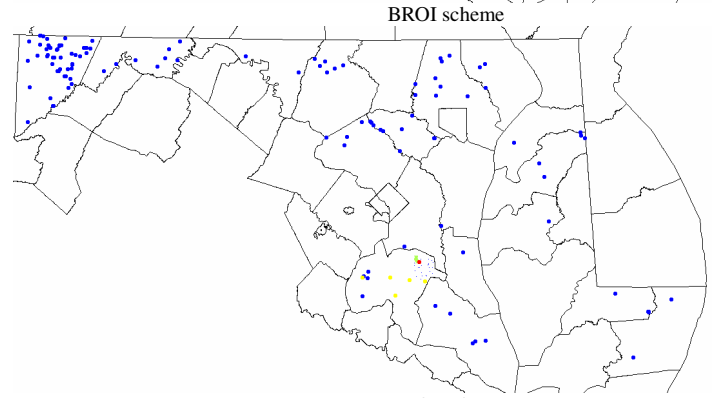
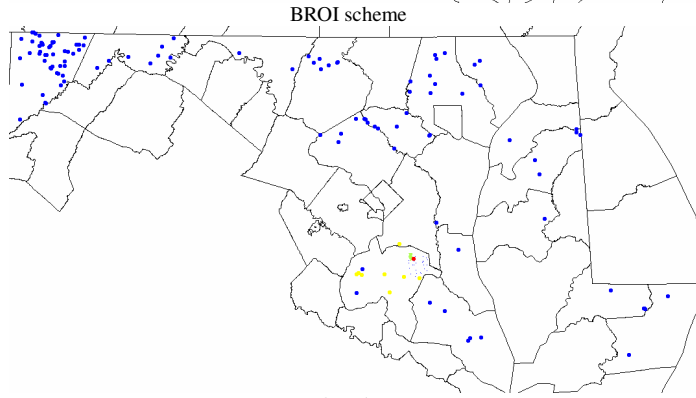
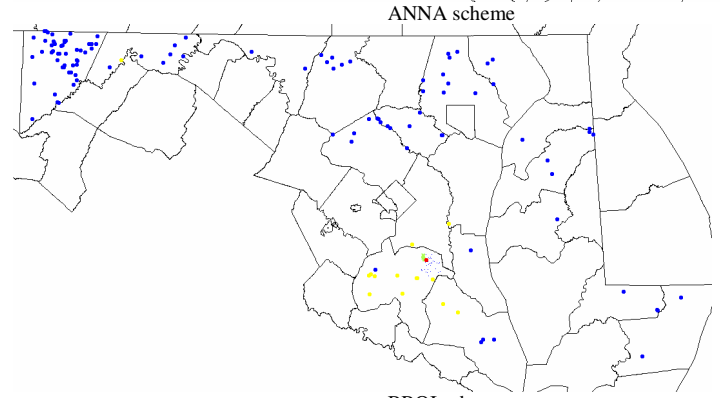
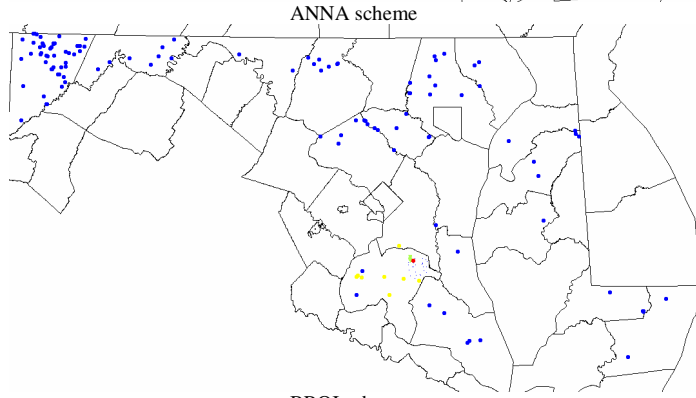
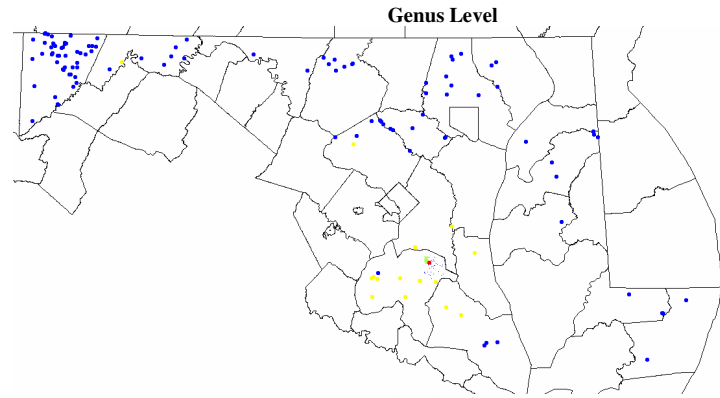
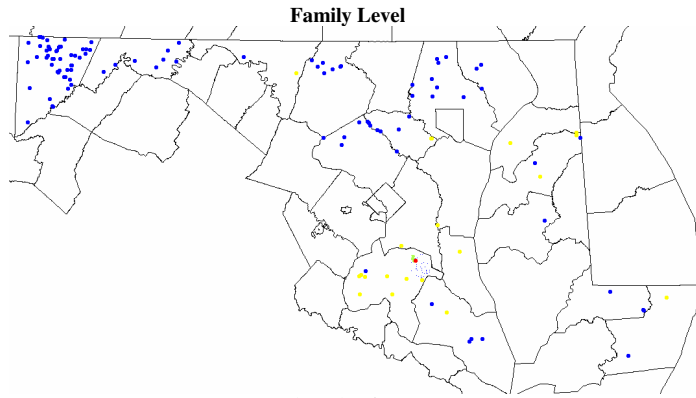
NROI scheme

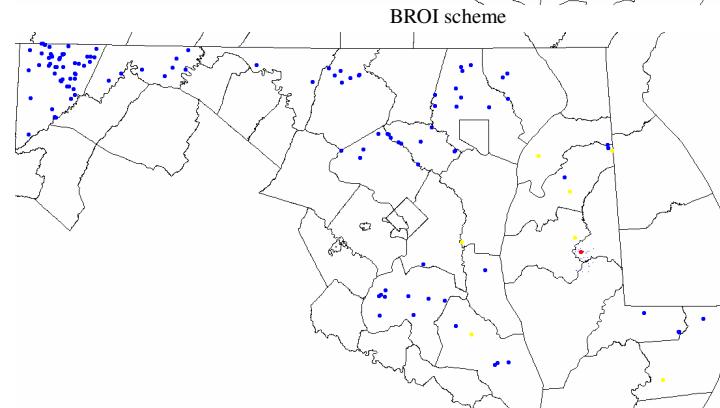
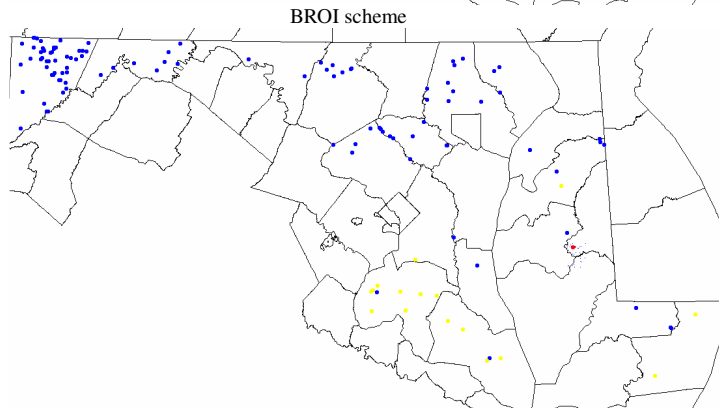
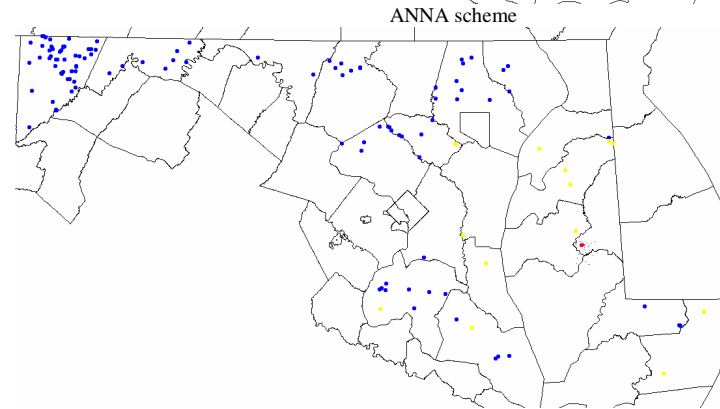
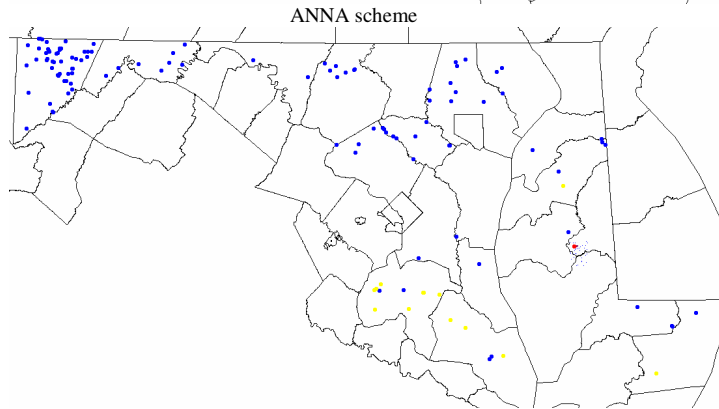
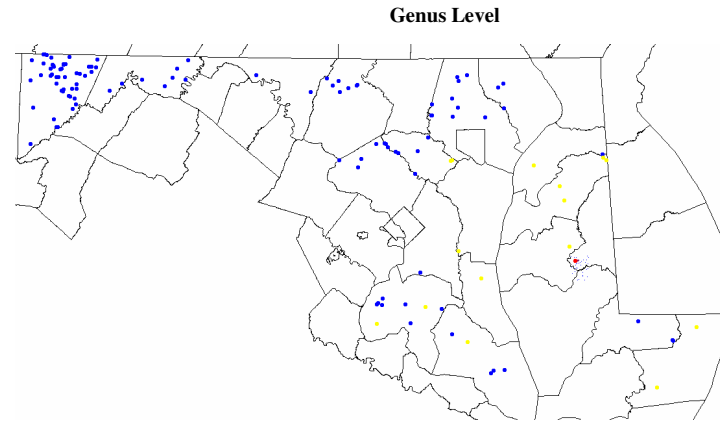
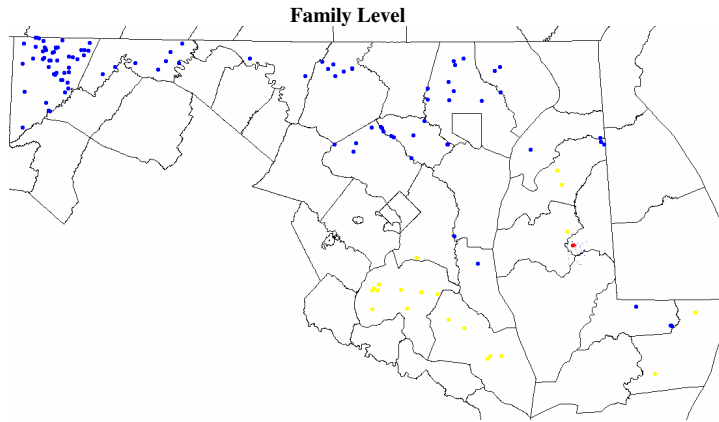
Cass-111-r-2000



NROI scheme

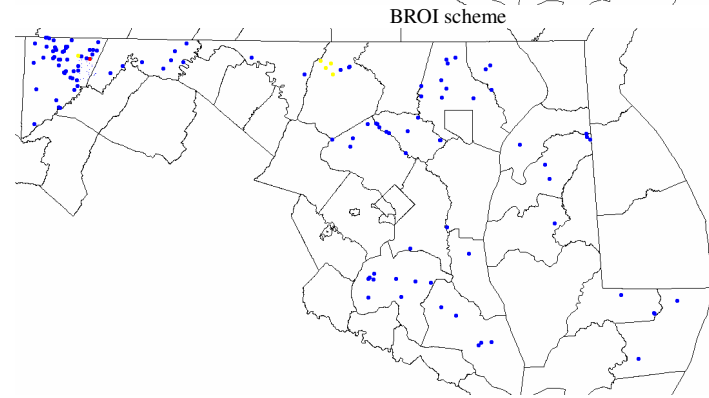
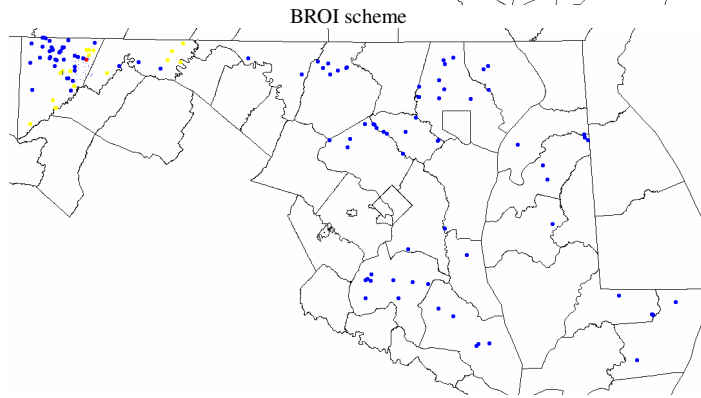
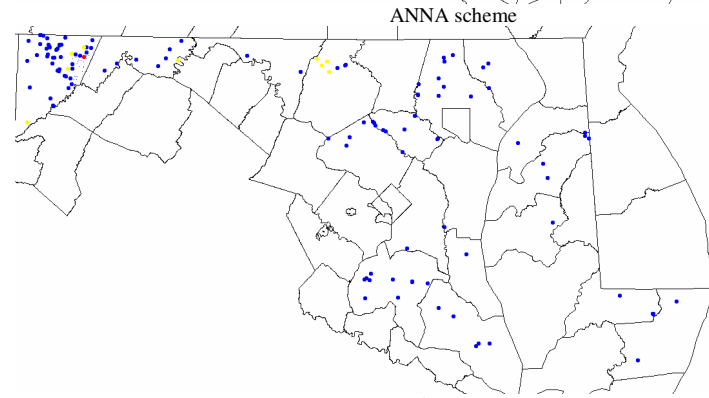
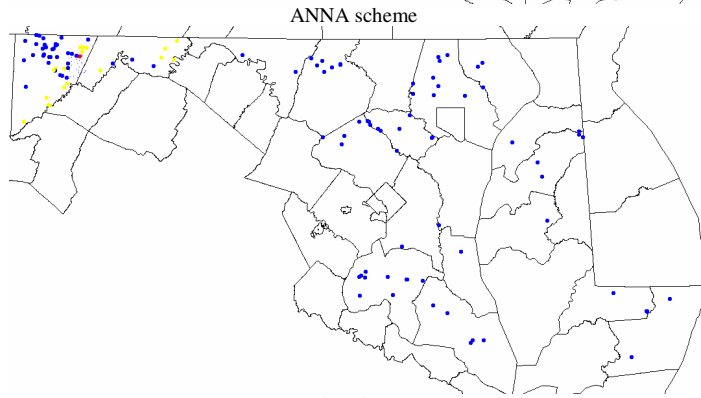
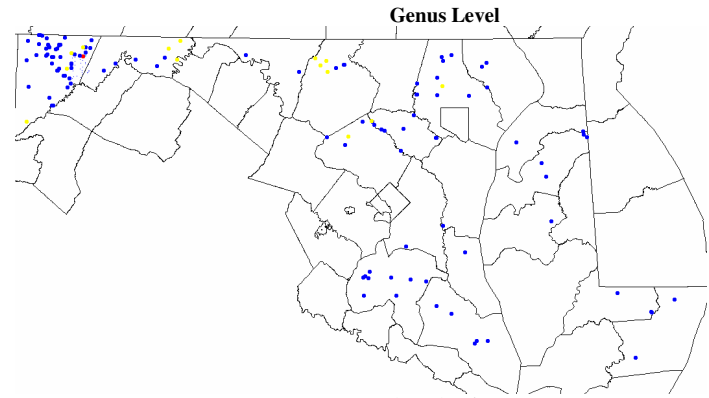
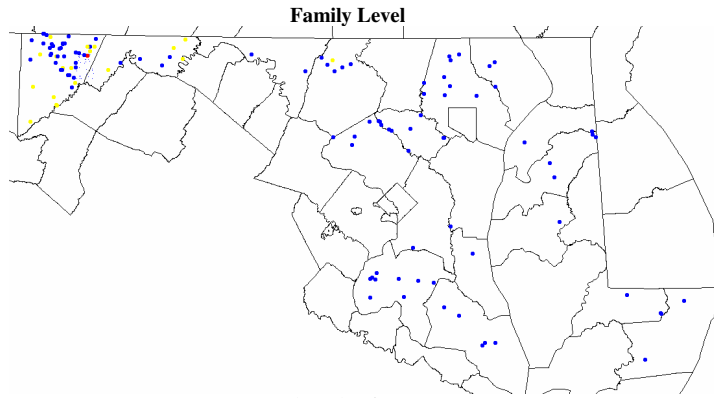
NROI scheme





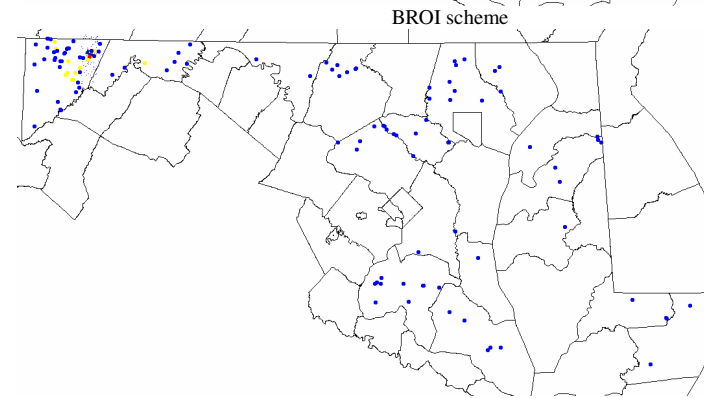
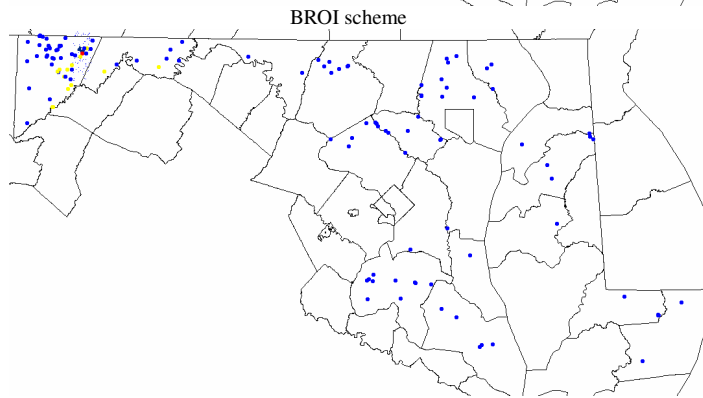
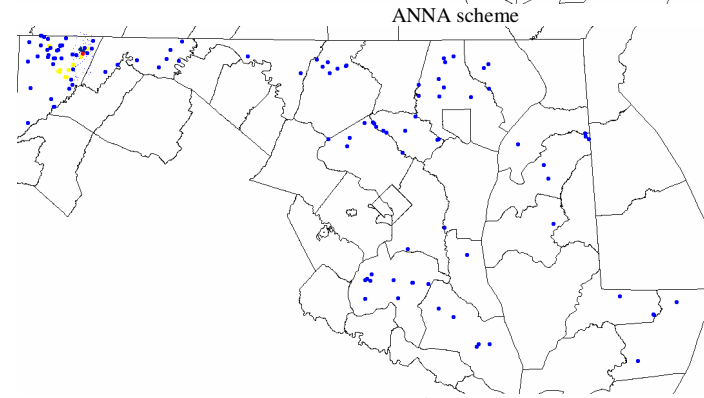
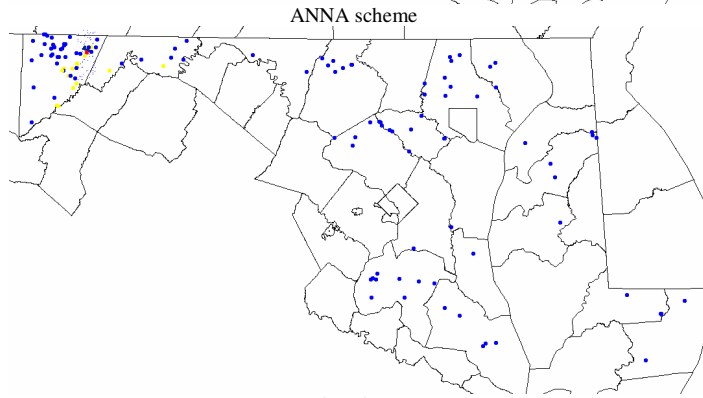
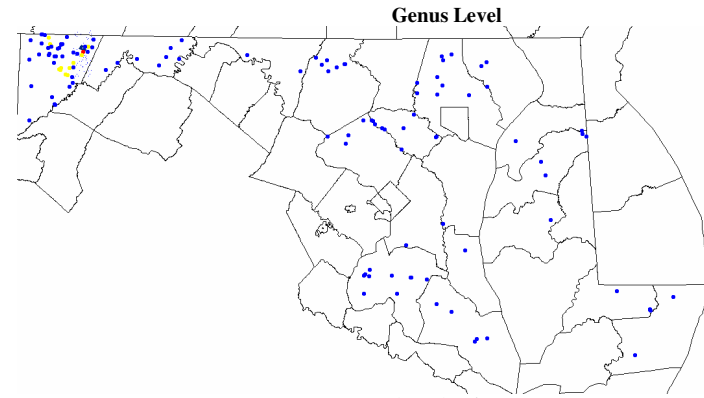
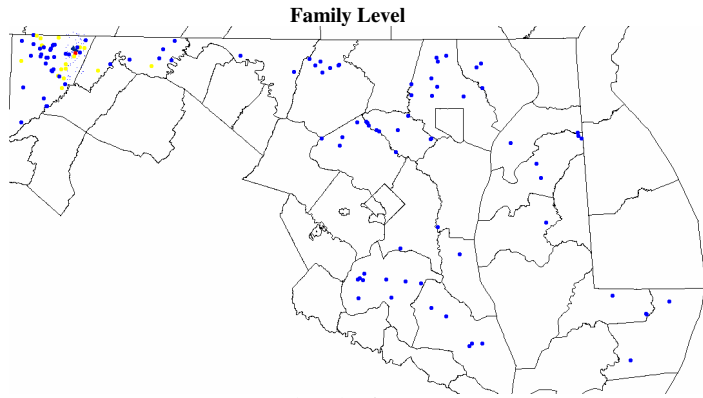
NROI scheme

NROI scheme



NROI scheme

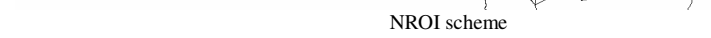
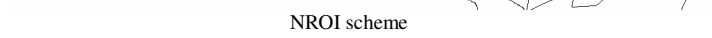
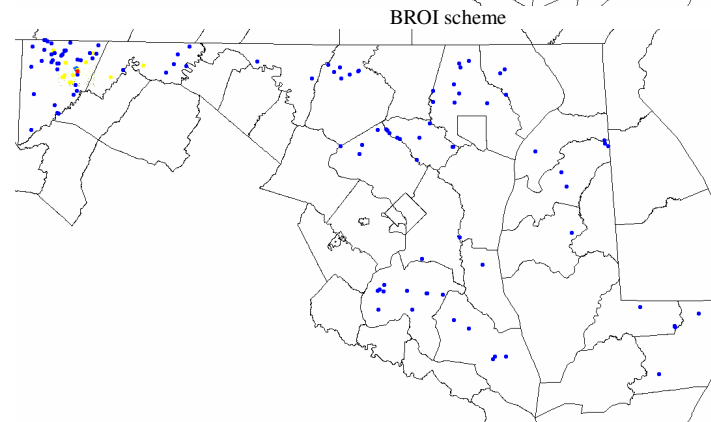
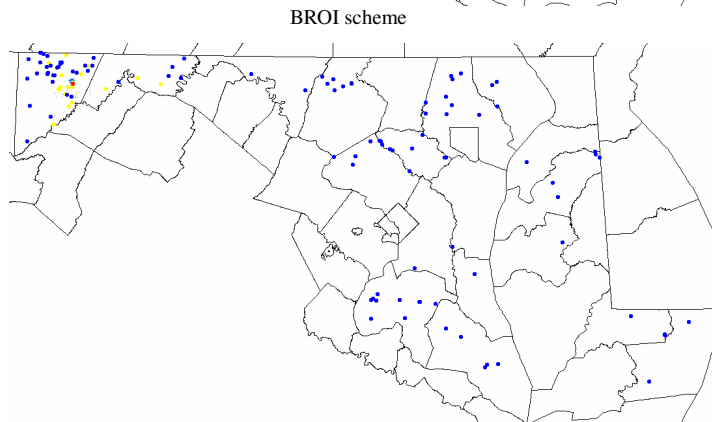
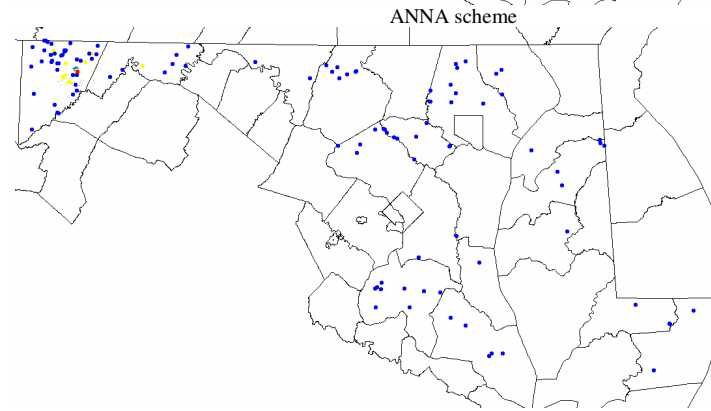
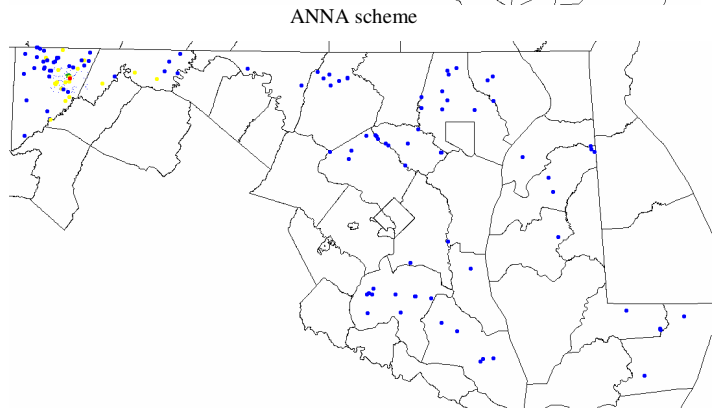
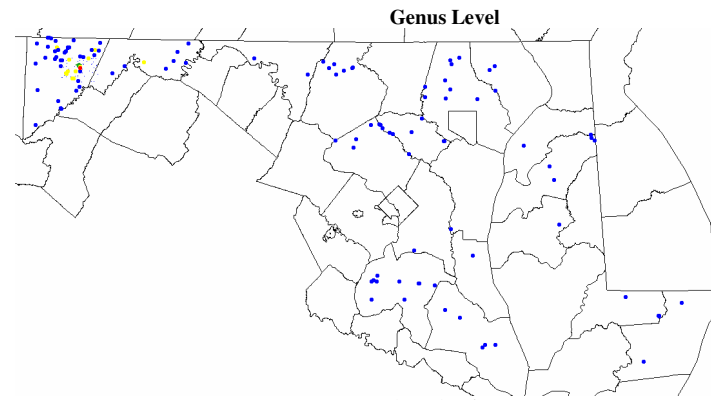
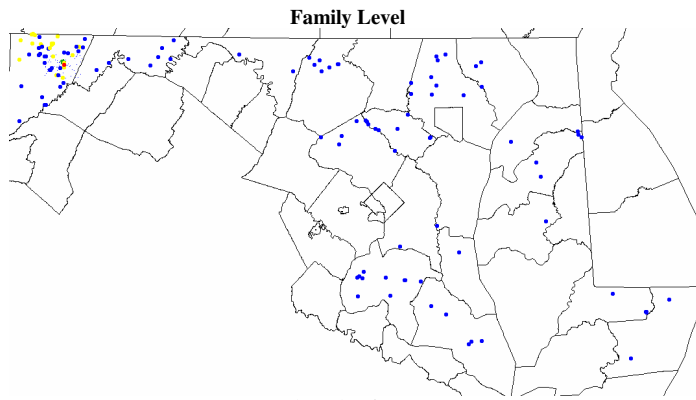
NROI scheme

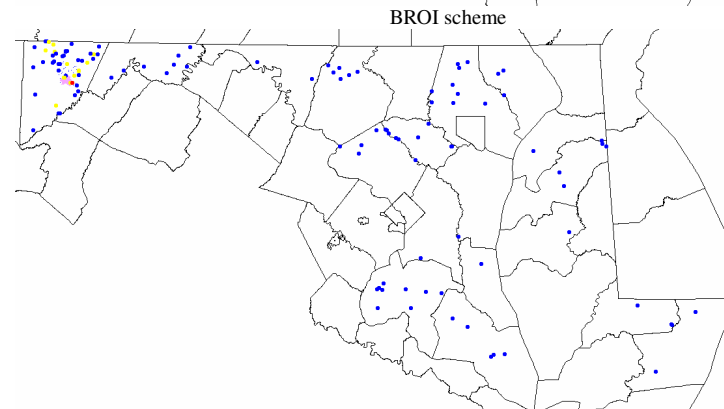
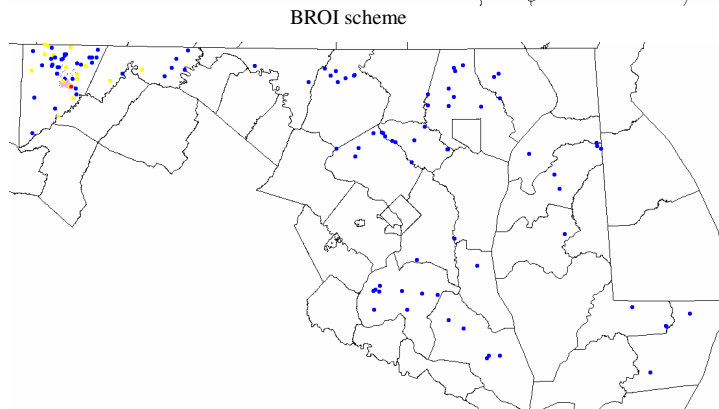
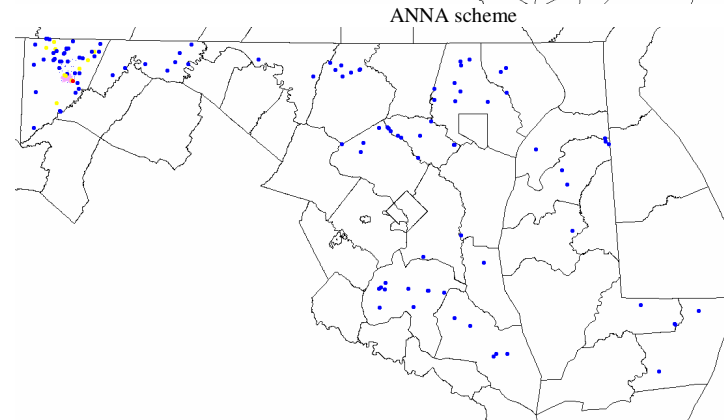
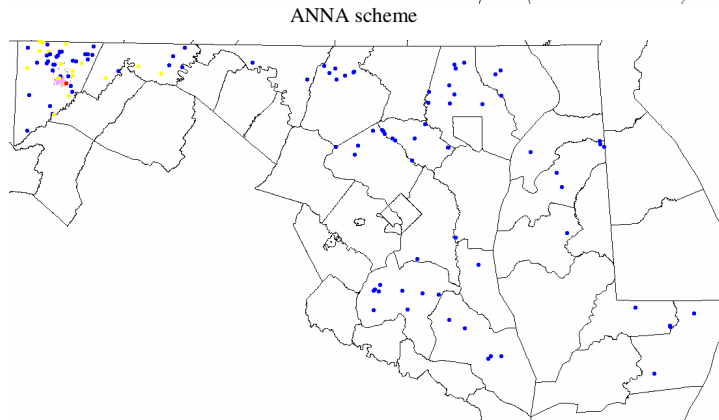
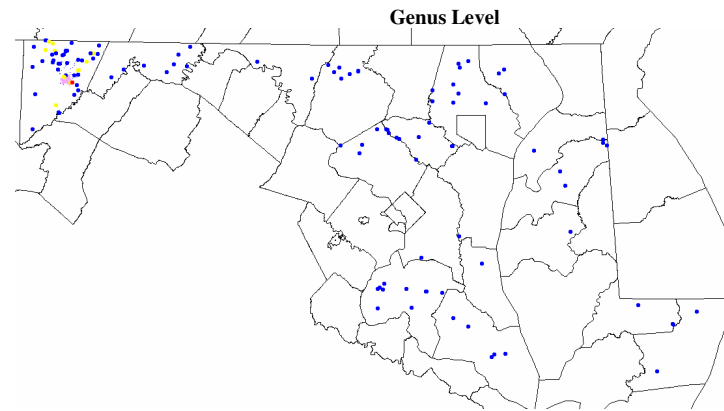
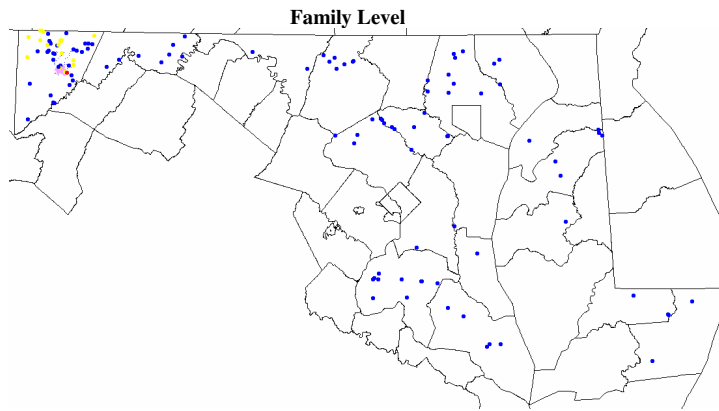


NROI scheme

NROI scheme

Ga-a-121-210-96



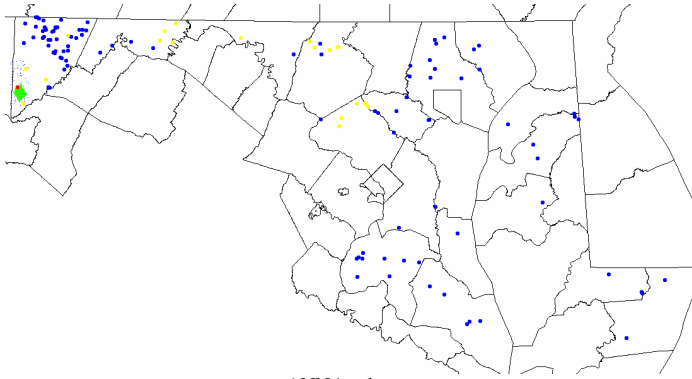


NROI scheme

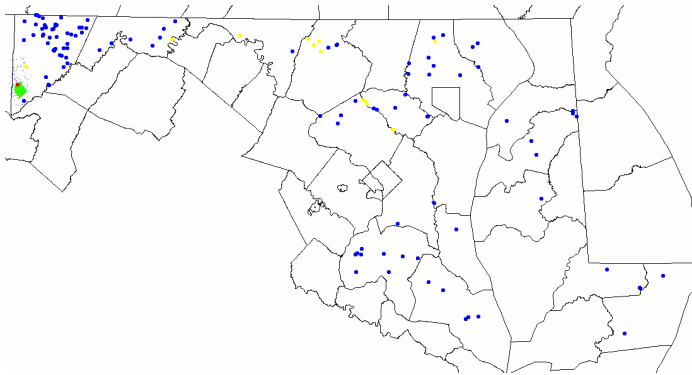
NROI scheme

Ga-a-185-309-95

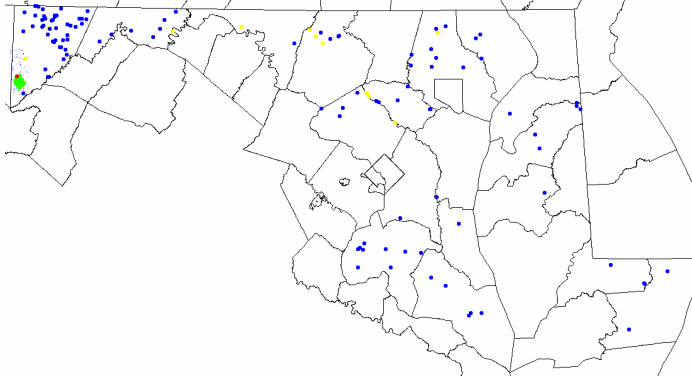
Family Level



ANNA scheme

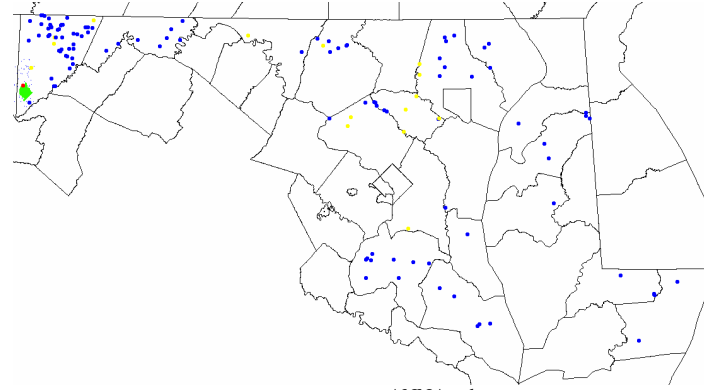


BROI scheme

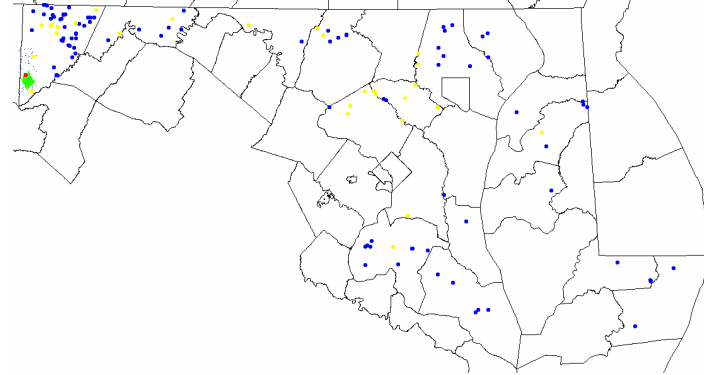


NROI scheme

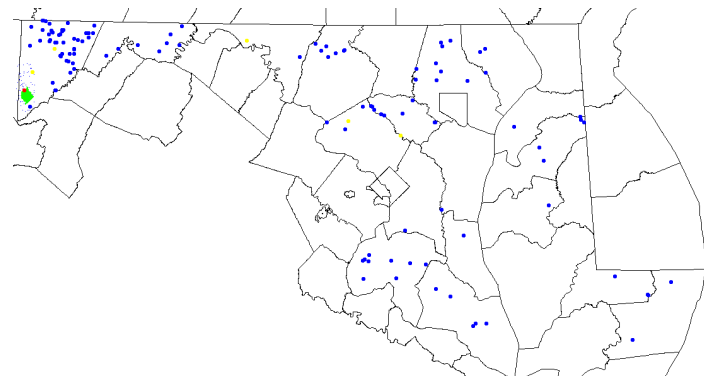
Genus Level



ANNA scheme

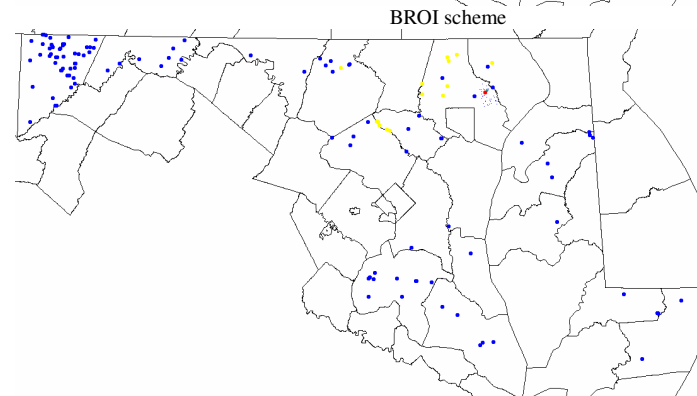
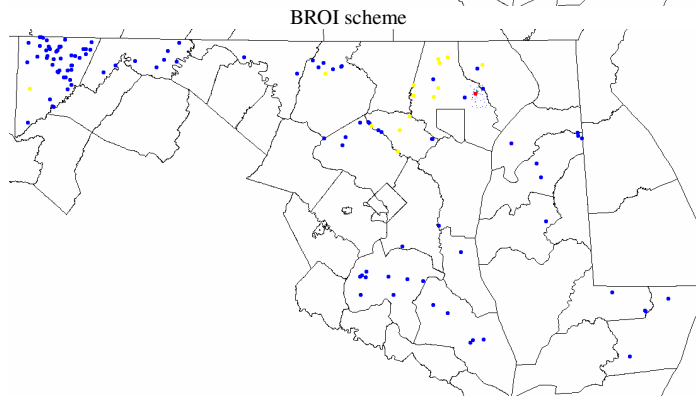
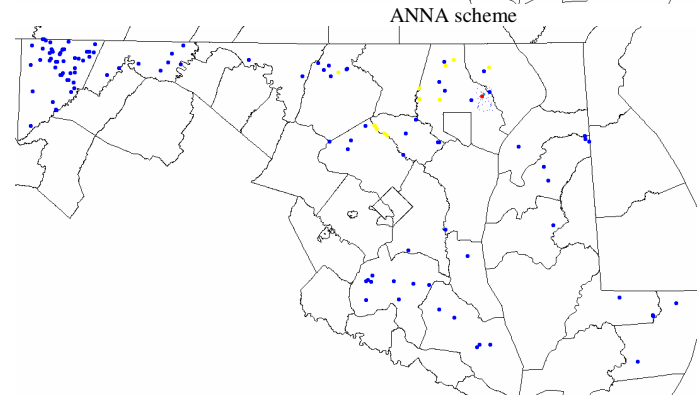
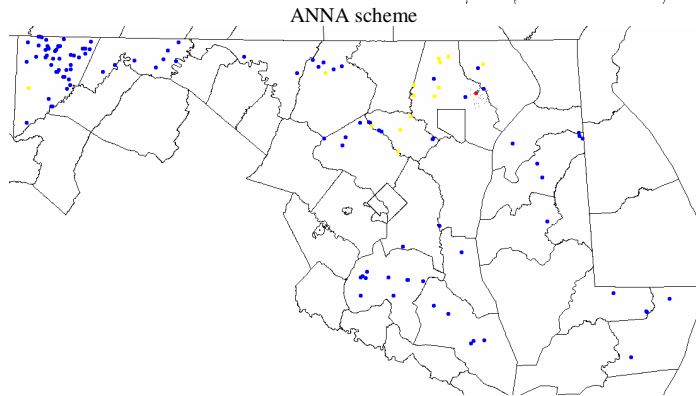
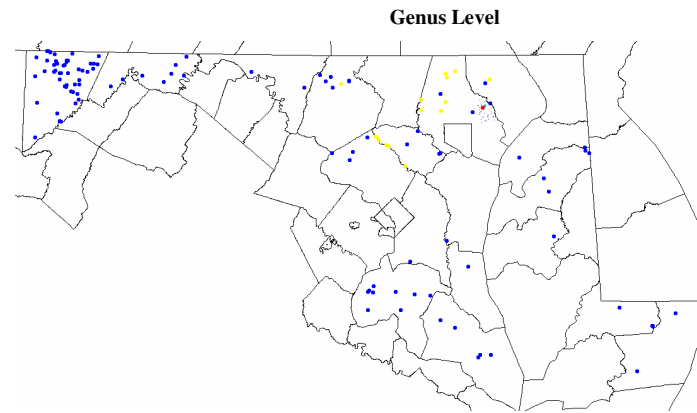
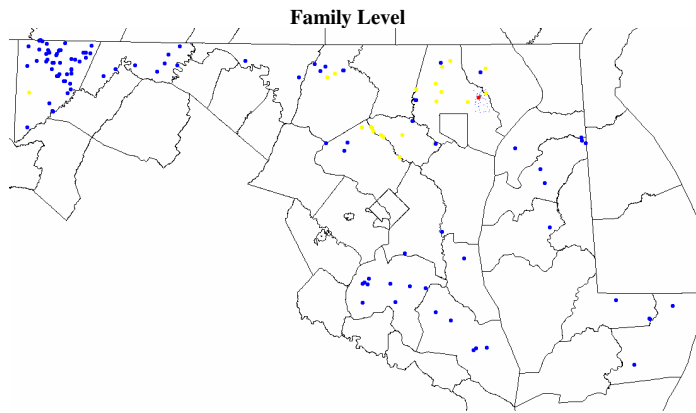


BROI scheme



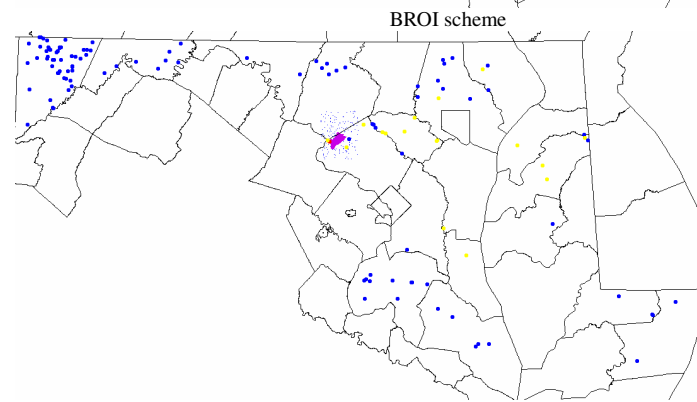
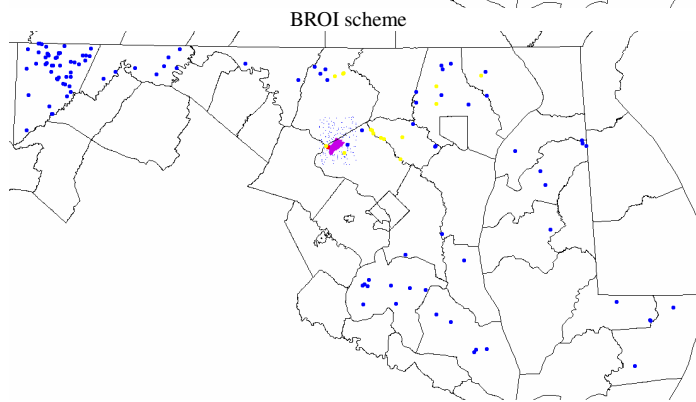
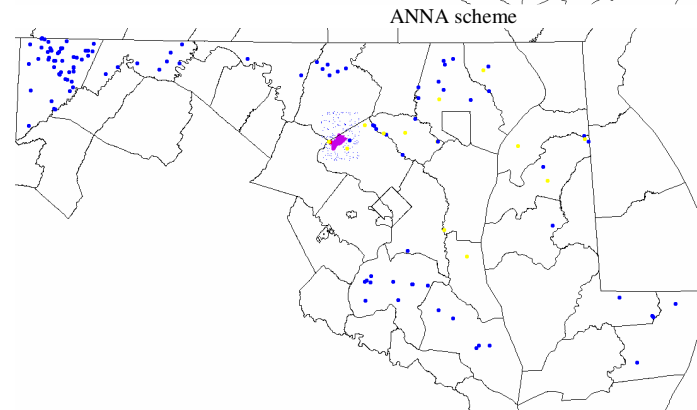
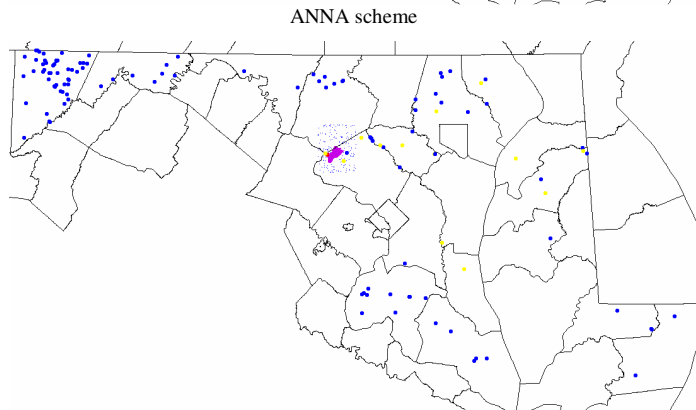
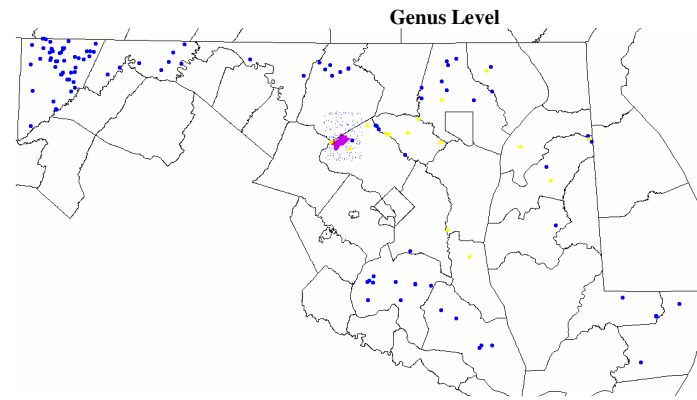
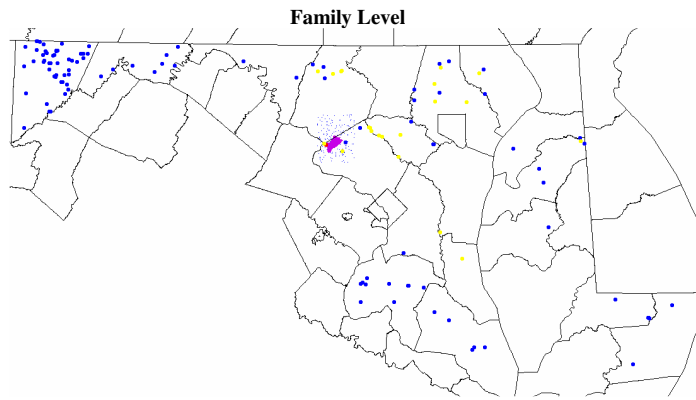
NROI scheme

Logu-106-r-2002



NROI scheme

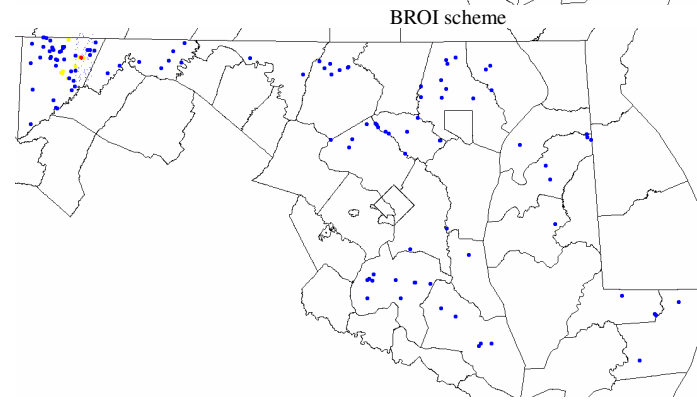
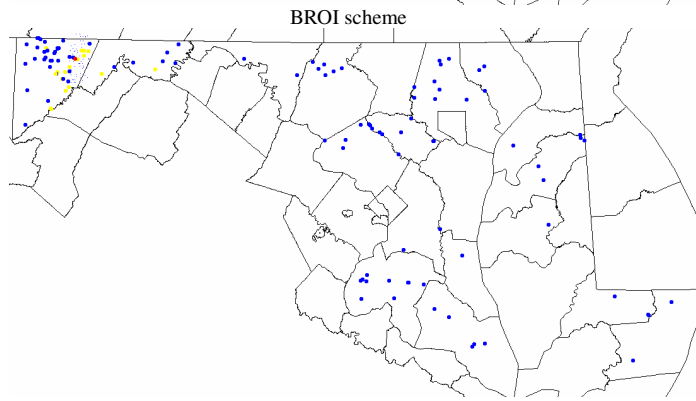
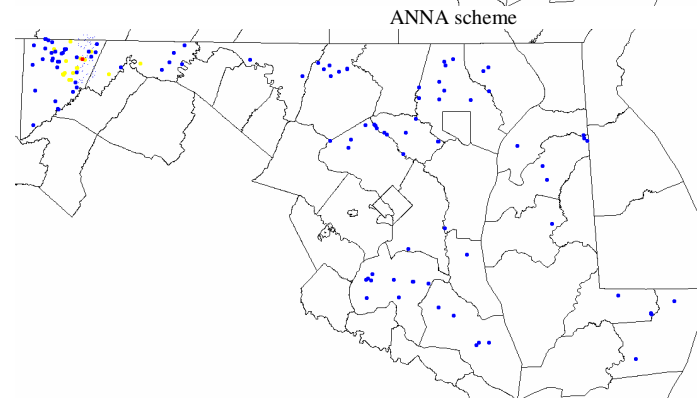
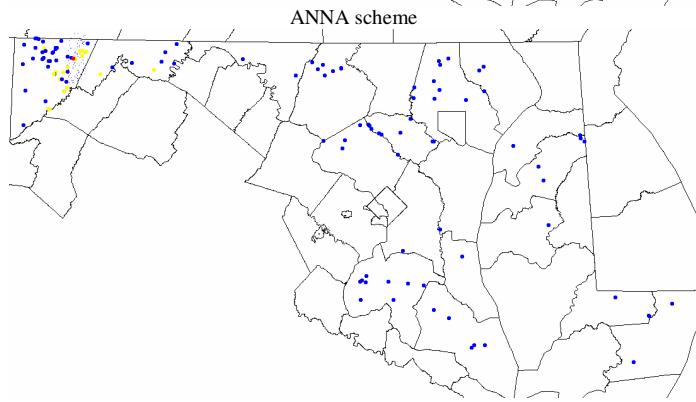
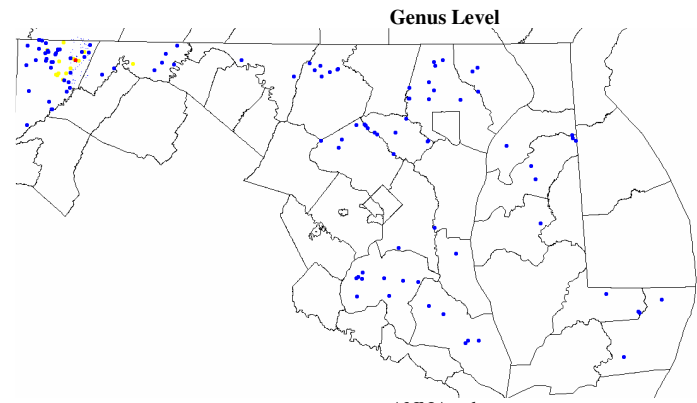
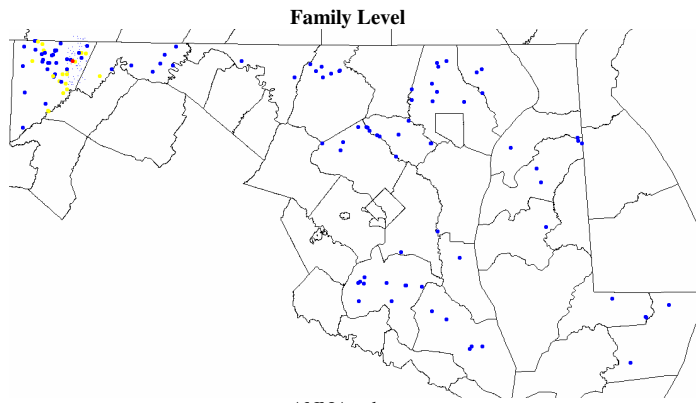
NROI scheme



NROI scheme

NROI scheme

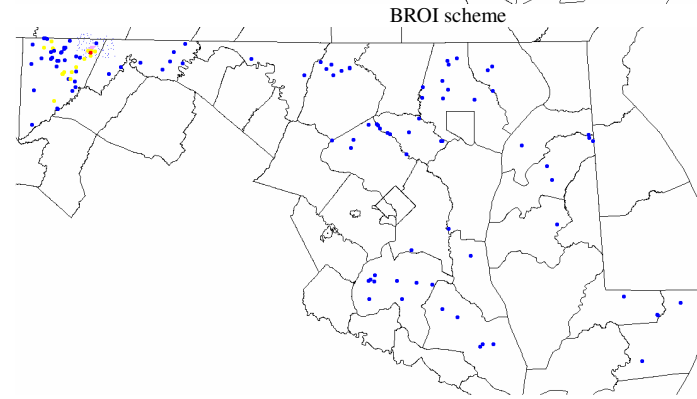
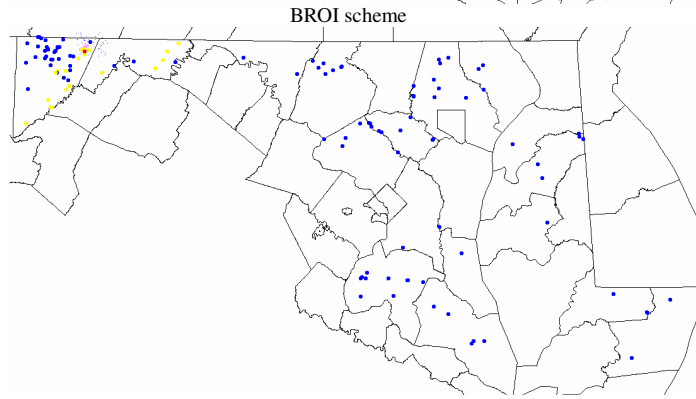
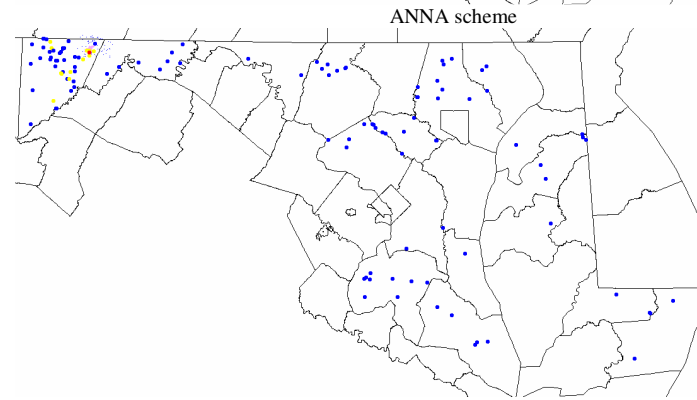
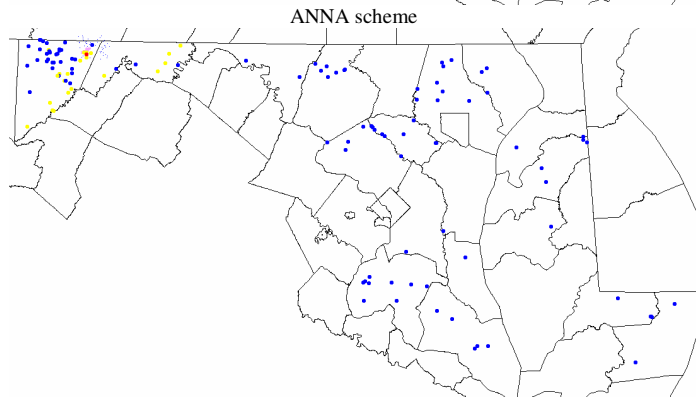
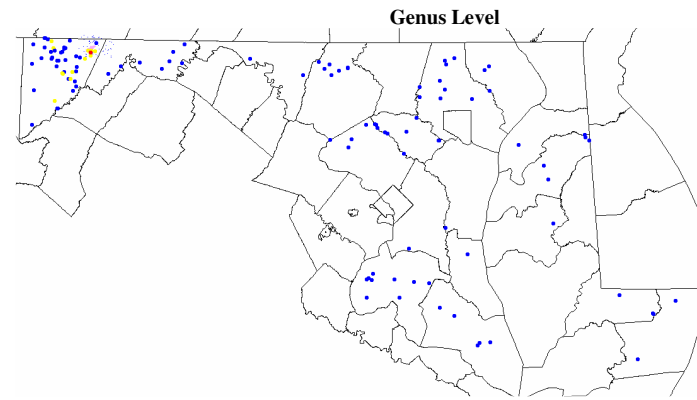
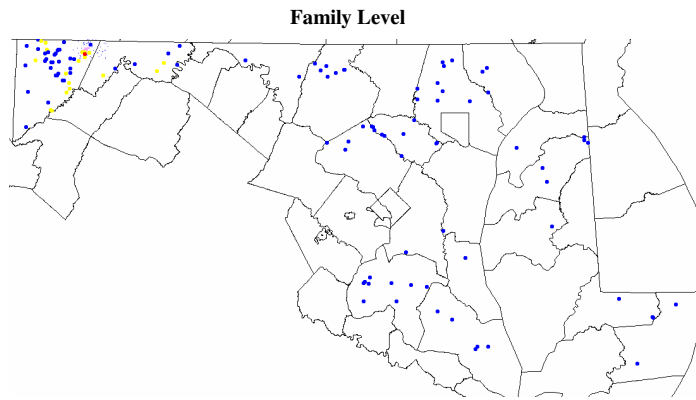
Sava-116-r-2002



NROI scheme

NROI scheme

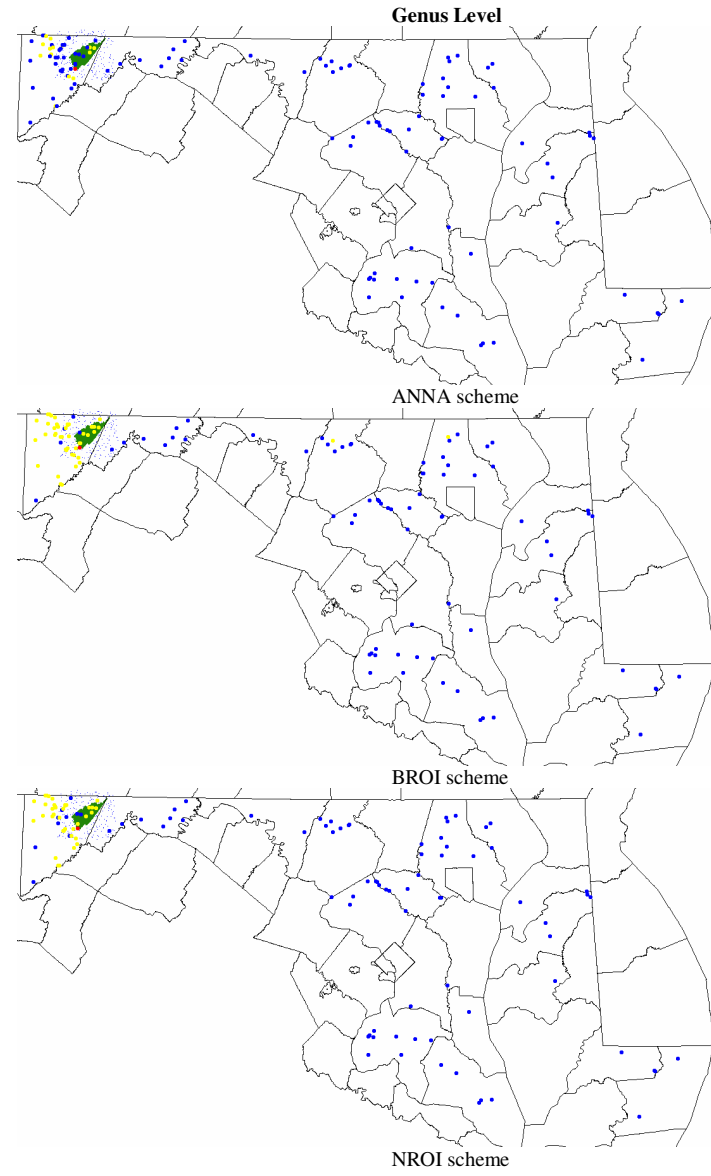
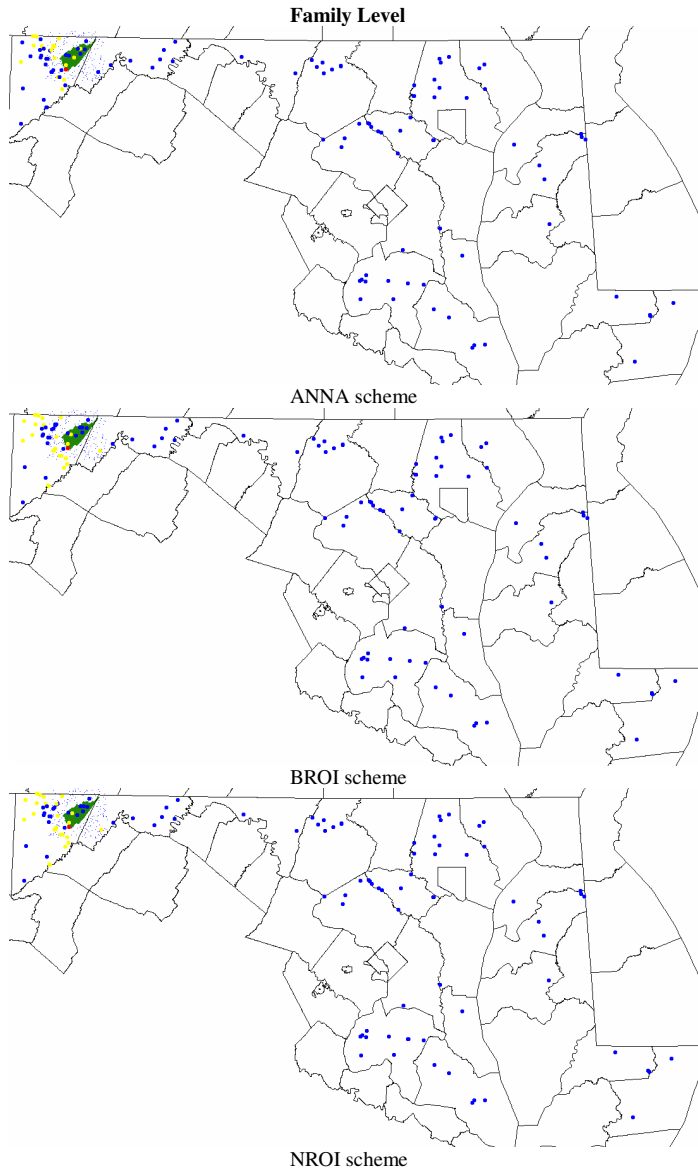
Sava-206-r-2002



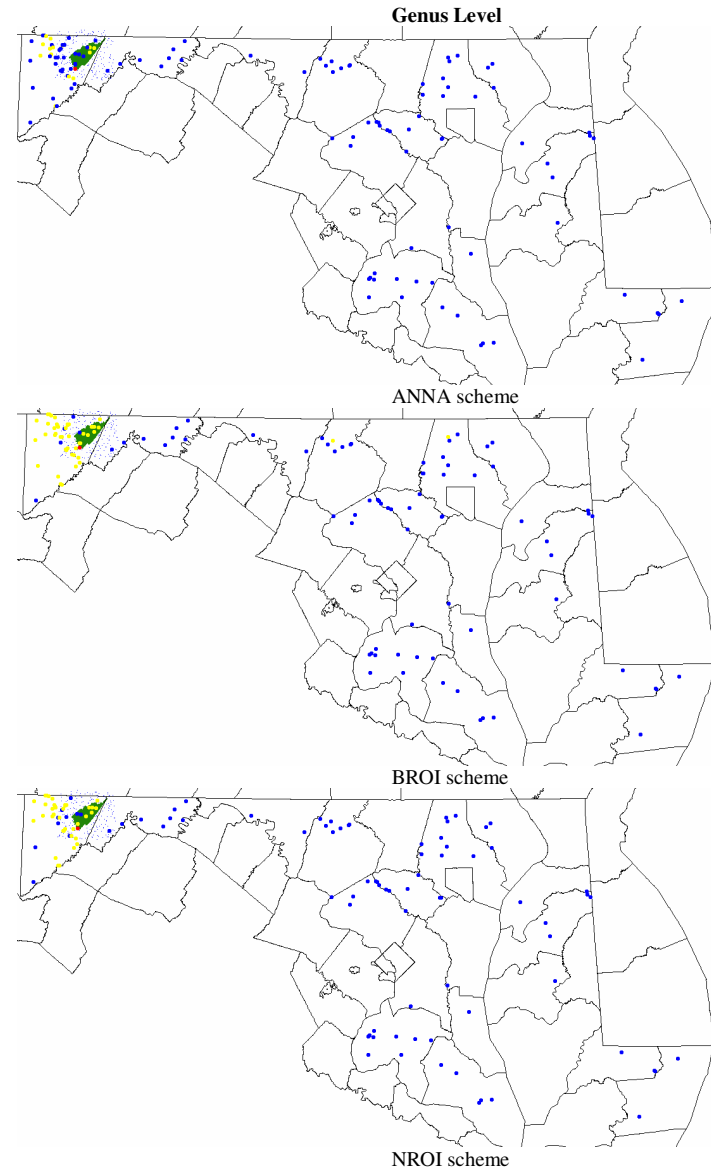
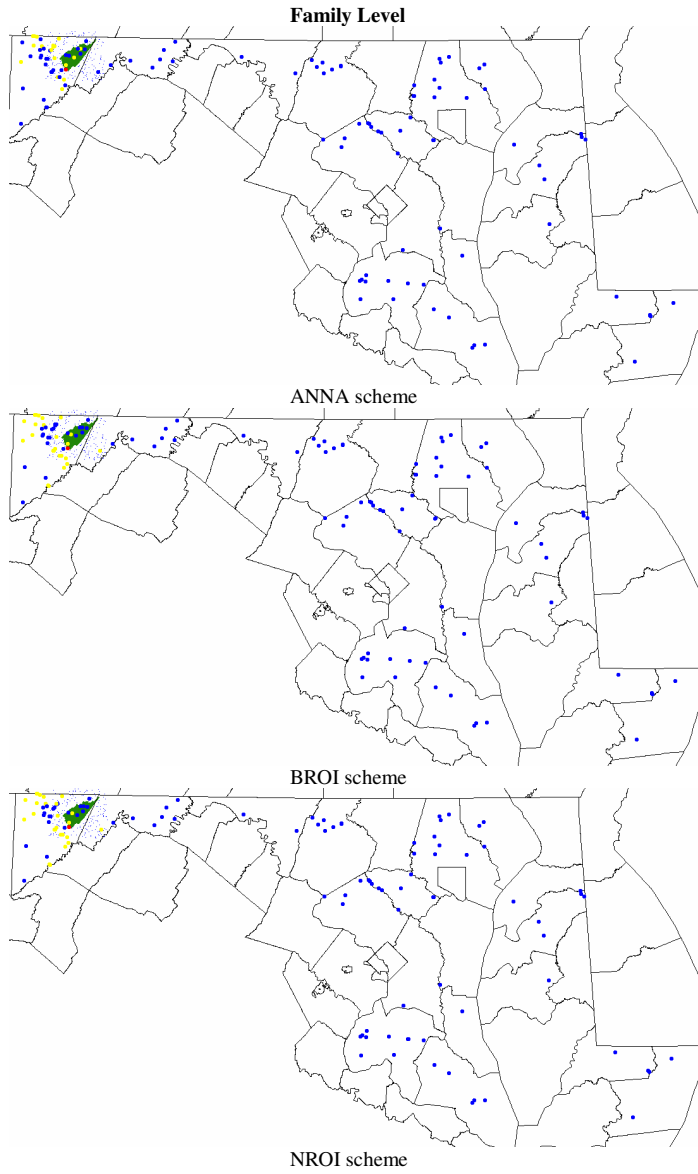
NROI scheme

NROI scheme

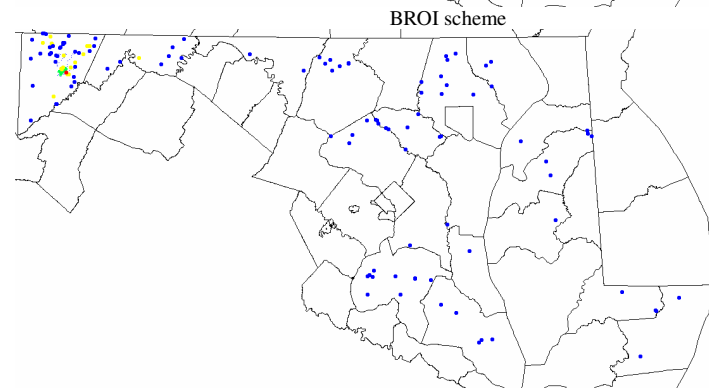
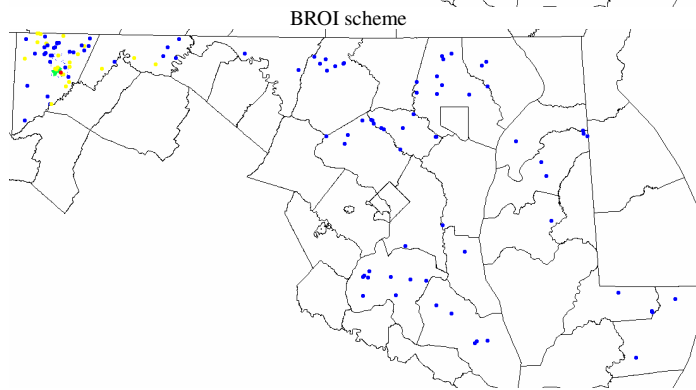
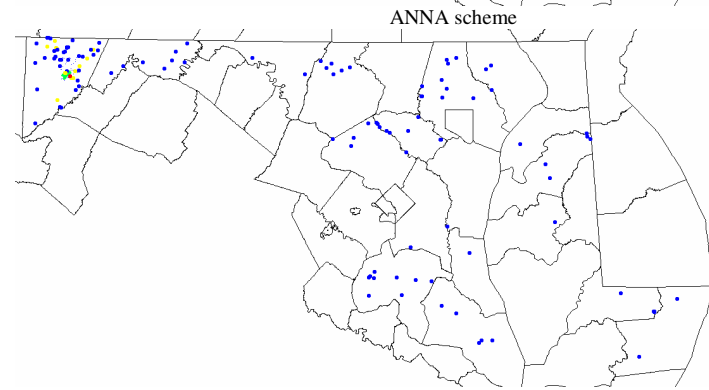
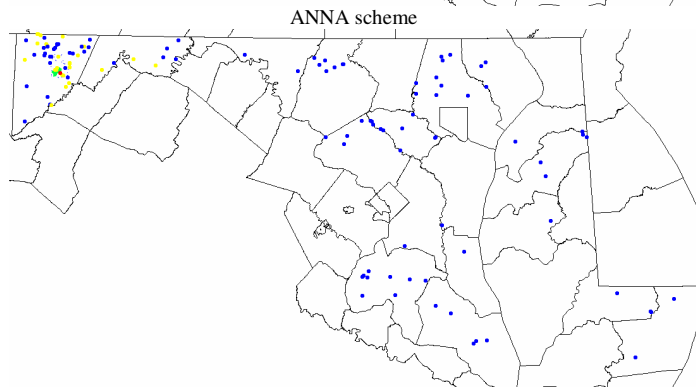
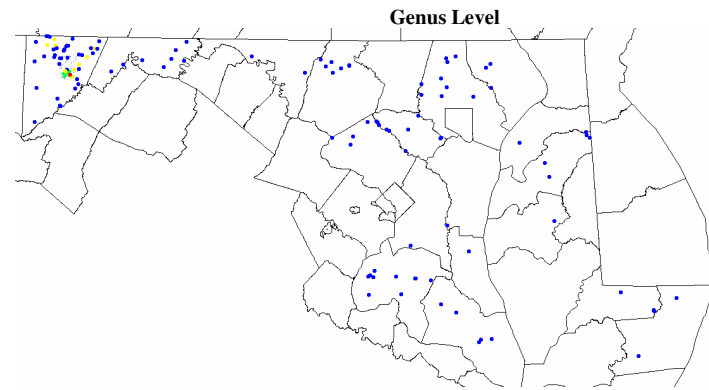
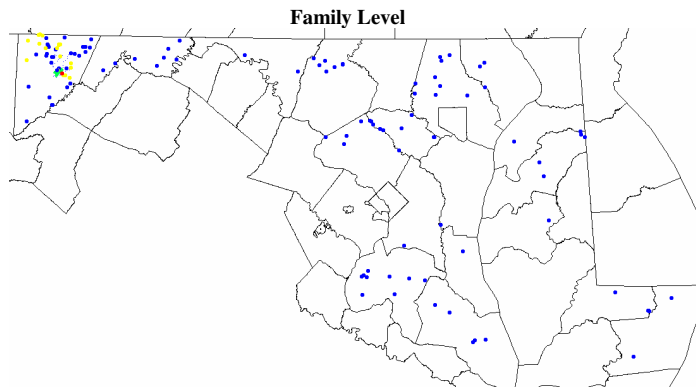
Sava-225-s-2001



Sava-225-s-2002



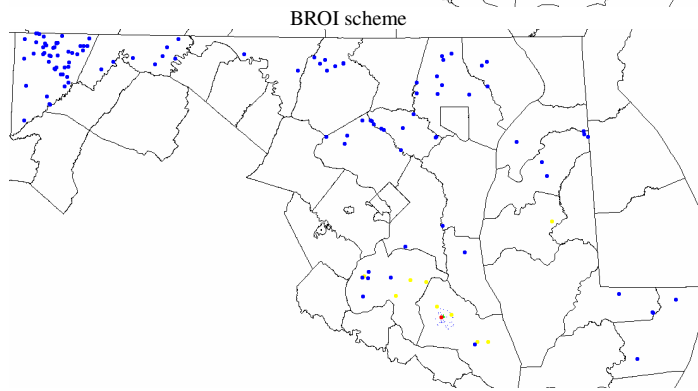
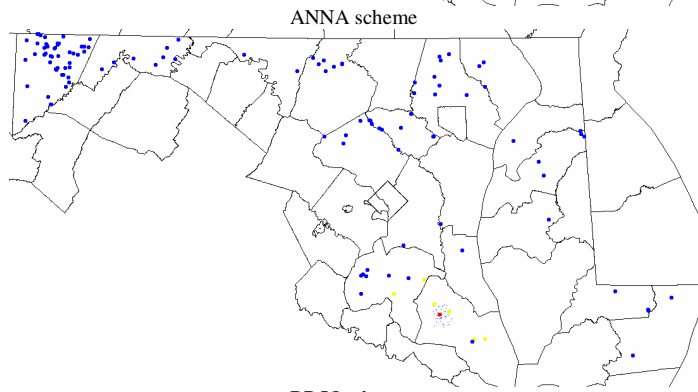
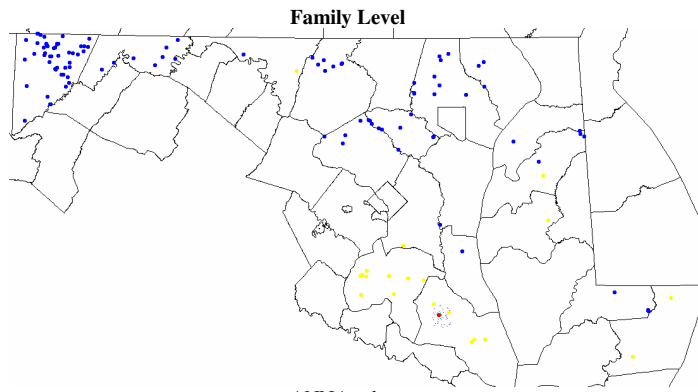
Sava-312-r-2002



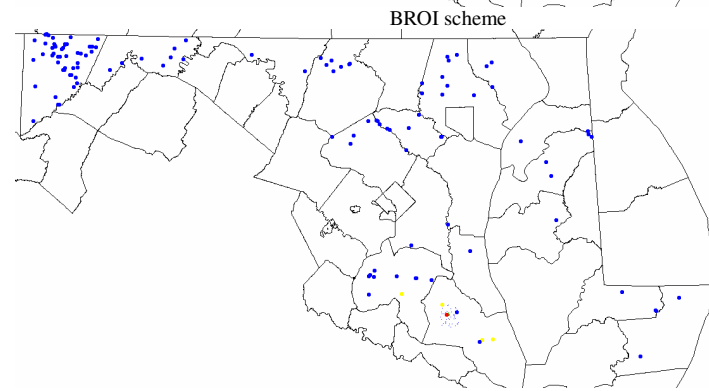
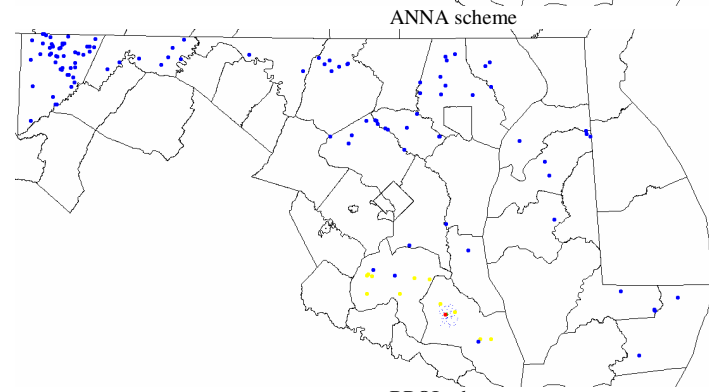
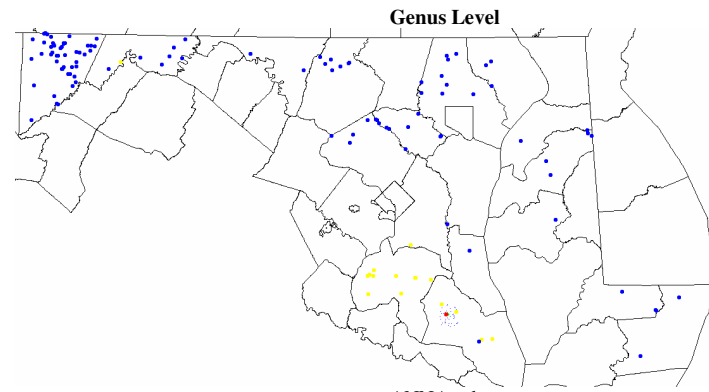
NROI scheme

NROI scheme

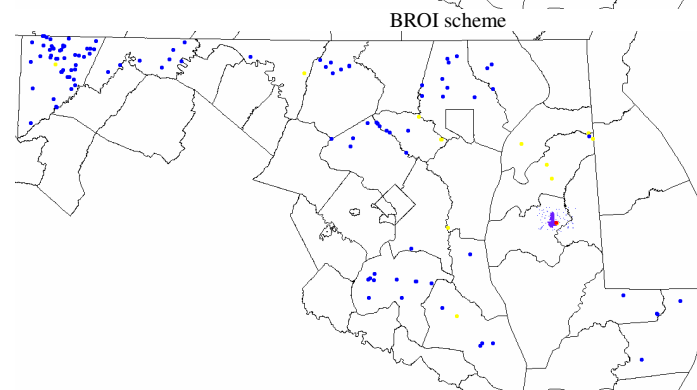
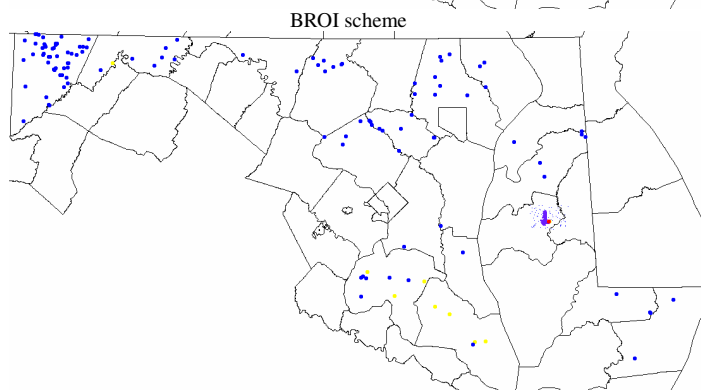
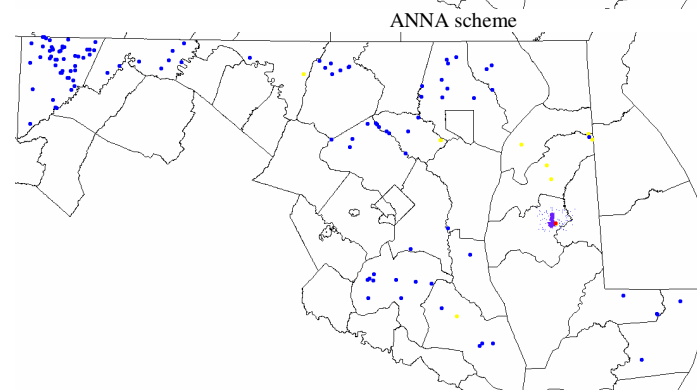
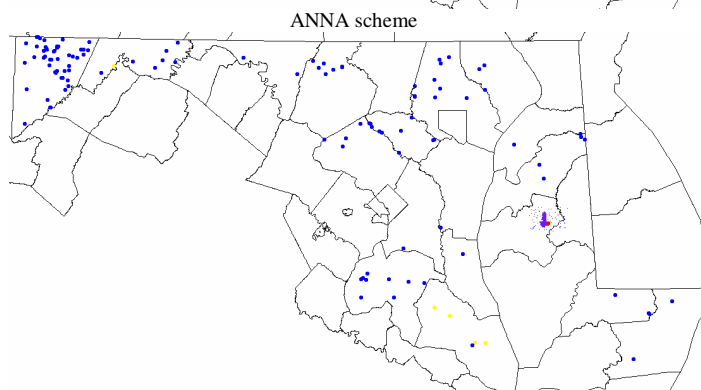
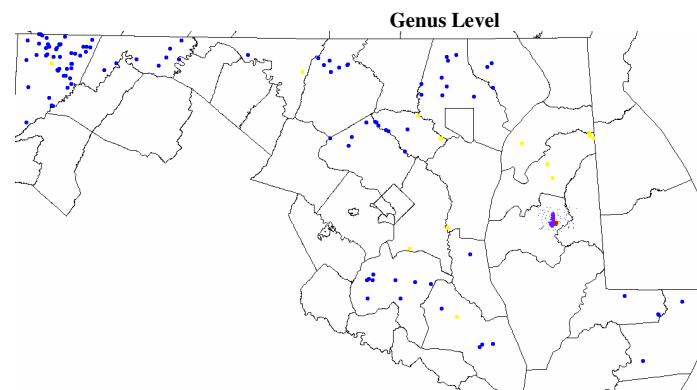
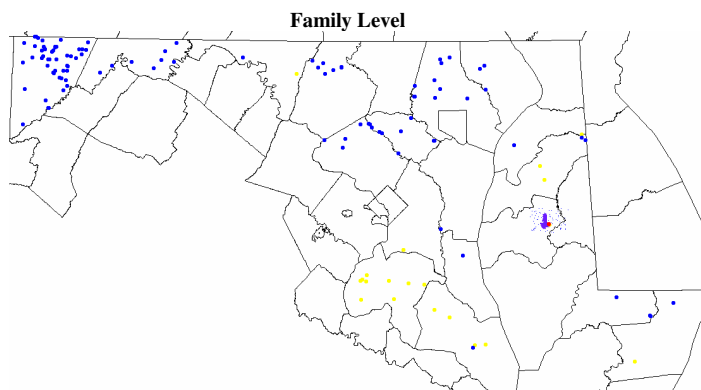
Sm-s-039-127-95



NROI scheme



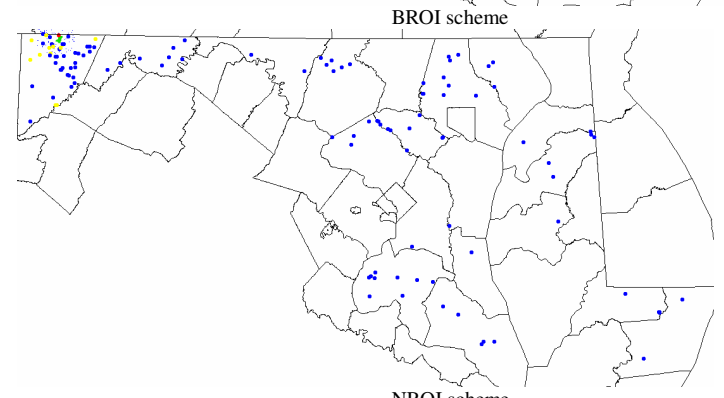
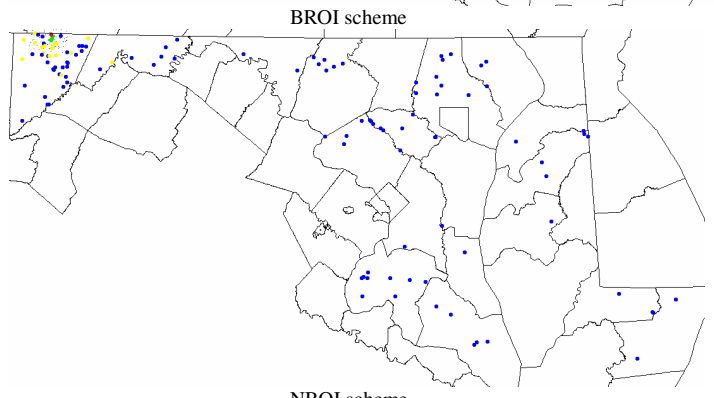
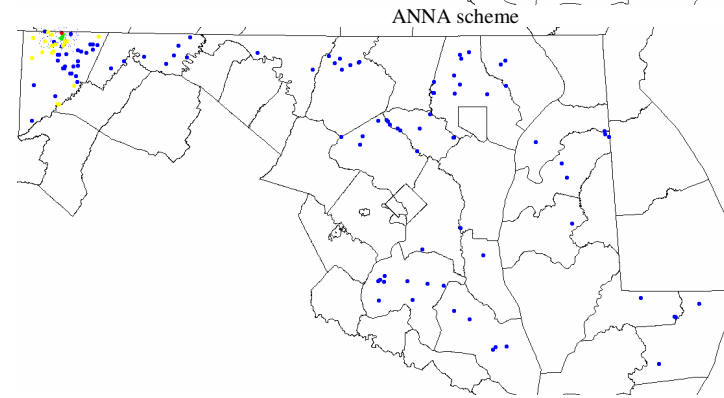
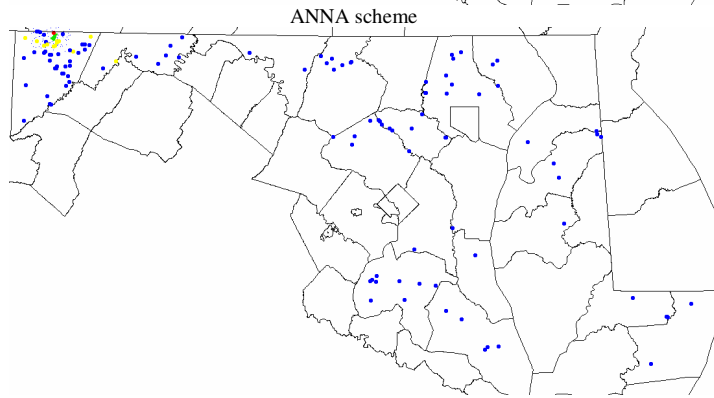
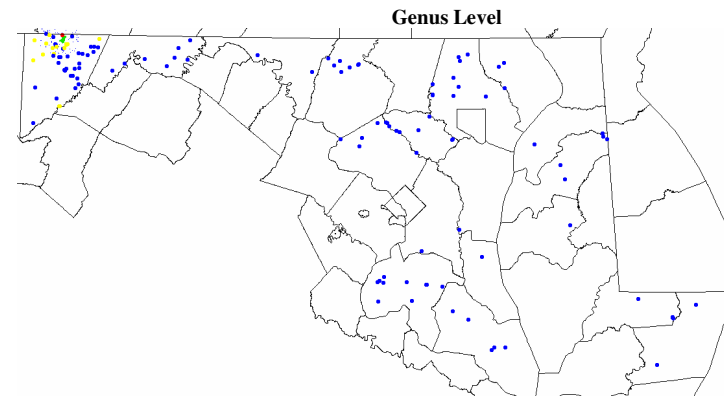
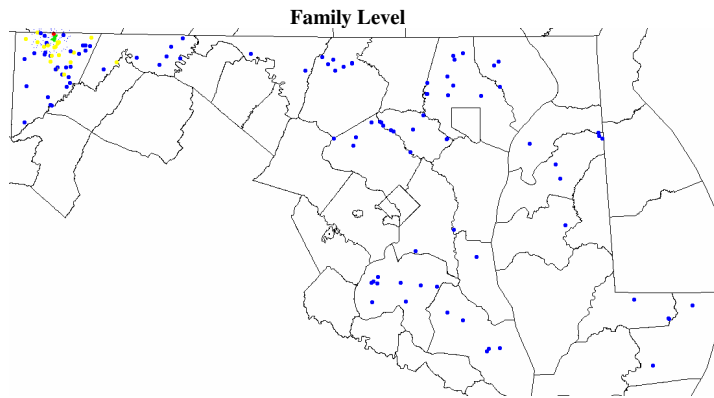
NROI scheme



NROI scheme

NROI scheme

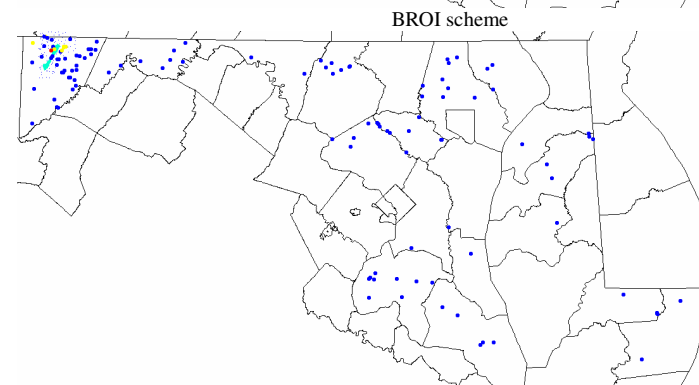
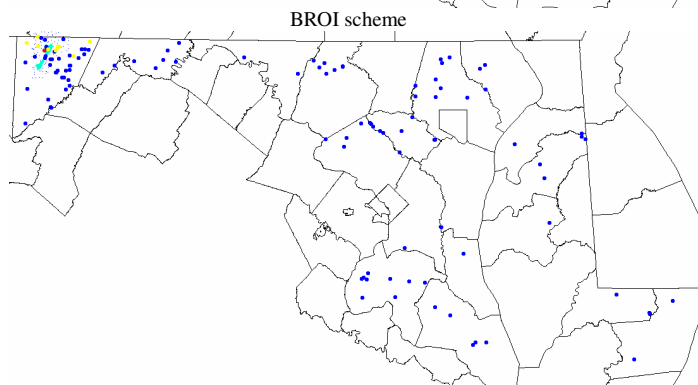
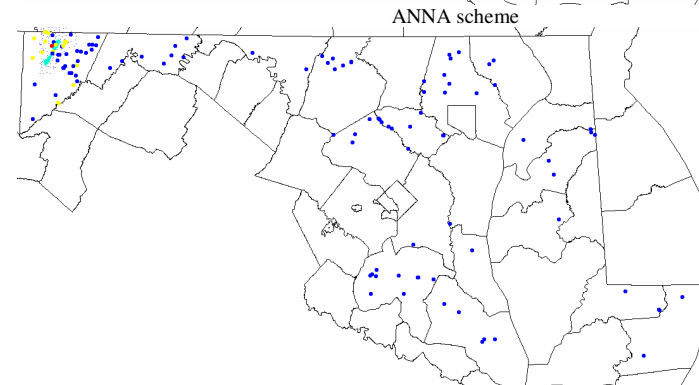
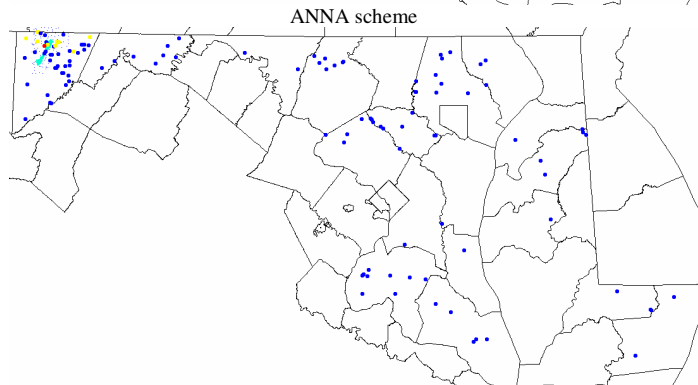
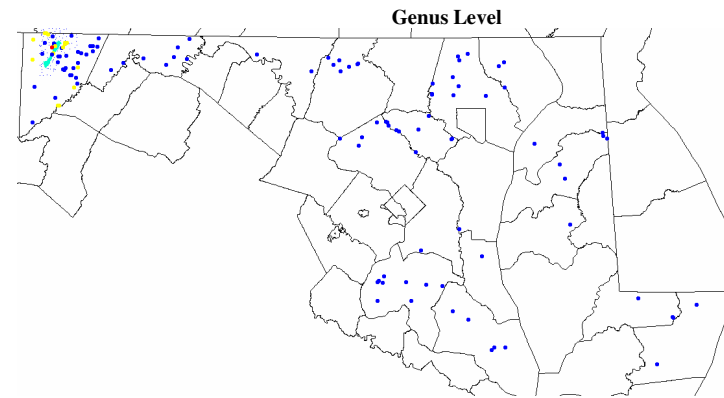
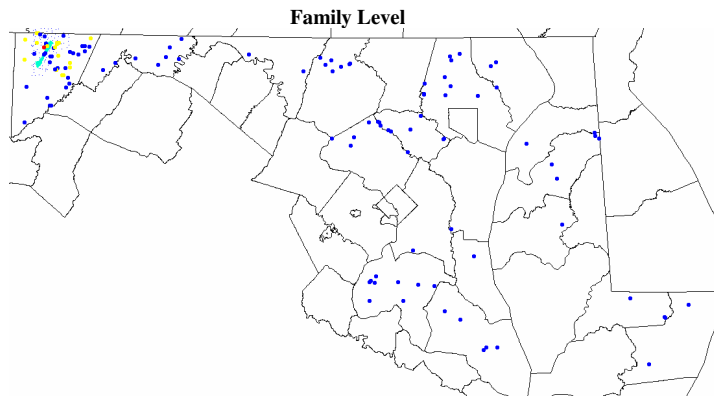
Young-203-c-2000



NROI scheme

NROI scheme

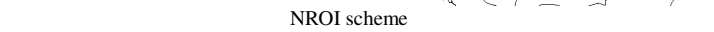
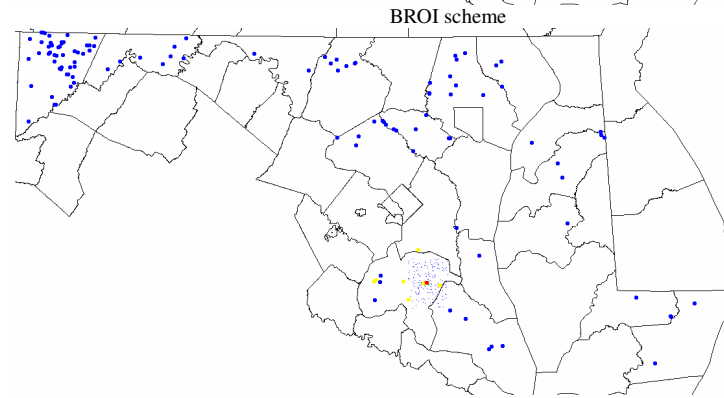
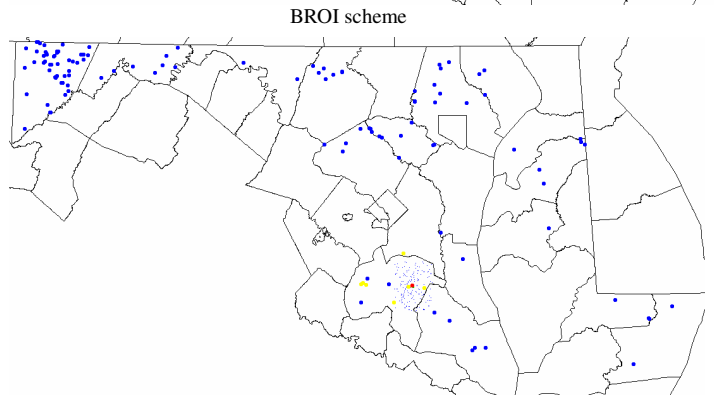
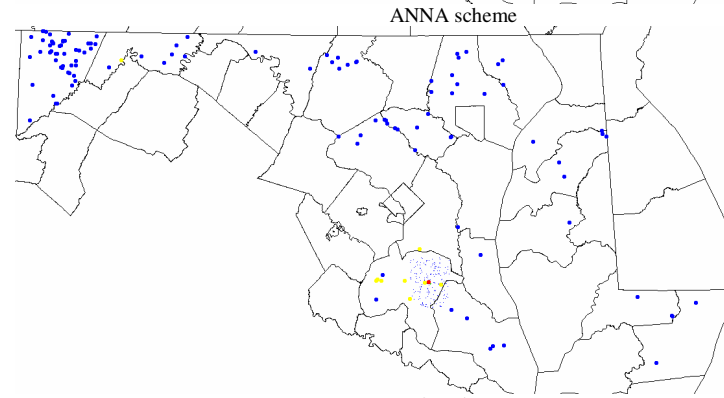
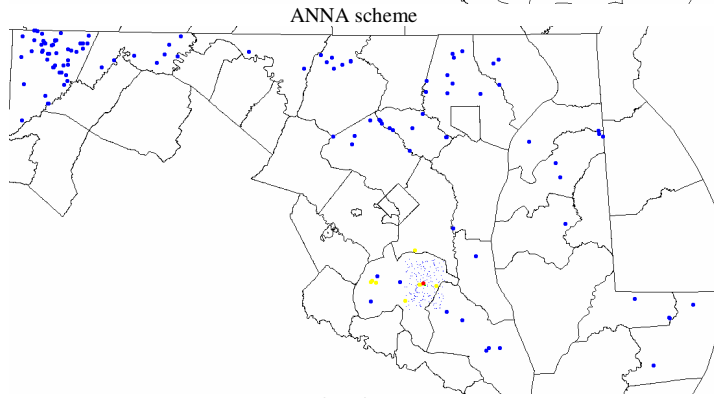
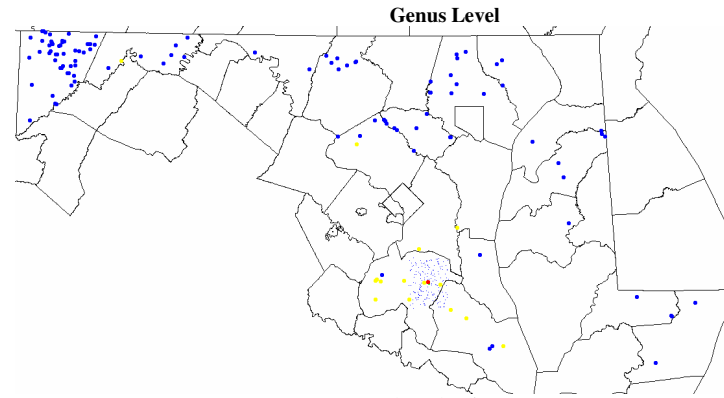
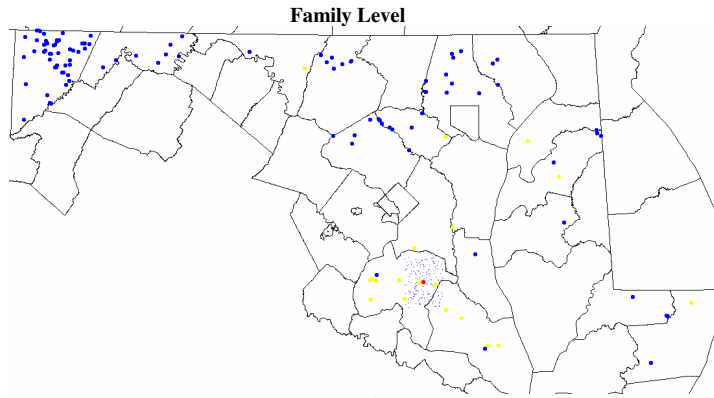
Young-432-s-2001



NROI scheme

NROI scheme

Zeki-106-r-2001



Appendix E. Original Predicted Taxa Probabilities At Stream Restoration Sites

Start from the following page.

Biodiversity for LFLF 301B

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TRCI AIC	Tubificidae	0.05	0.13	0.11	Lepidostoma	0.05	0.04	0
Aeshnidae	0	0	0	Uenoidae	0	0	0	Leptophlebia	0.12	0.11	0
Ameletidae	0	0	0	Unionidae	0	0	0	Leptoxis	0	0	0
Ancylidae	0	0	0	Valvatidae	0	0	0	Leucotrichia	0	0	0
Asellidae	0.11	0.25	0.12	Velidae	0	0	0	Leucrocota	0	0	0
Astacidae	0	0	0	Viviparidae	0	0	0	Leuctridae	0.22	0.21	0
Athericidae	0	0	0					Limmophila	0.06	0.04	0
Baelidae	0.59	0.5	0.64					Limmophyes	0.06	0.06	0
Baeliscidae	0	0	0					Lopescladius	0	0	0
Belostomatidae	0	0	0					Lumbriculidae	0.17	0.24	0
Bithyniidae	0	0	0					Lype	0	0.03	0
Blephariceridae	0	0	0					Macronychus	0	0	0
Brachycentridae	0.06	0.13	0.08					Menetus	0	0.03	0
Braconidae	0	0	0					Menopselopia	0.12	0.07	0
Branchiobdellidae	0	0	0					Micrasema	0	0	0
Caenidae	0	0	0					Microsectra	0.55	0.4	0.44
Calamoceratidae	0	0	0					Microtendipes	0.08	0.07	0.28
Calopterygidae	0.1	0	0.03					Microvelia	0	0	0
Cambariae	0	0	0					Molophilus	0	0	0
Capniidae	0	0	0.03					Musculium	0	0.03	0
Carabidae?	0	0	0					Naididae	0.38	0.29	0.45
Ceratopogonidae	0.11	0.12	0.1					Nanocladius	0.05	0.07	0
Chironomidae	1	1	1					Nematomorpha	0	0	0
Chloroperlidae	0.09	0.12	0.12					Nemoura	0	0	0
Coenagrionidae	0	0	0					Neophylax	0.49	0.53	0.45
Collembola	0	0	0					Nigronia	0.18	0.18	0
Corbiculidae	0	0	0					Nise	0	0	0
Cordulegastridae	0	0	0					Oecetis	0	0.03	0
Cordulidae	0	0	0					Oemopteryx	0	0	0
Coxidae	0	0	0					Optioservus	0.37	0.39	0.4
Corydalidae	0.2	0.26	0.23					Ormosia	0	0	0
Crangonyctidae	0	0	0					Orthoclaudiinae(A)	0	0.03	0
Culicidae	0	0	0					Ostrocerca	0	0	0
Curculionidae	0	0	0					Oulimnius	0.62	0.54	0.6
Diplostrota	0	0	0					Pagastia	0.07	0.04	0.01
Dividae	0.04	0	0.03					Palaemonetes	0	0	0
Dolichopodidae	0	0	0					Paracapnia	0	0	0
Dryopidae	0.15	0.13	0.13					Parachaeoclaadius	0	0.03	0
Dugesidae	0	0	0					Parachironomus	0.06	0.04	0
Dytiscidae	0	0	0					Paragnetra	0	0	0
Elmidae	0.86	0.88	0.9					Parakiefferiella	0	0.03	0
Empididae	0.55	0.62	0.61					Paraleptophlebia	0.13	0.11	0.01
Enchytraeidae	0	0	0					Paramerina	0.06	0.04	0
Ephemeroidea	0.94	0.87	0.92					Parametrichironomus	0.56	0.64	0.41
Ephemeroidea	0	0	0					Paraphaenoclaadius	0.06	0.03	0
Ephydriidae	0	0	0					Paratanytarsus	0.14	0.18	0.28
Eripodellidae	0	0	0					Pedicia	0	0.03	0
Gammaridae	0.26	0.25	0.13					Peltodytes	0.06	0.07	0
Gerridae	0	0	0					Peltoperla	0	0	0
Glossiphoniidae	0	0	0					Perlesta	0	0	0
Glossosomatidae	0.29	0.25	0.26					Phaenopspectra	0.08	0.04	0.28
Gomphidae	0.19	0.37	0.24					Phasganophora	0	0	0
Gordidae	0.05	0.12	0.04					Phylocentropus	0	0	0
Gynptidae	0	0	0					Physidae	0.12	0.11	0
Halipidae	0	0	0					Pisidium	0.07	0.04	0.01
Helodidae	0	0	0					Planorbella	0	0	0
Heptageniidae	0.62	0.63	0.65					Platycentropus	0	0.03	0
Hydrobiidae	0	0	0					Platysmittia	0.05	0.04	0
Hydrophilidae	0.04	0	0.04					Polycentropus	0.31	0.22	0.13
Hydropsychidae	0.82	0.75	0.75					Polypedium	0.3	0.32	0.13
Hydrophilidae	0.11	0.12	0.07					Pothastia	0.07	0.04	0.12
Lepidoptera	0	0	0					Probezia	0	0.07	0
Lepidostomatidae	0.04	0	0.03					Procladius	0	0.03	0
Leptoceridae	0	0	0					Promoresia	0	0	0
Leptophlebiidae	0.13	0	0.11					Prosimulium	0.69	0.61	0.87
Lestidae	0.09	0	0.08					Prostoia	0.25	0.25	0.32
Leucotrichia	0	0	0					Prostoma	0	0	0
Limnephilidae	0.53	0.38	0.56					Psephenus	0.05	0.04	0
Lumbriculidae	0.05	0	0.09					Pseudochironomus	0	0	0
Lymnaeidae	0.06	0.13	0.08					Pseudolimnophila	0.12	0.07	0
Macromiidae	0	0	0					Pseudorthoclaadius	0	0.03	0
Maxillopoda	0	0	0					Pseudosuccinea	0	0	0
Metretropodidae	0	0	0					Ptilotreta	0	0.03	0
Molannidae	0	0	0					Pteronarcys	0.05	0.04	0
Muscidae	0	0	0					Ptilostomis	0	0	0
Naididae	0.36	0.37	0.33					Pycnopsyche	0.12	0.11	0
Nemouridae	0.83	0.75	0.8					Radix	0	0	0
Nepidae	0	0	0					Rheocricotopus	0.11	0.11	0
Notonectidae	0	0	0					Rheotanytarsus	0.19	0.15	0.13
Odontoceridae	0	0	0					Rhyacophila	0.57	0.47	0.87
Oligoneuriidae	0.1	0	0.07					Serratella	0.22	0.18	0.01
Ostracoda	0	0	0					Sialis	0.06	0.04	0
Palaemonidae	0.04	0	0.01					Simulium	0.34	0.28	0.01
Pettoperilidae	0.05	0	0.04					Siphonurus	0	0	0
Perilidae	0.35	0.25	0.23					Siphonoplecton	0	0	0
Periodidae	0.44	0.38	0.4					Sphaeriidae	0.06	0.04	0
Philopotamidae	0.3	0.38	0.34					Sphaerium	0	0.03	0
Phyganeidae	0	0	0					Sphaeromias	0	0	0
Physidae	0.04	0	0.04					Stagopterna	0.36	0.36	0.12
Piscicolidae	0	0	0					Stempellina	0	0	0
Pisididae	0.1	0.13	0.13					Stempellina	0.05	0.07	0
Planariidae	0.09	0.12	0.07					Stenacron	0	0	0
Planorbidae	0	0	0					Steneimis	0.08	0.11	0.32
Pleuroceridae	0.1	0	0.03					Stenonema	0.42	0.29	0.13
Polycentropodidae	0.23	0.13	0.2					Stictochironomus	0.06	0.04	0
Potamanthidae?	0	0	0					Stilocladius	0	0.03	0
Psephenidae	0.05	0.12	0.04					Strophopteryx	0.12	0.11	0
Psychodidae	0	0	0					Stygobromus	0	0	0
Psychomyiidae	0.05	0.12	0.04					Symposiocladius	0	0	0
Pteronarcidae	0.09	0	0.06					Symphylotarsus	0.07	0.07	0.12
Ptiloctylidae?	0.13	0.12	0.16					Synurella	0	0.03	0
Ptychopteridae	0	0	0					Tabanus	0	0	0
Rhyacophilidae	0.58	0.5	0.55					Taeniopteryx	0	0	0
Sakidae	0	0	0					Talaperla	0	0	0
Scomyzidae	0	0	0					Tanytarsus	0	0.03	0
Sericostomatidae	0	0	0					Tanytarsus	0.3	0.35	0.13
Sialidae	0	0	0					Thienemannella	0.18	0.18	0.12
Simuliidae	0.95	0.87	0.91					Thienemannimyia(grp)	0.06	0.04	0
Siphonuridae	0.24	0.25	0.32					Tipula	0.35	0.36	0
Sisyridae	0	0	0					Trepobates	0	0	0
Sphaeriidae	0	0	0					Trianaodes	0.06	0.07	0
Stratiomyidae	0	0	0					Tribeles	0.06	0.04	0
Syrphidae	0	0	0					Trisopelopia	0.3	0.25	0.12
Tabanidae	0	0	0					Tubificidae	0.13	0.18	0.12
Taeniopterygidae	0.09	0	0.04					Tvetenia	0.07	0.07	0.12
Talitridae	0	0	0					Unniella	0	0	0
Tetrastemmatidae	0	0	0					Wormaldia	0.06	0.04	0
Tipulidae	0.78	0.51	0.68					Xyloptopus	0	0	0
								Zavreimyia	0.27	0.15	0.44

Biodiversity for LFLF 301C:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.11	0.22	0.15
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.6	0.56	0.59
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.06	0.11	0.11
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.09	0	0
Cambariae	0	0	0
Capniidae	0	0	0
Carabidae?	0	0	0
Ceratopogonidae	0.1	0.11	0.11
Chironomidae	1	1	1
Chloroperlidae	0.09	0.11	0.13
Coenagrionidae	0	0	0
Collembola	0	0	0
Corbiculidae	0	0	0
Cordulegasteridae	0	0	0
Cordulidae	0	0	0
Corixidae	0	0	0
Corydalidae	0.2	0.23	0.28
Crangonyctidae	0	0	0
Culicidae	0	0	0
Culicoidae	0	0	0
Diplostroca	0	0	0
Dixidae	0.04	0	0
Dolichopodidae	0	0	0
Dryopidae	0.15	0.11	0.11
Dugesidae	0	0	0
Dytiscidae	0	0	0
Etmidae	0.86	0.89	0.95
Empididae	0.55	0.66	0.62
Enchytraeidae	0	0	0
Ephemereididae	0.94	0.89	0.89
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ergobdellidae	0	0	0
Gammaridae	0.23	0.22	0.15
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.29	0.33	0.27
Gomphidae	0.16	0.33	0.33
Gordidae	0.05	0.11	0.04
Gyrinidae	0	0	0
Halplidae	0	0	0
Helodidae	0	0	0
Hepptagenidae	0.66	0.67	0.66
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0.02
Hydropsychidae	0.82	0.78	0.74
Hydroptilidae	0.1	0.11	0.04
Lepidoptera	0	0	0
Lepidostomatidae	0.04	0	0
Lepidostomatidae	0.04	0	0
Lepidostomatidae	0.13	0.11	0.08
Lestidae	0	0	0
Leuctridae	0.09	0	0
Libellulidae	0	0	0
Limnephilidae	0.49	0.44	0.51
Limnephilidae	0.05	0	0.07
Lymnaeidae	0.06	0.11	0.11
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.36	0.44	0.36
Nemouridae	0.79	0.67	0.76
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.13	0	0.01
Ostracoda	0	0	0
Palaemonidae	0.04	0	0
Peltoperlidae	0.05	0	0
Perlidae	0.36	0.33	0.23
Perlidae	0.43	0.34	0.36
Philopotamidae	0.3	0.34	0.36
Phryganeidae	0	0	0
Physidae	0.04	0	0.02
Piscicolidae	0	0	0
Pisidiidae	0.11	0.22	0.18
Planariidae	0.09	0.11	0.04
Planorbidae	0	0	0
Pleuroceridae	0.09	0	0
Polycentropodidae	0.2	0.22	0.21
Potamanthidae?	0	0	0
Psephenidae	0.05	0.11	0.04
Psychodidae	0	0	0
Psychomyiidae	0.05	0.11	0.04
Pteronarcidae	0.09	0	0
Ptilodactylidae?	0.13	0.22	0.18
Ptychopteridae	0	0	0
Rhyacophilidae	0.54	0.45	0.54
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.88	0.85
Siphonuridae	0.28	0.22	0.3
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.05	0	0
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.78	0.56	0.6
Tubificidae	0.09	0.12	0.15

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Uenoidea	0	0	0
Unionidae	0	0	0
Valvulidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Ablabesmyia	0.06	0.04	0
Acentrella	0.11	0.11	0
Acepernia	0.23	0.22	0
Acroneria	0.07	0.04	0.05
Agapetus	0	0.03	0
Allocapnia	0.12	0.15	0
Alotanypus	0	0.04	0
Ameletus	0.26	0.3	0.42
Amphimemura	0.72	0.7	0.42
Anchytarsus	0.19	0.15	0.31
Ancylidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.19	0.19	0.17
Aspectrotanypus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.06	0.04	0
Bezzia	0	0	0
Bittacomorpha	0.06	0.04	0
Blepharicera	0	0	0
Boyeria	0.11	0.11	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brillia	0	0.07	0
Caecidotea	0.06	0.11	0
Caenis	0.06	0.07	0
Callibaetis	0.06	0.04	0
Calopteryx	0.06	0.04	0
Cambariae	0.06	0.04	0
Centropilum	0.06	0.04	0
Ceratopogon	0.21	0.15	0.3
Chaetocladius	0.05	0.04	0
Chelifera	0.08	0.04	0.27
Chneumatopsyche	0.38	0.41	0.48
Chimarra	0.12	0.08	0.12
Chironomus	0	0.04	0
Chloroperlidae	0.13	0.18	0.27
Chrysops	0	0	0
Cinygmula	0	0	0
Cliocapnia	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.11	0.18	0
Clinotanytus	0	0	0
Cloperla	0	0.07	0
Cnephia	0	0	0
Collinella	0.06	0.04	0
Conchapelopia	0.63	0.56	0.74
Cordulegaster	0.06	0.04	0
Cordulidae	0.06	0.04	0
Corixidae	0.06	0.04	0
Corydalus	0	0	0
Corynoneura	0.33	0.23	0.42
Crangonyx	0.12	0.11	0
CricotopusOrthocladius	0.55	0.63	0.58
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.06	0.04	0
Diamesa	0.27	0.16	0.52
Dicranota	0.11	0.11	0
Dicrolentipes	0	0.04	0
Dineutus	0	0	0
Dipheter	0.06	0.04	0
Diploctroza	0.39	0.34	0.42
Diplocladius	0	0.07	0
Diploperla	0.07	0.08	0.12
Dixa	0.05	0.07	0
Dolichopodidae	0.06	0.04	0
Dolophlodes	0.11	0.07	0
Drunella	0.05	0.07	0
Dubiraphia	0.05	0.07	0
Dugesia	0.05	0.04	0
Ecoptura	0.06	0.11	0
Ectopria	0	0	0
Enallagma	0.06	0.04	0
Enchytraeidae	0.06	0.07	0
Endochironomus	0.06	0.07	0
Epeorus	0.24	0.19	0.3
Ephemera	0	0	0
Ephemerella	0.7	0.63	1
Erioptera	0	0	0
Eukiefferiella	0.25	0.36	0.27
Eurytophella	0.2	0.12	0.3
Falceon	0	0	0
Gammarus	0.06	0.07	0
Glossosoma	0.12	0.11	0.05
Gomphidae	0.22	0.12	0.56
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleniella	0	0.03	0
Helichus	0.05	0.04	0
Helius	0	0	0
Hemirodromia	0.34	0.23	0.7
Heptagenia	0	0	0
Heterotrissocladius	0.12	0.08	0
Hexagenia	0	0	0
Hexatoma	0.13	0.08	0.12
Hyalella	0	0.03	0
Hydrobaenus	0.14	0.08	0.27
Hydrobius	0.06	0.07	0
Hydroporus	0	0	0
Hydropsyche	0.3	0.26	0.17
Ironoquia	0	0.04	0
Isonychia	0	0.03	0
Isoperla	0.12	0.08	0
Kiefferulus	0.06	0.04	0
Krenopelopia	0.06	0.04	0
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0.06	0.04	0
Lepidostoma	0.05	0.04	0
Leptophlebia	0.12	0.11	0

Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.22	0.22	0
Limnophila	0.06	0.04	0
Limnophyes	0.06	0.04	0
Lopescladius	0	0	0
Lumbriculidae	0.16	0.25	0
Lype	0	0.04	0
Macronychus	0	0	0
Menetus	0	0.04	0
Meropelopia	0.12	0.08	0
Micrasema	0	0	0
Micropsectra	0.55	0.38	0.42
Microtendipes	0.08	0.07	0.27
Microvelia	0	0	0
Molophilus	0	0	0
Muscilium	0	0.04	0
Naididae	0.38	0.3	0.48
Nanocladius	0.05	0.04	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.49	0.55	0.48
Nigronia	0.18	0.19	0
Nixe	0	0	0
Ocetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.37	0.41	0.43
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0.03	0
Ostrocerca	0	0	0
Oulimnius	0.62	0.56	0.62
Pagastia	0.07	0.04	0.05
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaetocladius	0	0.03	0
Parachironomus	0.06	0.04	0
Paragnetmia	0	0	0
Parakiefferiella	0	0.03	0
Paraleptophlebia	0.13	0.11	0.05
Parameria	0.06	0.04	0
Parametricnemus	0.56	0.66	0.44
Paraphaenocladus	0	0.04	0
Paratanytarsus	0.14	0.18	0.27
Pedicia	0	0.04	0
Pelodytes	0.06	0.07	0
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.08	0.04	0.27
Phaenopsectra	0	0	0
Philocentropus	0	0	0
Physidae	0.12	0.11	0
Pisidium	0.07	0.04	0.05
Planorbella	0	0	0
Platycentropus	0	0	0
Platymitria	0.05	0.04	0
Platycentropus	0.31	0.18	0.17
Polypedilum	0.3	0.33	0.17
Pothastia	0.07	0.04	0.12
Probezzia	0	0.07	0
Procladius	0	0.04	0
Promeresia	0	0	0
Prosimulium	0.69	0.63	0.83
Prostia	0.25	0.26	0.3
Prostoma	0	0	0
Psephenus	0.05	0.04	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.12	0.08	0
Pseudorthocladius	0	0.04	0
Pseudosuccinea	0	0	0
Psilotreta	0	0.03	0
Pteronarcys	0.05	0.04	0
Ptilostomis	0	0	0
Pycnopsyche	0.12	0.11	0
Radix	0	0	0
Rhaeciclotopus	0.11	0.11	0
Rheotanytarsus	0.19	0.15	0.17
Rhyacophila	0.57	0.48	0.83
Serratella	0.23	0.18	0.05
Sialis	0.06	0.04	0
Simulium	0.34	0.29	0.05
Siphonurus	0	0	0
Siphonoclepton	0	0	0
Sphaeriidae	0.06	0.04	0
Sphaerium	0	0.03	0
Sphaeromias	0	0	0
Stegopterna	0.36	0.37	0.12
Stenopelopia	0	0	0
Stempellina	0.05	0.07	0
Stenacron	0	0	0
Stenelmis	0.08	0.11	0.3
Stenonema	0.42	0.3	0.17
Stictochironomus	0.06	0.04	0
Stilocladus	0	0.04	0
Strophopteryx	0.12	0.11	0
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphothastia	0.07	0.07	0.12
Synurella	0	0.03	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytarsus	0	0.04	0
Tanytarsus	0.3	0	

Biodiversity for LRCR 101A:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Aselidae	0.1	0.09	0.02
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.6	0.72	0.61
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.1	0	0
Cambariae	0	0	0
Capniidae	0	0	0
Carabidae?	0	0	0
Ceratopogonidae	0.11	0.09	0.15
Chironomidae	1	1	1
Chloroperlidae	0.08	0.09	0.09
Coenagrionidae	0	0	0
Collembola	0	0	0
Corticulidae	0	0	0
Cordulegasteridae	0	0	0
Corduliidae	0	0	0
Corixidae	0	0	0
Corydalidae	0.22	0.28	0.32
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0.08	0	0.04
Dolichopodidae	0	0	0
Dryopidae	0.14	0.09	0.1
Dugesidae	0	0	0
Dytiscidae	0	0	0
Etmidae	0.82	0.91	1
Empididae	0.5	0.54	0.5
Enchytraeidae	0	0	0
Ephemereidae	0.95	1	1
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ergasilidae	0	0	0
Gammaridae	0.25	0.09	0.02
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.32	0.27	0.06
Gomphidae	0.13	0.09	0.19
Gordidae	0.05	0.09	0.02
Gyrinidae	0	0	0
Halplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.64	0.73	0.76
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0.09	0.08
Hydropsychidae	0.83	0.91	0.8
Hydroptilidae	0.11	0.09	0.02
Lepidoptera	0	0	0
Lepidostomatidae	0.08	0	0.04
Lepidostomatidae	0	0	0
Lepidostomatidae	0.17	0.27	0.16
Lestidae	0	0	0
Leuctridae	0.12	0	0.09
Libellulidae	0	0	0
Limnephilidae	0.48	0.54	0.68
Lumbriculidae	0.05	0.09	0.1
Lymnaeidae	0.05	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Muscidae	0.59	0.37	0.47
Nemouridae	0.84	0.82	0.95
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.1	0.09	0.06
Ostracoda	0	0	0
Palaemonidae	0.04	0	0
Peltoperlidae	0.05	0.09	0.05
Perlidae	0.37	0.36	0.1
Periodidae	0.42	0.45	0.46
Philopotamidae	0.33	0.37	0.43
Phryganeidae	0	0	0
Physidae	0.04	0.09	0.08
Piscicolidae	0	0	0
Pisidiidae	0.09	0.09	0.03
Planariidae	0.1	0.09	0.06
Planorbidae	0	0	0
Pleuroceridae	0.1	0	0
Polycentropodidae	0.2	0.19	0.26
Potamanthidae?	0	0	0
Psephenidae	0.05	0.09	0.02
Psychodidae	0	0	0
Psychomyiidae	0.05	0.09	0.02
Pteronarcidae	0.09	0.09	0.1
Ptilodactylidae?	0.12	0.09	0.12
Ptychopteridae	0	0	0
Rhyacophilidae	0.56	0.37	0.6
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.95	0.9	0.82
Siphonuridae	0.24	0.28	0.39
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.06	0	0
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.76	0.82	0.79
Tubificidae	0.05	0.1	0.18

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Ablabesmyia	0.07	0.06	0.13
Acentrella	0.12	0.12	0.11
Acepernia	0.19	0.18	0.18
Acroneria	0.07	0.06	0.13
Agapetus	0	0	0
Allocapnia	0.24	0.23	0.18
Alotanypus	0	0	0
Ameletus	0.4	0.36	0.56
Amphinemura	0.75	0.76	0.74
Anchytarsus	0.19	0.18	0.21
Ancylidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.26	0.24	0.34
Aspectrotanypus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.07	0.06	0.11
Bezzia	0	0	0
Bittacomorpha	0	0.06	0
Blepharicera	0	0	0
Boyeria	0	0.06	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brillia	0.05	0.06	0
Caecidotea	0.05	0.06	0
Caenis	0.05	0.06	0
Callibaetis	0.05	0.06	0
Calopteryx	0.05	0.06	0
Cambariae	0.07	0.06	0.13
Centropilum	0.07	0.06	0.13
Ceratopogon	0.22	0.18	0.42
Chaetocladius	0.05	0.06	0
Chelifera	0.06	0.06	0.05
Chusumatopsyche	0.41	0.46	0.57
Chimarra	0.14	0.12	0.21
Chironomus	0	0	0
Chloroperlidae	0.06	0.06	0.05
Chrysops	0	0	0
Cinygmula	0	0	0
Ciliocpeima	0	0	0
Ciadorantarsus	0	0	0
Clinocera	0.17	0.17	0
Clinotanypus	0	0	0
Cloperla	0	0	0
Cnephia	0	0	0
Collembola	0.07	0.06	0.13
Conchapelopia	0.58	0.59	0.73
Cordulegaster	0.07	0.06	0.13
Corduliidae	0.05	0.06	0
Corixidae	0.05	0.06	0
Corydalus	0	0	0
Corynoneura	0.3	0.3	0.67
Crangonyx	0.06	0.12	0
CricotopusOrthocladius	0.54	0.53	0.33
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.05	0.06	0
Diamesa	0.26	0.24	0.23
Dicranota	0.12	0.12	0.11
Dicrolentipes	0	0	0
Dineutus	0	0	0
Dipheteron	0.05	0.06	0
Diplectrona	0.35	0.36	0.62
Diplocladius	0	0	0
Diploperla	0.08	0.06	0.21
Dixa	0.05	0.06	0
Dolichopodidae	0.07	0.06	0.13
Dolophlodes	0.11	0.11	0
Drunella	0.11	0.11	0.16
Dubiraphia	0.05	0.06	0
Dugesia	0.05	0.06	0
Ecoptura	0.07	0.06	0.07
Ectopria	0	0	0
Enallagma	0.05	0.06	0
Enchytraeidae	0.07	0.06	0.13
Endochironomus	0.05	0.06	0
Epeorus	0.25	0.23	0.23
Ephemera	0	0	0
Ephemerella	0.69	0.65	0.69
Erioptera	0	0	0
Eukiefferiella	0.24	0.24	0.16
Eurytophella	0.21	0.18	0.35
Falceon	0	0	0
Gammarus	0.06	0.06	0
Glossosoma	0.18	0.18	0.13
Gomphidae	0.21	0.18	0.3
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleniella	0	0	0
Helichus	0.05	0.06	0
Helius	0	0	0
Hemirodromia	0.33	0.3	0.47
Heptagenia	0	0	0
Heterotrissocladius	0.07	0.12	0.13
Hexagenia	0	0	0
Hexatoma	0.15	0.12	0.29
Hyalella	0	0	0
Hydrobaenus	0.12	0.12	0.05
Hydrobius	0.07	0.06	0.13
Hydroporus	0	0	0
Hydropsyche	0.26	0.24	0.34
Ironoquia	0	0	0
Isonychia	0	0	0
Isoperla	0.14	0.12	0.23
Kiefferulus	0.05	0.06	0
Krenopelopia	0.07	0.06	0.07
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0.07	0.06	0.13
Lepidostoma	0.05	0.06	0
Leptophlebia	0.12	0.12	0.13

Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.16	0.23	0
Limnophila	0	0.06	0
Limnophyes	0.07	0.06	0.07
Lopescladius	0	0	0
Lumbriculidae	0.18	0.17	0.13
Lype	0	0	0
Macronychus	0	0	0
Menetus	0	0	0
Meropelopia	0.13	0.12	0.2
Micrasema	0	0	0
Nanocladius	0.53	0.48	0.74
Microtendipes	0.06	0.06	0.05
Microvelia	0	0	0
Molophilus	0	0	0
Muscilium	0	0	0
Naididae	0.46	0.42	0.68
Nanocladius	0.05	0.06	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.63	0.59	0.68
Nigronia	0.19	0.18	0.2
Nixe	0	0	0
Ocetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.43	0.41	0.42
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostrocerca	0	0	0
Oulimnius	0.57	0.53	0.59
Pagastia	0.07	0.06	0.13
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaetocladius	0	0	0
Parachironomus	0.05	0.06	0
Paragnetmia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.13	0.12	0.2
Paramerina	0.05	0.06	0
Parametricnemus	0.62	0.65	0.57
Paraphaenocladus	0	0	0
Paralutyrus	0.13	0.12	0.16
Pedicia	0	0	0
Pelodytes	0.05	0.06	0
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.06	0.06	0.05
Phaeganophora	0	0	0
Phycocentropus	0	0	0
Physidae	0.12	0.12	0.13
Pisidium	0.07	0.06	0.13
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0.05	0.06	0
Prosimulium	0.26	0.29	0.34
Polypedilum	0.38	0.35	0.45
Pothastia	0.08	0.06	0.21
Probezzia	0	0	0
Procladius	0	0	0
Promeresia	0	0	0
Prosimulium	0.69	0.65	0.66
Prostoma	0.32	0.3	0.33
Pseudocricotopus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.13	0.12	0.2
Pseudorthocladius	0.05	0.06	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0.05	0.06	0
Ptilostomis	0	0	0
Pycnopsyche	0.07	0.12	0.07
Radix	0	0	0
Rhaecricotopus	0	0.06	0
Rheotanytarsus	0.2	0.18	0.34
Rhyacophila	0.61	0.59	0.42
Serratella	0.23	0.23	0.13
Sialis	0.07	0.06	0.07
Simulium	0.28	0.35	0.13
Siphonurus	0	0	0
Siphonoclepton	0	0	0
Sphaeriidae	0.07	0.06	0.13
Sphaerium	0	0	0
Sphaeromias	0	0	0
Stegopterna	0.27	0.3	0.41
Stenopelopia	0	0	0
Stempellina	0.05	0.06	0
Stenacron	0	0	0
Stenelmis	0.09	0.06	0.23
Stenonema	0.44	0.41	0.54
Stictochironomus	0.07	0.06	0.07
Stilocladus	0.05	0.06	0
Strophopteryx	0.18	0.18	0.11
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphothastia	0.08	0.06	0.21
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytarsus	0	0	0
Tanytarsus	0.26	0.29	0.34
Thienemanniella	0.2	0.18	0.32
Thienemanniella(grp)	0.07	0.06	0.11
Tipula	0.35	0.41	0.2
Trepobates	0	0	0

Biodiversity for LRJB 203A:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.11	0.1	0.05
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.64	0.6	0.63
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.05	0	0
Cambaridae	0	0	0
Capniidae	0	0.19	0.05
Carabidae?	0	0	0
Ceratopogonidae	0.09	0	0
Chironomidae	1	1	1
Chloroperlidae	0.08	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corticulidae	0	0	0
Cordulegastriidae	0	0	0
Cordulidae	0	0	0
Corydalidae	0.19	0.1	0.18
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0.04	0	0
Dolichopodidae	0	0	0
Dryopidae	0.13	0	0
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.9	0.8	0.91
Empididae	0.61	0.8	0.77
Enchytraeidae	0	0	0
Ephemereilidae	0.95	1	1
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ergobdellidae	0	0	0
Gammaridae	0.18	0.1	0.07
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.35	0.5	0.44
Gomphidae	0.13	0	0
Gordiidae	0.06	0.1	0.05
Gyrinidae	0	0	0
Haliplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.67	0.61	0.73
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.84	0.9	0.95
Hydroptilidae	0.11	0.1	0.05
Lepidoptera	0	0	0
Lepidostomatidae	0.04	0	0
Lepidoceridae	0	0	0
Leptophlebiidae	0.14	0.1	0.18
Lestidae	0	0	0
Leuctridae	0.08	0.1	0.02
Libellulidae	0	0	0
Limnephilidae	0.5	0.5	0.45
Lumbriculidae	0.08	0.2	0.11
Lymnaeidae	0.05	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.36	0.4	0.38
Nemouridae	0.81	0.7	0.64
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.14	0.1	0.11
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.05	0	0
Perlidae	0.37	0.3	0.32
Perlotidae	0.38	0.2	0.23
Philopotamidae	0.36	0.4	0.47
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.1	0.1	0.15
Planariidae	0.1	0.1	0.05
Planorbidae	0	0	0
Pleuroceridae	0.05	0	0
Polycentropodidae	0.2	0.21	0.31
Potamanthidae?	0	0	0
Psephenidae	0.06	0.1	0.05
Psychodidae	0	0	0
Psychomyiidae	0.06	0.1	0.05
Pteronarcidae	0.08	0	0
Ptilodactylidae?	0.13	0.1	0.15
Ptychopteridae	0	0	0
Rhyacophilidae	0.55	0.4	0.41
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.9	0.82
Siphonuridae	0.32	0.6	0.53
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.05	0.1	0.03
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.78	0.8	0.81
Tubificidae	0.1	0.2	0.29

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Ablabesmyia	0.06	0.05	0
Acetrotella	0.06	0.15	0
Acerpenna	0.18	0.15	0.12
Acroneria	0.08	0.05	0.23
Agapetus	0	0.05	0
Allocapnia	0.23	0.2	0.12
Aklotrypys	0	0	0
Ameletus	0.37	0.4	0.33
Amphinemura	0.73	0.75	0.64
Anchytarsus	0.15	0.15	0.25
Ancyidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.26	0.2	0.36
Apsectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.06	0.05	0
Bezzia	0	0	0
Bitacomorpha	0.06	0.05	0
Blepharicera	0	0	0
Boeyia	0.11	0.1	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brilia	0.05	0.1	0
Caecidotea	0.05	0.05	0
Caenis	0.05	0.05	0
Calibaetis	0.05	0.05	0
Calopteryx	0.05	0.05	0
Cambaridae	0.06	0.05	0
Centroptilum	0.06	0.05	0
Ceratopogon	0.15	0.15	0.32
Chaetocladius	0	0.05	0
Chellifera	0.07	0.05	0.11
Cheumatopsyche	0.41	0.4	0.75
Chimarra	0.12	0.1	0.13
Chironomus	0	0	0
Chloroperlidae	0.07	0.05	0.11
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.18	0.25	0.19
Clinotanytarsus	0	0	0
Cloperia	0	0	0
Cnephia	0	0	0
Collembola	0.06	0.05	0
Conchapelopia	0.57	0.65	0.67
Cordulegaster	0.06	0.05	0
Cordulidae	0.05	0.05	0
Corydidae	0.05	0.05	0
Corydatus	0	0	0
Corynoneura	0.32	0.3	0.33
Crangonyx	0.13	0.1	0.19
Cricotopus/Orthocladus	0.57	0.6	0.67
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.05	0.05	0
Diamesa	0.26	0.21	0.32
Dicranota	0.06	0.15	0
Dicrolenidipes	0	0	0
Dineutus	0	0	0
Diphetor	0.06	0.05	0
Diplectrona	0.33	0.3	0.45
Diplocladius	0	0	0
Diploperla	0.07	0.05	0.13
Dixa	0	0.05	0
Dolichopodidae	0.06	0.05	0
Dolophilodes	0.06	0.1	0
Drunella	0.11	0.1	0
Dubiraphia	0.05	0.05	0
Dugesia	0	0.05	0
Ecopotlura	0.12	0.1	0.12
Ectopria	0	0	0
Enallagma	0.05	0.05	0
Enchytraeidae	0.06	0.05	0
Endochironomus	0.05	0.1	0
Epeorus	0.19	0.25	0.2
Ephemera	0	0	0
Ephemerella	0.6	0.65	0.88
Erioptera	0	0	0
Eukiefferiella	0.25	0.3	0.3
Eurylophella	0.21	0.16	0.39
Fallicon	0	0	0
Gammarus	0.13	0.1	0.19
Glossosoma	0.19	0.15	0.23
Gomphidae	0.21	0.16	0.41
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Helemiella	0	0	0
Helictus	0	0.05	0
Helius	0	0	0
Hemerodromia	0.28	0.3	0.49
Heptagenia	0	0	0
Heterotrissocladius	0.12	0.1	0
Hexagenia	0	0	0
Hexatoma	0.14	0.1	0.25
Hyalella	0	0	0
Hydrobaenus	0.12	0.1	0.11
Hydrobius	0.06	0.05	0
Hydroporus	0	0	0
Hydropsyche	0.26	0.3	0.36
Ironoquia	0	0	0
Isonychia	0	0.05	0
Isoperla	0.12	0.1	0
Kiefferulus	0.05	0.05	0
Krenopelopia	0.07	0.05	0.12
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0.06	0.05	0
Lepidostoma	0	0.05	0

Leptophlebia	0.11	0.1	0
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.11	0.19	0
Limnophila	0.06	0.05	0
Limnophyes	0.07	0.05	0.12
Lopescladius	0	0	0
Lumbriculidae	0.22	0.2	0
Lype	0.05	0.05	0
Macronychus	0	0	0
Menetus	0	0	0
Meropelopia	0.13	0.1	0.12
Microasema	0	0	0
Microprosectra	0.46	0.41	0.64
Microtendipes	0.07	0.1	0.11
Microvelia	0	0	0
Molophilus	0	0	0
Musculium	0	0	0
Naididae	0.34	0.35	0.56
Nanocladius	0	0.05	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.57	0.55	0.75
Nigronia	0.24	0.2	0.12
Nixa	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.34	0.4	0.57
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostrocerca	0	0	0
Oulimnius	0.48	0.55	0.85
Pagastia	0.08	0.05	0.23
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaelocladus	0	0	0
Parachironomus	0.05	0.05	0
Paragnetia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.15	0.1	0.35
Paramerina	0.05	0.05	0
Parameitricnemus	0.69	0.65	0.78
Paraphaenocladus	0	0	0
Paratanytarsus	0.13	0.2	0.11
Pedicia	0	0	0
Pelodytes	0.05	0.05	0
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.07	0.05	0.11
Phaenoglyphia	0	0	0
Phycocentropus	0	0	0
Physidae	0.11	0.1	0
Pisidium	0.08	0.05	0.23
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0.05	0
Polycentropus	0.28	0.25	0.55
Polypedium	0.37	0.4	0.36
Pothastia	0.07	0.05	0.13
Probezzia	0	0	0
Procladius	0	0	0
Pronesia	0	0	0
Prosimulium	0.58	0.6	0.64
Prostia	0.32	0.3	0.39
Prostoma	0	0	0
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.13	0.1	0.12
Pseudorthocladus	0.05	0.05	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0	0.05	0
Ptilostomis	0	0	0
Pycnopsyche	0.12	0.1	0.12
Radix	0	0	0
Rheocricotopus	0.06	0.05	0
Rheotanytarsus	0.21	0.15	0.36
Rhyacophila	0.51	0.55	0.64
Serratella	0.14	0.2	0.23
Sialis	0.07	0.05	0.12
Simulium	0.25	0.3	0.23
Siphonurus	0	0	0
Siphonocentrus	0	0	0
Sphaeridium	0.06	0.05	0
Sphaerium	0	0	0
Sphaeromias	0.33	0.26	0.44
Stegopterna	0	0	0
Stempellina	0	0	0
Stempellinella	0.05	0.05	0
Stenacron	0	0	0
Stenelmis	0.08	0.05	0.2
Stenonema	0.39	0.4	0.48
Stictochironomus	0.07	0.05	0.12
Stilocladius	0.05	0.05	0
Strophopteryx	0.19	0.15	0.19
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphitostoma	0.07	0.1	0.13
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytarsus	0	0	0
Tanytarsus	0.31	0.35	0.36
Thienemannella	0.13	0.24	0.13
Thienemannimyia(grp)	0.06	0.05	0
Tipula	0.37	0.35	0.31
Trepobates	0	0	0
Trienodes	0.06	0.05	0
Tribelos	0.06	0.05	0
Trisopelopia	0.36	0.3	0.25
Tubificidae	0		

Biodiversity for LRJB 203B:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.11	0.09	0.05
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.63	0.64	0.65
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.05	0	0
Cambaridae	0	0	0
Capniidae	0	0.18	0.02
Carabidae?	0	0	0
Ceratopogonidae	0.1	0	0
Chironomidae	1	1	1
Chloroperlidae	0.08	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corticulidae	0	0	0
Contullegastriidae	0	0	0
Cordulidae	0	0	0
Corixidae	0	0	0
Corydalidae	0.2	0.09	0.21
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0.04	0	0
Dolichopodidae	0	0	0
Dryopidae	0.13	0	0
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.9	0.82	0.93
Empididae	0.6	0.73	0.76
Enchytraeidae	0	0	0
Ephemereidae	0.95	1	1
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ergobdellidae	0	0	0
Gammaridae	0.18	0.09	0.05
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.35	0.54	0.44
Gomphidae	0.13	0	0
Gordidae	0.06	0.09	0.05
Gyrinidae	0	0	0
Halplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.67	0.64	0.81
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.84	0.91	1
Hydroptilidae	0.12	0.09	0.05
Lepidoptera	0	0	0
Lepidostomatidae	0.04	0	0
Lepidoceridae	0	0	0
Lepidostomatidae	0.14	0.09	0.16
Lestidae	0	0	0
Leuctridae	0.08	0.09	0
Libellulidae	0	0	0
Limnephilidae	0.5	0.45	0.46
Lumbriculidae	0.08	0.18	0.09
Lymnaeidae	0.05	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metreopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.37	0.4	0.4
Nemouridae	0.82	0.72	0.68
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.14	0.09	0.1
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.05	0	0
Perlidae	0.36	0.36	0.33
Perlodidae	0.38	0.18	0.25
Philopotamidae	0.36	0.46	0.55
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.1	0.09	0.16
Planariidae	0.1	0.09	0.05
Planorbidae	0	0	0
Pleuroceridae	0.05	0	0
Polycentropodidae	0.2	0.19	0.35
Potamanthidae?	0	0	0
Psephenidae	0.06	0.09	0.05
Psychodidae	0	0	0
Psychomyiidae	0.06	0.09	0.05
Pteronarcidae	0.09	0	0
Ptilodactylidae?	0.13	0.09	0.16
Ptychopteridae	0	0	0
Rhyacophilidae	0.56	0.37	0.41
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.91	0.79
Siphonuridae	0.32	0.54	0.5
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.05	0.09	0.02
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.79	0.82	0.83
Tubificidae	0.1	0.19	0.29

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Ablabesmyia	0.06	0.06	0
Acentrota	0.11	0.12	0
Acerpenna	0.18	0.18	0.1
Acroneria	0.08	0.06	0.21
Agapetus	0	0	0
Allocapnia	0.23	0.23	0.1
Alotarypus	0	0	0
Ameletus	0.38	0.35	0.38
Amphinemura	0.73	0.76	0.6
Anchyrtarus	0.14	0.18	0.28
Ancyliidae	0	0	0
Ancronyx	0	0	0
Antocha	0.26	0.24	0.38
Apsectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.06	0.06	0
Bezzia	0	0	0
Bitacomorpha	0.05	0.06	0
Blepharicera	0	0	0
Boyeria	0.05	0.06	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brilia	0.05	0.06	0
Caecidotea	0.05	0.06	0
Caenis	0.05	0.06	0
Calibaetis	0.05	0.06	0
Calopteryx	0.05	0.06	0
Cambaridae	0.06	0.06	0
Centropilum	0.06	0.06	0
Ceratopogon	0.21	0.18	0.31
Chaeocladus	0.05	0.06	0
Chelitera	0.07	0.06	0.12
Cheumatopsyche	0.46	0.42	0.72
Chimarra	0.13	0.12	0.17
Chironomus	0	0	0
Chloroperlidae	0.07	0.06	0.12
Chrysops	0	0	0
Chrysmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.18	0.18	0.12
Clinotanytus	0	0	0
Cloperla	0	0	0
Cnephia	0	0	0
Collembola	0.06	0.06	0
Conchapelopia	0.58	0.59	0.71
Cordulegaster	0.06	0.06	0
Corduliidae	0.05	0.06	0
Corixidae	0.05	0.06	0
Corydalus	0	0	0
Corynoneura	0.33	0.3	0.38
Crangonyx	0.12	0.12	0.12
Cricotopus/Orthocladus	0.51	0.53	0.62
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.05	0.06	0
Diamesa	0.26	0.24	0.31
Dicranota	0.06	0.12	0
Dicrotendipes	0	0	0
Dineutus	0	0	0
Dipheter	0.06	0.06	0
Diplectrona	0.39	0.35	0.48
Diplocladius	0	0	0
Diploperla	0.07	0.06	0.17
Dixa	0	0.06	0
Dolichopodidae	0.06	0.06	0
Dolophloides	0.11	0.11	0
Drunella	0.11	0.11	0
Dubiraphia	0.05	0.06	0
Dugesia	0	0.06	0
Ecoptura	0.07	0.06	0.1
Ectopria	0	0	0
Enallagma	0.05	0.06	0
Enchytraeidae	0.06	0.06	0
Endochironomus	0.05	0.06	0
Epeorus	0.24	0.23	0.21
Ephemera	0	0	0
Ephemerella	0.65	0.65	0.9
Erioptera	0	0	0
Eukiefferella	0.25	0.24	0.24
Eurylophella	0.21	0.18	0.33
Fallicon	0	0	0
Gammarus	0.07	0.06	0.12
Glossosoma	0.19	0.18	0.21
Gomphidae	0.21	0.18	0.41
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Helieniella	0	0	0
Helichus	0.05	0.06	0
Helius	0	0	0
Hemerodromia	0.28	0.3	0.57
Heptagenia	0	0	0
Heterotrissocladius	0.12	0.12	0
Hexagenia	0	0	0
Hexatoma	0.14	0.12	0.27
Hyalella	0	0	0
Hydrobaenus	0.12	0.12	0.12
Hydrobius	0.06	0.06	0
Hydroponus	0	0	0
Hydropsyche	0.26	0.24	0.38
Ironoquia	0	0	0
Isopychia	0	0	0
Isoperla	0.12	0.12	0
Kiefferulus	0.05	0.06	0
Krenopelopia	0.07	0.06	0.1
Labrundinia	0	0	0

Larsia	0	0	0
Lepidoptera	0.06	0.06	0
Lepidostoma	0	0.06	0
Leptophlebia	0.12	0.12	0
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocuta	0	0	0
Leuctridae	0.16	0.23	0
Limnophila	0.05	0.06	0
Limnophyes	0.07	0.06	0.1
Lopescladius	0	0	0
Lumbriculidae	0.17	0.17	0
Lype	0	0	0
Macropnychus	0	0	0
Menetus	0	0	0
Meropelopia	0.13	0.12	0.1
Micrasema	0	0	0
Micropectra	0.46	0.47	0.6
Microtendipes	0.07	0.06	0.12
Microvelia	0	0	0
Molophilus	0	0	0
Musculium	0	0	0
Naididae	0.4	0.41	0.59
Nanocladius	0	0.06	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.58	0.58	0.72
Nigrinia	0.18	0.18	0.1
Nixe	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.39	0.41	0.58
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostrocerca	0	0	0
Oulimnius	0.53	0.53	0.76
Pagasia	0.08	0.06	0.21
Palaemonetes	0	0	0
Paracania	0	0	0
Parachaeocladus	0	0	0
Parachironomus	0.05	0.06	0
Paragnetina	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.15	0.12	0.31
Parameira	0	0.05	0
Parametrioctenemus	0.63	0.65	0.72
Paraphaenocladus	0	0	0
Paratanytarsus	0.13	0.12	0.12
Pedicia	0	0	0
Peltopteryx	0.05	0.06	0
Peltoperla	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.07	0.06	0.12
Phasganophora	0	0	0
Phylocentropus	0	0	0
Physidae	0.12	0.12	0
Pisidium	0.08	0.06	0.21
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0.06	0
Polycentropus	0.27	0.3	0.51
Polydellium	0.32	0.35	0.38
Poithastia	0.07	0.06	0.17
Probaetia	0	0	0
Procladius	0	0	0
Promesia	0	0	0
Prosimulium	0.63	0.65	0.62
Prostia	0.32	0.3	0.33
Prostoma	0	0	0
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.13	0.12	0.1
Pseudorthocladus	0.05	0.06	0
Pseudosuccinea	0	0	0
Psilotritia	0	0	0
Pteronarcys	0	0.06	0
Ptilostomis	0	0	0
Pycnopsyche	0.12	0.12	0.1
Radix	0	0	0
Rheocricotopus	0.05	0.06	0
Rheotanytarsus	0.21	0.18	0.38
Rhyacophila	0.56	0.59	0.62
Serratella	0.19	0.23	0.21
Sialis	0.07	0.06	0.1
Simulium	0.3	0.35	0.21
Siphonurus	0	0	0
Siphonopteron	0	0	0
Sphaeriidae	0.06	0.06	0
Sphaerium	0	0	0
Sphaeromias	0	0	0
Stegopterna	0.32	0.3	0.39
Stempellina	0	0	0
Stempellinella	0.05	0.06	0
Stenacron	0	0	0
Stenelmis	0.08	0.06	0.21
Stenonema	0.44	0.41	0.48
Stictochironomus	0.07	0.06	0.1
Stilocladus	0.05	0.06	0
Strophopteryx	0.18	0.18	0.12
Stygobromus	0	0	0
Symptocladus	0	0	0
Symplothastia	0.07	0.06	0.17
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Talitria	0	0	0
Tanytarsus	0	0	0
Tanytarsus	0.26	0.29	0.38
Thienemannella	0.13	0.18	0.17
Thienemannimyia(grp)	0.06	0.06	0
Tipula	0.41	0.41	0.22
Trepobates	0	0	0
Trisenodes	0.05	0.06	0
Tribelos	0.05	0.06	0

Biodiversity for LRJB 204:

Family:	Ragan ANNA BIC	Ragan BROJ StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.11	0.09	0.05
Astacidae	0	0	0
Athericidae	0	0	0
Baeliscidae	0.64	0.64	0.64
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.05	0	0
Cambariae	0	0	0
Capniidae	0	0.18	0.04
Carabidae?	0	0	0
Ceratopogonidae	0.09	0	0
Chironomidae	1	1	1
Chloroperlidae	0.08	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corticulidae	0	0	0
Coruliagastridae	0	0	0
Corulidae	0	0	0
Coryidae	0	0	0
Corydalidae	0.2	0.09	0.18
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0.04	0	0
Dolichopodidae	0	0	0
Dryopidae	0.13	0	0
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.9	0.82	0.92
Empididae	0.61	0.73	0.75
Enchytraeidae	0	0	0
Ephemereidae	0.95	1	1
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ergobdellidae	0	0	0
Gammaridae	0.18	0.09	0.06
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.35	0.54	0.45
Gomphidae	0.13	0	0
Gordidae	0.06	0.09	0.05
Gyrinidae	0	0	0
Haliplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.67	0.64	0.77
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.84	0.91	0.96
Hydroptilidae	0.11	0.09	0.05
Lepidoptera	0	0	0
Lepidostomatidae	0.04	0	0
Lepiceridae	0	0	0
Leptophlebiidae	0.14	0.09	0.17
Lestidae	0.08	0.09	0.02
Leuctridae	0.08	0.09	0.02
Libellulidae	0	0	0
Limnephilidae	0.5	0.46	0.46
Lumbriculidae	0.08	0.18	0.11
Lymnaeidae	0.05	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naxidae	0.36	0.36	0.38
Nemouridae	0.81	0.72	0.67
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.14	0.09	0.1
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.05	0	0
Perlidae	0.36	0.36	0.33
Perlodidae	0.38	0.18	0.23
Philopotamidae	0.36	0.46	0.51
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.1	0.09	0.15
Planariidae	0.1	0.09	0.05
Planorbidae	0	0	0
Pleuroceridae	0.05	0	0
Polycentropodidae	0.21	0.19	0.32
Potamanthidae?	0	0	0
Psephenidae	0.06	0.09	0.05
Psychodidae	0	0	0
Psychomyiidae	0.06	0.09	0.05
Pteronarcidae	0.08	0	0
Ptilodactylidae?	0.13	0.09	0.15
Ptychopteridae	0	0	0
Rhyacophilidae	0.55	0.37	0.4
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.93	0.91	0.82
Siphonuridae	0.32	0.55	0.52
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.05	0.09	0.03
Talitridae	0	0	0
Tetrammatidae	0	0	0
Tipulidae	0.78	0.82	0.83
Tubificidae	0.1	0.19	0.29

Ulenidae	0	0	0
Unionidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0

Biodiversity for LRJB 204:

Genus:	Ragan ANNA AIC	Ragan BROJ AIC	Ragan TROI AIC
Ablabesmyia	0.06	0.06	0
Acentrella	0.11	0.12	0
Acerpenna	0.18	0.18	0.08
Acroneria	0.08	0.06	0.21
Agapetus	0	0	0
Allocapnia	0.23	0.23	0.08
Aloctanytus	0	0	0
Ameletus	0.38	0.35	0.39
Amphinemura	0.73	0.76	0.59
Anchyrtarus	0.15	0.18	0.29
Ancylidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.27	0.24	0.39
Assectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.06	0.06	0
Bezzia	0	0	0
Bitiacomorpha	0.05	0.06	0
Blepharocera	0	0	0
Boyeria	0.05	0.06	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brilia	0.05	0.06	0
Caecidotea	0.05	0.06	0
Caenis	0.05	0.06	0
Callibaetis	0.05	0.06	0
Calopteryx	0.05	0.06	0
Cambariae	0.06	0.06	0
Centropilum	0.06	0.06	0
Ceratopogon	0.21	0.18	0.29
Chaetocladius	0.05	0.06	0
Chelifera	0.07	0.06	0.12
Cheumatopsyche	0.47	0.42	0.73
Chimarra	0.13	0.12	0.18
Chironomus	0	0	0
Chloroperlidae	0.07	0.06	0.12
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.18	0.18	0.13
Clinotanytus	0	0	0
Cloptera	0	0	0
Cnephia	0	0	0
Collembola	0.06	0.06	0
Conchapelopia	0.58	0.59	0.72
Cordulegaster	0.06	0.06	0
Coruliidae	0.05	0.06	0
Corixidae	0.05	0.06	0
Corydalis	0	0	0
Corynoneura	0.33	0.3	0.39
Crangonyx	0.12	0.12	0.13
Cricotopus	0.51	0.53	0.61
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.05	0.06	0
Diamesa	0.26	0.24	0.32
Dicranota	0.06	0.12	0
Dicrolaelidipes	0	0	0
Dineutus	0	0	0
Dipheter	0.06	0.06	0
Diplectrona	0.38	0.35	0.46
Diplocladius	0	0	0
Diploperla	0.07	0.06	0.18
Dixa	0	0.06	0
Dolichopodidae	0.06	0.06	0
Dolophilodes	0.11	0.11	0
Drunella	0.11	0.12	0
Dubiraphia	0.06	0.06	0
Dugesia	0	0.06	0
Eccoptura	0.06	0.06	0.08
Ectopria	0	0	0
Enallagma	0.05	0.06	0
Enchytraeidae	0.06	0.06	0
Endochironomus	0.05	0.06	0
Epeorus	0.24	0.23	0.21
Ephemerella	0	0	0
Ephemerella	0.65	0.65	0.82
Erioptera	0	0	0
Eukiefferiella	0.25	0.24	0.25
Eurylophella	0.21	0.18	0.34
Falceon	0	0	0
Gammarus	0.07	0.06	0.13
Glossosoma	0.19	0.18	0.21
Gomphidae	0.21	0.18	0.42
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleiniella	0	0	0
Helicidius	0.05	0.06	0
Hellus	0	0	0
Hemerodromia	0.29	0.3	0.58
Heptagenia	0	0	0
Heterotrissocladius	0.11	0.12	0
Hexagenia	0	0	0
Hexatoma	0.14	0.12	0.25
Hyleiella	0	0	0
Hydrobaenus	0.12	0.12	0.12
Hydrobius	0.06	0.06	0
Hydroporus	0	0	0
Hydropsyche	0.27	0.24	0.39
Itronquia	0	0	0
Isonychia	0	0	0
Isoperla	0.12	0.12	0
Kiefferulus	0.05	0.06	0
Krenopelopia	0.06	0.06	0.08

Labrundinia	0	0	0
Larisa	0	0	0
Lepidoptera	0.06	0.06	0
Lepidostoma	0	0.06	0
Leptophlebia	0.12	0.12	0
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucocuta	0	0	0
Leuctridae	0.16	0.23	0
Limnophila	0.05	0.06	0
Limnophyes	0.06	0.06	0.08
Lopescladius	0	0	0
Lumbriculidae	0.17	0.17	0
Lype?	0	0	0
Macronychus	0	0	0
Menetus	0	0	0
Meropelopia	0.13	0.12	0.08
Micrasema	0	0	0
Micropectra	0.46	0.47	0.59
Microtendipes	0.07	0.06	0.12
Microvelia	0	0	0
Molophilus	0	0	0
Muscilium	0	0	0
Naididae	0.4	0.41	0.6
Nanocladius	0	0.06	0
Nematomorphia	0	0	0
Nemoura	0	0	0
Neophylax	0.58	0.59	0.73
Nigronia	0.18	0.18	0.08
Nixe	0	0	0
Oecetis	0	0	0
Oernopteryx	0	0	0
Optioservus	0.39	0.41	0.6
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostrocerca	0	0	0
Oulimnius	0.53	0.53	0.75
Pagastia	0.08	0.06	0.21
Palaeonetes	0	0	0
Paracapnia	0	0	0
Parachaetocladius	0	0	0
Parachironomus	0.05	0.06	0
Paragnetina	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.15	0.12	0.29
Paramerina	0.05	0.06	0
Parametricnemus	0.63	0.65	0.71
Paraphaenocladus	0	0	0
Paratanytarsus	0.13	0.12	0.12
Pedicia	0	0	0
Pelodytes	0.05	0.06	0
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.07	0.06	0.12
Phasganophora	0	0	0
Phycocentropus	0	0	0
Physidae	0.12	0.12	0.08
Pisidium	0.08	0.06	0.21
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0.06	0
Polycentropus	0.28	0.3	0.52
Polypedilum	0.32	0.35	0.39
Pothosia	0.07	0.06	0.18
Probezia	0	0	0
Procladius	0	0	0
Promoresia	0	0	0
Prosimulium	0.63	0.65	0.61
Prostia	0.32	0.3	0.34
Prostoma	0	0	0
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.13	0.12	0.08
Pseudorthocladius	0.05	0.06	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0	0.06	0
Ptilostomis	0	0	0
Pycnopsyche	0.12	0.12	0.08
Radix	0	0	0
Rheocricotopus	0.05	0.06	0
Rheotanytarsus	0.21	0.18	0.39
Rhyacophila	0.56	0.59	0.61
Serratella	0.19	0.23	0.21
Sialis	0.06	0.06	0.08
Simulium	0.3	0.35	0.21
Siphonurus	0	0	0
Siphonopteron	0	0	0
Sphaeriidae	0.06	0.06	0
Sphaerium	0	0	0
Sphaeromias	0	0	0
Stegopterna	0.32	0.3	0.38
Stempellina	0	0	0
Stempellinella	0.05	0.06	0
Stenacron	0	0	0
Stenelmis	0.08	0.06	0.21
Stenonema	0.44	0.41	0.47
Stictochironomus	0.06	0.06	0.08
Stitocladius	0.05	0.06	0
Strophopteryx	0.18	0.18	0.13
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphothastia	0.07	0.06	0.18
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Talaperla	0	0	0
Tanytus	0	0	0
Tanytarsus	0.26	0	

Biodiversity for LRSB 101A:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.09	0	0
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.58	0.62	0.59
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.04	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.04	0	0
Cambardae	0	0	0
Capniidae	0.17	0.25	0.3
Carabidae?	0	0	0
Ceratopogonidae	0.04	0	0
Chironomidae	1	1	1
Chironomulidae	0	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corbiculidae	0	0	0
Corulegastridae	0	0	0
Conduidae	0	0	0
Corixidae	0	0	0
Corydidae	0	0	0
Corydalidae	0.14	0.12	0.09
Crangonyctidae	0.04	0	0.06
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostrota	0	0	0
Dixidae	0	0	0
Dolichopodidae	0	0	0
Dryopidae	0.08	0	0
Dugesidae	0	0	0
Dytiscidae	0	0	0
Elmidae	0.87	0.88	0.82
Empididae	0.64	0.75	0.74
Enchytraeidae	0	0	0
Ephemereidae	0.84	1	0.87
Ephemeridae	0	0	0
Ephyrididae	0	0	0
Ephobellidae	0	0	0
Gammaridae	0.22	0	0.06
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.41	0.5	0.45
Gomphidae	0.08	0	0
Gonidae	0.05	0	0
Gyrinidae	0	0	0
Halplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.63	0.62	0.66
Hydrobiidae	0	0	0
Hydrophilidae	0	0	0
Hydropsychidae	0.79	0.88	0.73
Hydroptilidae	0.09	0	0
Lepidoptera	0	0	0
Lepidostomatidae	0	0	0
Leptoceridae	0	0	0
Leptophlebiidae	0.16	0.18	0.18
Lesidae	0	0	0
Leuctridae	0.04	0.12	0.08
Libellulidae	0	0	0
Limnephilidae	0.5	0.63	0.53
Lumbriculidae	0.13	0.25	0.19
Lymnaeidae	0.04	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.31	0.38	0.27
Nemouridae	0.77	0.75	0.71
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.23	0.13	0.24
Ostracoda	0	0	0
Palaeomonidae	0	0	0
Peltoperidae	0.04	0	0
Perliidae	0.36	0.25	0.26
Perlotidae	0.26	0.12	0.07
Phlebotomidae	0.33	0.37	0.37
Phygadeuonidae	0	0	0
Physidae	0	0	0
Piscocidae	0	0	0
Pisidiidae	0.12	0.13	0.12
Planariidae	0.05	0	0
Planorbidae	0	0	0
Pleurocentidae	0.04	0	0
Polycentropodidae	0.21	0.25	0.19
Potamanthidae?	0	0	0
Psephenidae	0.05	0	0
Psychodidae	0	0	0
Psychomyiidae	0.09	0	0
Pteronarcidae	0.04	0	0
Ptilodactylidae?	0.08	0.13	0.12
Ptychopteridae	0	0	0
Rhycophyllidae	0.42	0.5	0.39
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericoisomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.88	0.91
Siphonuridae	0.39	0.75	0.58
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.25	0.12	0.21
Talitridae	0	0	0
Tetrammatidae	0	0	0
Tipulidae	0.81	0.87	0.83
Tubificidae	0.16	0.25	0.25

Ulenidae	0	0	0
Unionidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Genus:			
Ablabesmyia	0.07	0.05	0.12
Acetrotella	0.12	0.1	0.06
Acerpenna	0.19	0.15	0.2
Acroneria	0.08	0.05	0.2
Agapetus	0	0	0
Allocapnia	0.25	0.2	0.2
Alotrypys	0	0	0
Ameletus	0.38	0.35	0.44
Amphinemura	0.8	0.75	0.8
Anchytarsus	0.19	0.15	0.2
Ancyliidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.27	0.25	0.37
Aspetrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.06	0.05	0.06
Bezzia	0	0	0
Bitacomorpha	0	0.05	0
Blepharicera	0	0	0
Boyeria	0	0.05	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brillia	0.06	0.05	0
Caesiodotea	0	0.05	0
Caenis	0	0.05	0
Callibaetis	0	0.05	0
Calopteryx	0	0.05	0
Cambaridae	0.07	0.05	0.12
Centroptilum	0.07	0.05	0.12
Ceratopogon	0.21	0.16	0.42
Chaetocladius	0.05	0.05	0
Chelifera	0.11	0.1	0
Cheumatopsyche	0.47	0.4	0.68
Chimarra	0.13	0.1	0.17
Chironomus	0	0	0
Chironomulidae	0.11	0.15	0
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.19	0.2	0.16
Clinotanytus	0	0	0
Cloperia	0	0	0
Cnephia	0	0	0
Collembole	0.07	0.05	0.12
Conchapelopia	0.51	0.55	0.58
Cordulegaster	0.07	0.05	0.12
Cordulidae	0	0.05	0
Corixidae	0	0.05	0
Corydalis	0	0	0
Corynoneura	0.28	0.26	0.5
Crangonyx	0.07	0.1	0.16
Cricotopus/Orthocladus	0.56	0.55	0.5
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0	0.05	0
Diamesa	0.25	0.2	0.22
Dicranota	0.17	0.2	0.06
Dicrolentipes	0	0	0
Dineutus	0	0	0
Diphetor	0.05	0.05	0
Diplectrona	0.39	0.4	0.53
Diplocladius	0	0	0
Diploperia	0.07	0.1	0.17
Dixa	0.05	0.05	0
Dolichopodidae	0.07	0.05	0.12
Dolophibodes	0.16	0.15	0
Drunella	0.11	0.1	0
Dubiraphia	0.06	0.05	0
Dugesia	0.05	0.05	0
Ecocoptura	0.07	0.05	0.14
Eclopria	0	0	0
Enallagma	0	0.05	0
Enchytraeidae	0.07	0.05	0.12
Endochironomus	0	0.05	0
Epeorus	0.29	0.3	0.16
Ephemerella	0.74	0.7	0.68
Erioptera	0	0	0
Eukiefferiella	0.25	0.3	0.22
Eurylophella	0.21	0.16	0.43
Falceon	0	0	0
Gammarus	0.07	0.05	0.16
Glossosoma	0.2	0.15	0.2
Gomphidae	0.2	0.15	0.31
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleliella	0	0	0
Helichus	0.05	0.05	0
Hellus	0	0	0
Hemerodromia	0.33	0.3	0.37
Heptagenia	0	0	0
Heterotrissocladus	0.07	0.1	0.12
Hexagenia	0	0	0
Hexatoma	0.2	0.15	0.31
Hyalella	0	0	0
Hydrobaenus	0.06	0.1	0
Hydrobius	0.07	0.1	0.12
Hydroporus	0	0	0
Hydropsyche	0.27	0.25	0.37
Isonychia	0	0.05	0
Isonychia	0	0.05	0
Isoperla	0.18	0.15	0.18
Kiefferulus	0	0.05	0
Krenopelopia	0.07	0.05	0.14
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0.07	0.05	0.12
Lepidostoma	0.05	0.05	0

Leptophlebia	0.07	0.1	0.12
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0.05	0.05	0
Leuctridae	0.21	0.24	0
Limnophila	0	0.05	0
Limnophyes	0.07	0.05	0.14
Lopesocladus	0	0	0
Lumbriculidae	0.18	0.15	0.12
Lype	0	0	0
Macronychus	0	0	0
Menetus	0	0	0
Meropelopia	0.19	0.15	0.26
Micrasema	0	0	0
Micropsectra	0.58	0.46	0.8
Microtendipes	0.06	0.05	0
Microvelia	0	0	0
Molophilus	0	0	0
Muscullum	0	0	0
Naididae	0.45	0.36	0.58
Nanocladus	0.05	0.05	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.69	0.6	0.74
Nigronia	0.24	0.25	0.26
Nixa	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.45	0.4	0.52
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostracerca	0	0	0
Oulimnius	0.63	0.6	0.71
Pagasia	0.08	0.05	0.2
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaeocladus	0.05	0.05	0
Parachironomus	0	0.05	0
Paragnetia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.21	0.2	0.34
Paramerina	0	0.05	0
Parameitricnemus	0.7	0.7	0.73
Paraphaenocladus	0	0	0
Paratanytarsus	0.12	0.15	0.06
Pedicia	0	0	0
Pelodytes	0	0.05	0
Peloperla	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.06	0.05	0
Phaeganophora	0	0	0
Phycocentropus	0	0	0
Physidae	0.07	0.15	0.12
Pisidium	0.08	0.05	0.2
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0.05	0.05	0
Polycentropus	0.28	0.25	0.52
Polypedium	0.39	0.3	0.43
Pothastia	0.07	0.05	0.17
Probezzia	0	0	0
Procladius	0	0	0
Promesesa	0	0	0
Prosimulium	0.73	0.7	0.63
Prostoia	0.32	0.3	0.37
Prostoma	0	0	0
Psephenus	0.05	0.05	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.14	0.1	0.26
Pseudorthocladus	0.06	0.05	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0.1	0.1	0
Ptilostomis	0	0	0
Pycnopsyche	0.07	0.1	0.14
Radix	0	0	0
Rheocricotopus	0	0.05	0
Rheotanytarsus	0.21	0.15	0.37
Rhyacophila	0.65	0.6	0.46
Serratella	0.24	0.25	0.2
Sialis	0.07	0.05	0.14
Simulium	0.3	0.3	0.2
Siphonurus	0	0	0
Siphonocentrus	0	0	0
Sphaeridium	0.07	0.05	0.12
Sphaerium	0	0	0
Sphaeromias	0	0	0
Siegesophlebia	0.28	0.26	0.59
Stempellina	0.05	0.05	0
Stempellinella	0.06	0.05	0
Stenacron	0	0	0
Stenelmis	0.07	0.1	0.16
Stenonema	0.51	0.45	0.63
Stictochironomus	0.07	0.05	0.14
Stilocladus	0.06	0.05	0
Strophopteryx	0.19	0.15	0.22
Stygobromus	0	0	0
Symposiocladus	0	0	0
Symphostoma	0.13	0.1	0.17
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytus	0	0	0
Tanytarsus	0.27	0.35	0.37
Thienemannella	0.19	0.2	0.23
Thienemannimyia(grp)	0.06	0.05	0.06
Tipula	0.38	0.4	0.4

Biodiversity for LRSB 101C:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0.04	0	0
Ameletidae	0	0	0
Ancyliidae	0	0	0
Asellidae	0.08	0	0
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.58	0.62	0.57
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.04	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0	0	0
Cambardae	0	0	0
Capniidae	0.28	0.5	0.53
Carabidae?	0	0	0
Ceratopogonidae	0.04	0	0
Chironomidae	1	1	1
Chloroperlidae	0.04	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corbiculidae	0	0	0
Contulegastridae	0	0	0
Corduliidae	0	0	0
Corixidae	0	0	0
Corydalidae	0.24	0	0
Crangonyctidae	0.05	0.12	0.25
Culiidae	0	0	0
Culicoidae	0	0	0
Diplostrota	0	0	0
Dividae	0	0	0
Dolichopodidae	0	0	0
Dryopidae	0.04	0	0
Dugesidae	0	0	0
Ectidae	0	0	0
Ectisidae	0.87	0.88	0.98
Empididae	0.58	0.88	0.98
Enchytraeidae	0	0	0
Ephemereidae	0.7	0.75	0.51
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ephobellidae	0	0	0
Gammaridae	0.25	0	0
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.4	0.5	0.33
Gomphidae	0.08	0	0
Gonidae	0.04	0	0
Gyrinidae	0	0	0
Halilidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.64	0.63	0.8
Hydrobiidae	0	0	0
Hydrophilidae	0	0	0
Hydropsychidae	0.8	0.63	0.49
Hydroptilidae	0.04	0	0
Lepidoptera	0	0	0
Lepidostomatidae	0.04	0	0
Leptophlebiidae	0.19	0.13	0.09
Lestidae	0	0	0
Leuctridae	0.13	0.12	0.02
Libellulidae	0	0	0
Limnephilidae	0.49	0.62	0.44
Lumbriculidae	0.09	0.25	0.08
Lymnaeidae	0.04	0	0
Macromitidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.3	0.25	0.1
Nemouridae	0.79	0.75	0.75
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.23	0.38	0.66
Ostracoda	0	0	0
Paiaemonidae	0	0	0
Peltoperlidae	0	0	0
Perlidae	0.33	0.25	0.24
Periodidae	0.17	0	0
Philopotamidae	0.36	0.25	0.31
Phygadeuonidae	0	0	0
Physidae	0	0	0
Piscicolidae	0	0	0
Pisidiidae	0.09	0.13	0.09
Planariidae	0.04	0	0
Planorbidae	0	0	0
Pleurocentidae	0	0	0
Polycentropodidae	0.18	0.13	0.09
Potamanthidae?	0	0	0
Psephenidae	0.09	0	0
Psychodidae	0	0	0
Psychomyiidae	0.11	0	0
Pteronarcidae	0.04	0	0
Ptilodactylidae?	0.05	0.13	0.09
Ptychopterygidae	0.04	0	0
Rhycophyllidae	0.43	0.37	0.23
Saldidae	0	0	0
Sciomyzidae	0	0	0
Senecostomatidae	0	0	0
Sialidae	0.04	0	0
Simuliidae	0.95	1	1
Siphonuridae	0.3	0.62	0.43
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.26	0.37	0.51
Talitridae	0	0	0
Tetrasemmatidae	0	0	0
Tipulidae	0.77	0.75	0.6
Tubificidae	0.15	0.13	0.17

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Ulenidae	0	0	0
Unionidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Ablabesmyia	0.07	0.04	0.07
Acetrotella	0.06	0.12	0
Acerpenna	0.19	0.17	0.23
Acroneria	0.08	0.04	0.27
Agapetus	0	0.04	0
Allocapnia	0.25	0.17	0.23
Alotanytus	0	0	0
Ameletus	0.42	0.38	0.21
Amphinemura	0.75	0.79	0.73
Anchytarsus	0.14	0.13	0.27
Ancyliidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.26	0.21	0.37
Apsectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.06	0.04	0
Bezzia	0	0	0
Bitacomorpha	0.05	0.04	0
Blepharicera	0	0	0
Boyeria	0.05	0.12	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brillia	0.06	0.08	0
Caesiodotea	0.04	0	0
Caenis	0	0.04	0
Callibaetis	0	0.04	0
Calopteryx	0	0.04	0
Cambaridae	0.07	0.04	0.07
Centroptilum	0.07	0.04	0.07
Ceratopogon	0.21	0.13	0.35
Chaetocladius	0	0.04	0
Chelifera	0.11	0.08	0
Cheumatopsyche	0.42	0.38	0.71
Chimarra	0.12	0.09	0.09
Chironomus	0	0	0
Chloroperlidae	0.11	0.16	0
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.25	0.21	0.28
Clinotanytus	0	0	0
Cloperia	0	0	0
Cnephia	0	0	0
Collembola	0.07	0.08	0.07
Conchapelopia	0.49	0.54	0.42
Cordulegaster	0.07	0.04	0.07
Corduliidae	0.04	0	0
Corixidae	0	0.04	0
Corydatus	0	0	0
Corynoneura	0.31	0.25	0.21
Crangonyx	0.14	0.09	0.28
CricobolusOrthocladus	0.58	0.58	0.79
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0	0.04	0
Diamesa	0.26	0.21	0.28
Dicranota	0.16	0.16	0
Dicrolenidipes	0	0	0
Dineutus	0	0	0
Diphetor	0.05	0.04	0
Diplectrona	0.37	0.34	0.38
Diplocladius	0	0	0
Diploperla	0.07	0.08	0.09
Dixa	0	0.04	0
Dolichopodidae	0.07	0.04	0.07
Dolophibodes	0.11	0.12	0
Drunella	0.11	0.08	0
Dubiraphia	0.06	0.04	0
Dugesia	0	0.04	0
Ecoptura	0.06	0.08	0.23
Ectopria	0	0	0
Enallagma	0	0.04	0
Enchytraeidae	0.07	0.04	0.07
Endochironomus	0	0.08	0
Epeorus	0.28	0.25	0.06
Ephemera	0	0	0
Ephemerella	0.69	0.67	0.71
Erioptera	0	0	0
Eukiefferiella	0.31	0.33	0.28
Eurylophella	0.22	0.13	0.41
Fallicon	0	0	0
Gammarus	0.09	0.09	0.28
Glossosoma	0.2	0.13	0.27
Gomphidae	0.21	0.13	0.34
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleliella	0	0	0
Helichus	0	0.04	0
Hellus	0	0	0
Hemerodromia	0.31	0.25	0.37
Heptagenia	0	0	0
Heterotrissocladius	0.12	0.08	0.07
Hexagenia	0	0	0
Hexatoma	0.19	0.13	0.32
Hyalella	0	0	0
Hydrobaenus	0.06	0.08	0
Hydrobius	0.07	0.08	0.07
Hydroporus	0	0	0
Hydropsyche	0.31	0.29	0.37
Isonychia	0	0.04	0
Isonychia	0.05	0.04	0
Isoperla	0.18	0.13	0.07
Kiefferulus	0	0.04	0
Krenopelopia	0.08	0.04	0.23
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0.07	0.04	0.07
Lepidostoma	0	0.04	0

Leptophlebia	0.07	0.08	0.07
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0.05	0.04	0
Leuctridae	0.16	0.24	0
Limnophila	0.05	0.04	0
Limnophyes	0.08	0.04	0.23
Lopesocladus	0	0	0
Lumbriculidae	0.18	0.21	0.07
Lype	0	0.04	0
Macronychus	0	0	0
Menetus	0	0	0
Meropelopia	0.19	0.13	0.29
Microasema	0	0	0
Microsectra	0.53	0.43	0.73
Microtendipes	0.06	0.08	0
Microvelia	0	0	0
Molophilus	0	0	0
Muscullum	0	0	0
Naididae	0.33	0.3	0.42
Nanocladus	0	0.04	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.59	0.55	0.71
Nigronia	0.25	0.25	0.29
Nixa	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.35	0.38	0.65
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostrocerca	0	0	0
Oulimnius	0.59	0.58	0.84
Pagastia	0.08	0.04	0.27
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaeocladus	0.05	0.04	0
Parachironomus	0	0.04	0
Paragnetia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.21	0.17	0.5
Paramerina	0	0.04	0
Parametrichonemus	0.76	0.67	0.88
Paraphaenocladus	0	0	0
Paratanytarsus	0.17	0.17	0
Pedicia	0	0	0
Pelodytes	0	0.04	0
Peloperla	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.06	0.04	0
Phaenoglyphora	0	0	0
Phycocentropus	0	0	0
Physidae	0.07	0.12	0.07
Pisidium	0.08	0.04	0.27
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0.04	0
Polycentropus	0.29	0.21	0.65
Polypedium	0.32	0.34	0.37
Pothastia	0.07	0.04	0.09
Probezzia	0	0	0
Procladius	0	0	0
Promesia	0	0	0
Prosimulium	0.69	0.67	0.63
Prostoma	0.33	0.29	0.34
Pseudocricotopus	0	0	0
Psephenus	0.05	0.08	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.14	0.09	0.29
Pseudorthocladus	0.06	0.04	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0.05	0.08	0
Ptilostomis	0	0	0
Ptycopsyche	0.13	0.09	0.23
Radix	0	0	0
Rheocricotopus	0.05	0.08	0
Rheotanytarsus	0.21	0.17	0.37
Rhyacophila	0.56	0.54	0.57
Serratella	0.14	0.21	0.27
Sialis	0.08	0.04	0.23
Simulium	0.25	0.33	0.27
Siphonurus	0	0	0
Siphonoplecton	0	0	0
Sphaeridium	0.07	0.04	0.07
Sphaerium	0	0	0
Sphaeromias	0	0	0
Synurella	0.35	0.26	0.67
Stempellina	0.05	0.04	0
Stempellinella	0.06	0.08	0
Stenacron	0	0	0
Stenelmis	0.07	0.08	0.06
Stenonema	0.51	0.38	0.66
Stictochironomus	0.08	0.04	0.23
Stilocladus	0.06	0.04	0
Strophopteryx	0.21	0.13	0.28
Stygobromus	0	0	0
Symposiocladus	0	0	0
Symphotha	0.12	0.12	0.09
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytus	0	0	0
Tanytarsus	0.31	0.33	0.37
Thienemannella	0.18	0.25	0.09
Thienemannimyia(grp)	0.06	0.04	0
Tipula			

Biodiversity for LRTB 203A:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TRCI AIC	Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TRCI AIC				
Aeshnidae	0	0	0	Ablesmyia	0.06	0.04	0	Kiefferulus	0.06	0.04	0
Ameletidae	0	0	0	Acentrella	0.05	0.12	0	Krenopelopia	0.06	0.04	0
Ancylidae	0	0	0	Acerpenna	0.17	0.2	0	Labrundinia	0	0	0
Asellidae	0.1	0	0	Acronaeria	0.08	0.08	0.32	Larsia	0	0	0
Astacidae	0	0	0	Agapetus	0	0.04	0	Lepidoptera	0.06	0.04	0
Athericidae	0	0	0	Allocaenia	0.23	0.17	0	Lepidostoma	0	0.04	0
Baelidae	0.61	0.67	0.63	Alotanypus	0	0.04	0	Leptophlebia	0.11	0.08	0
Baetiscidae	0	0	0	Ameletus	0.35	0.33	0.17	Leptoxis	0	0	0
Belostomatidae	0	0	0	Amphinemura	0.73	0.71	0.48	Leucotrichia	0	0	0
Bithyniidae	0	0	0	Anchytarsus	0.15	0.13	0.32	Leucrocota	0	0	0
Blephariceridae	0	0	0	Ancylidae	0	0	0	Leuctridae	0.12	0.2	0
Brachycentridae	0.05	0	0	Ancynonyx	0	0	0	Limnophila	0.06	0.04	0
Braconidae	0	0	0	Antocha	0.26	0.21	0.32	Limnophyes	0.06	0.04	0
Branchiobdellidae	0	0	0	Apsectrotanypus	0	0	0	Lopescladius	0.22	0.24	0
Caeniidae	0	0	0	Argia	0	0	0	Lype	0.06	0.04	0
Calamoceratidae	0	0	0	Atherix	0	0	0	Macronychus	0	0.04	0
Calopterygidae	0.05	0	0	Baetis	0.05	0.04	0	Menetus	0	0.04	0
Cambardae	0	0	0	Bezzia	0.06	0.04	0	Meropelopia	0.12	0.09	0
Capniidae	0.12	0.22	0.11	Bittacomorpha	0.06	0.04	0	Microasema	0	0	0
Carabidae?	0	0	0	Blepharicera	0	0	0	Micropsocetra	0.45	0.38	0.48
Ceratopogonidae	0.05	0	0	Boyeria	0.12	0.12	0	Microtendipes	0.07	0.08	0.2
Chironomidae	1	1	1	Brachycentrus	0	0.04	0	Microvelia	0	0	0
Chiroperlidae	0.04	0	0	Branchiobdellidae	0	0	0	Molophilus	0	0	0
Coenagrionidae	0	0	0	Brillia	0.05	0.12	0	Muscium	0	0.04	0
Collembola	0	0	0	Caecidotea	0.06	0.12	0	Naicidae	0.32	0.33	0.49
Corbiculidae	0	0	0	Caenis	0.06	0.08	0	Nanocladus	0	0.04	0
Cordulegastriidae	0	0	0	Callibaetis	0.06	0.04	0	Nematomorpha	0	0	0
Cordulidae	0	0	0	Calopteryx	0.06	0.04	0	Nemoura	0	0	0
Corixidae	0	0	0	Cambardae	0.06	0.04	0	Neophylax	0.57	0.54	0.8
Corydalidae	0.19	0.11	0.12	Centropilum	0.06	0.04	0	Nigronia	0.24	0.25	0
Crangonyctidae	0	0	0	Ceratopogon	0.19	0.13	0.17	Nixa	0	0	0
Culicidae	0	0	0	Chaetocladius	0	0.04	0	Oecetis	0	0	0
Curculionidae	0	0	0	Chelifera	0.07	0.04	0.2	Oemopteryx	0	0	0
Diplostroca	0	0	0	Cheumatopsyche	0.41	0.42	0.8	Optioservus	0.34	0.38	0.62
Dixidae	0	0	0	Chimarra	0.12	0.08	0	Ormosia	0	0	0
Dolichopodidae	0	0	0	Chironomus	0.04	0.04	0	Orthoclaadiinae(A)	0	0	0
Dryopidae	0.09	0	0	Chloroperlidae	0.07	0.12	0.2	Ostrocerca	0	0	0
Dugesidae	0	0	0	Chrysops	0	0	0	Oulimnius	0.48	0.54	1
Dytiscidae	0	0	0	Cinygmula	0	0	0	Pagastia	0.08	0.04	0.32
Elmidae	0.86	0.78	0.9	Cladopelma	0	0	0	Palaeomonetes	0	0	0
Empididae	0.63	0.78	0.76	Cladotanytarsus	0	0	0	Paracapia	0	0	0
Empytridae	0	0	0	Clinocera	0.19	0.21	0.3	Parachaetocladius	0	0	0
Ephemerellidae	0.87	1	0.97	Clinotanypus	0	0	0	Parachironomus	0.06	0.04	0
Ephemeridae	0	0	0	Cloperla	0	0	0	Paragnetia	0	0	0
Ephydriidae	0	0	0	Cnephia	0	0	0	Parakiefferiella	0	0	0
Erebodellidae	0	0	0	Collembola	0.06	0.04	0	Paraleptophlebia	0.15	0.13	0.32
Gammaridae	0.2	0	0.07	Conchapelopia	0.57	0.62	0.7	Parameria	0.06	0.04	0
Gerridae	0	0	0	Cordulegaster	0.06	0.04	0	Parametricnemus	0.69	0.63	0.83
Glossiphoniidae	0	0	0	Cordulidae	0.06	0.04	0	Paraphaenocladus	0	0.04	0
Glossosomatidae	0.44	0.45	0.45	Corixidae	0.06	0.04	0	Paratanytarsus	0.13	0.21	0.2
Gomphidae	0.05	0	0	Corydalus	0	0	0	Pedicia	0	0.04	0
Goniatidae	0.06	0	0	Corynoneura	0.3	0.25	0.17	Peltodytes	0.06	0.08	0
Gyrinidae	0	0	0	Crangonyx	0.14	0.09	0.3	Peltopera	0	0	0
Halplidae	0	0	0	Cricobolus	0.59	0.66	0.63	Perlesta	0	0	0
Helodidae	0	0	0	Cryptochironomus	0	0	0	Phaenopsectra	0.07	0.04	0.2
Heptageniidae	0.63	0.56	0.63	Culicoides	0	0	0	Phaenopsectra	0	0	0
Hydrobiidae	0	0	0	Cultus	0	0	0	Phagangophora	0	0	0
Hydrophiliidae	0	0	0	Cura	0	0	0	Phycotropus	0	0	0
Hydropsychidae	0.79	0.89	0.86	Deronectes	0.06	0.04	0	Physidae	0.11	0.12	0
Hydroptilidae	0.1	0	0	Diamasa	0.27	0.17	0.51	Pisidium	0.08	0.08	0.32
Lepidoptera	0	0	0	Dicranota	0.05	0.12	0	Planorbella	0	0	0
Lepidostomatidae	0	0	0	Dicrotendipes	0	0.04	0	Platycotropus	0	0	0
Lepidostoma	0	0	0	Dineutus	0	0	0	Platysmittia	0	0.04	0
Lepidostoma	0	0	0	Dipheter	0.06	0.04	0	Polycentropus	0.28	0.21	0.62
Lepidostoma	0	0	0	Dipletrona	0.31	0.29	0.17	Polypedium	0.36	0.33	0.32
Lepidostoma	0	0	0	Diplocladius	0	0.04	0	Pothastia	0.06	0.04	0
Lepidostoma	0	0	0	Diploperla	0.06	0.08	0	Probezzia	0	0	0
Lepidostoma	0	0	0	Dixa	0	0.04	0	Procladius	0	0.04	0
Lepidostoma	0	0	0	Dolichopodidae	0.06	0.04	0	Promotesia	0	0	0
Lepidostoma	0	0	0	Dolophlodes	0.06	0.08	0	Prosimulium	0.57	0.58	0.68
Lepidostoma	0	0	0	Drunella	0.11	0.08	0	Prostoma	0.31	0.33	0.48
Lepidostoma	0	0	0	Dubirapha	0.05	0.04	0	Prostoma	0	0	0
Lepidostoma	0	0	0	Dugesia	0.05	0.04	0	Psephenus	0	0.04	0
Lepidostoma	0	0	0	Eccoptura	0.12	0.09	0	Pseudochironomus	0	0	0
Lepidostoma	0	0	0	Ectopria	0	0	0	Pseudolimnophila	0.12	0.09	0
Lepidostoma	0	0	0	Enallagma	0.06	0.04	0	Pseudorthocladus	0.05	0.04	0
Lepidostoma	0	0	0	Enchytraeidae	0.06	0.08	0	Pseudosuccinea	0	0	0
Lepidostoma	0	0	0	Endochironomus	0.06	0.08	0	Psilotreta	0	0	0
Lepidostoma	0	0	0	Epeorus	0.18	0.21	0.17	Pteronarcys	0	0.04	0
Lepidostoma	0	0	0	Ephemer	0	0	0	Ptilostomis	0	0	0
Lepidostoma	0	0	0	Ephemerella	0.6	0.63	1	Ptycopsycha	0.13	0.09	0
Lepidostoma	0	0	0	Erioptera	0	0	0	Radix	0	0.04	0
Lepidostoma	0	0	0	Eukiefferiella	0.26	0.33	0.51	Rheocricotopus	0.06	0.12	0
Lepidostoma	0	0	0	Eurytophella	0.2	0.13	0.48	Rheotanytarsus	0.2	0.21	0.32
Lepidostoma	0	0	0	Fallicon	0	0	0	Rhyacophila	0.52	0.5	0.68
Lepidostoma	0	0	0	Gammarus	0.14	0.13	0.3	Serratella	0.14	0.21	0.32
Lepidostoma	0	0	0	Glossosoma	0.19	0.13	0.32	Sialis	0.06	0.04	0
Lepidostoma	0	0	0	Gomphidae	0.21	0.17	0.48	Simulium	0.25	0.33	0.32
Lepidostoma	0	0	0	Goniobasis	0	0	0	Siphonurus	0	0	0
Lepidostoma	0	0	0	Gyrinus	0	0	0	Siphonurinae	0	0	0
Lepidostoma	0	0	0	Habrophiella	0	0	0	Siphonurinae	0	0	0
Lepidostoma	0	0	0	Heleniella	0	0	0	Siphonurinae	0	0	0
Lepidostoma	0	0	0	Helichus	0	0.08	0	Sphaerididae	0.06	0.04	0
Lepidostoma	0	0	0	Helius	0	0	0	Sphaerium	0	0.04	0
Lepidostoma	0	0	0	Hemerodromia	0.28	0.26	0.52	Sphaeromias	0	0	0
Lepidostoma	0	0	0	Heptagenia	0	0	0	Siegeoptera	0.32	0.25	0.3
Lepidostoma	0	0	0	Heterotrissocladius	0.12	0.08	0	Stempellina	0	0	0
Lepidostoma	0	0	0	Hexagenia	0	0	0	Stempellinella	0.05	0.08	0
Lepidostoma	0	0	0	Hexatoma	0.13	0.13	0	Stenacron	0	0	0
Lepidostoma	0	0	0	Hyalella	0	0	0	Stenelmis	0.07	0.12	0.17
Lepidostoma	0	0	0	Hydrobaenus	0.13	0.09	0.2	Stenonema	0.38	0.38	0.32
Lepidostoma	0	0	0	Hydrobius	0.06	0.08	0	Stichochironomus	0.06	0.04	0
Lepidostoma	0	0	0	Hydroponus	0	0	0	Stilocladus	0.05	0.04	0
Lepidostoma	0	0	0	Hydropsyche	0.26	0.33	0.32	Strophopteryx	0.19	0.13	0.3
Lepidostoma	0	0	0	Ironoquia	0	0.04	0	Stygobromus	0	0	0
Lepidostoma	0	0	0	Isonychia	0	0.04	0	Symposiocladius	0	0	0
Lepidostoma	0	0	0	Isoperla	0.11	0.12	0	Symphotastia	0.06	0.08	0
Lepidostoma	0	0	0					Synurella	0	0	0
Lepidostoma	0	0	0					Tabanus	0	0	0
Lepidostoma	0	0	0					Taeniopteryx	0	0	0
Lepidostoma	0	0	0					Tallaperla	0	0	0
Lepidostoma	0	0	0					Tanypus	0	0.04	0
Lepidostoma	0	0	0					Tanytarsus	0.32	0.33	0.32
Lepidostoma	0	0	0					Thienemannella	0.12	0.21	0
Lepidostoma	0	0	0					Thienemannimyia(gp)	0.05	0.04	0
Lepidostoma	0	0	0					Tipula	0.37	0.33	0.3
Lepidostoma	0	0	0					Trepobates	0	0	0
Lepidostoma	0	0	0					Trianaodes	0.06	0.04	0
Lepidostoma	0	0	0					Tribelos	0.06	0.04	0
Lepidostoma	0	0	0					Trisopelopia	0.35	0.25	0
Lepidostoma	0	0	0					Tubificidae	0.18	0.21	0
Lepidostoma	0	0	0					Tvetenia	0.06	0.04	0
Lepidostoma	0	0	0					Unniella	0	0	0
Lepidostoma	0	0	0					Wormaldia	0.06	0.04	0
Lepidostoma	0	0	0					Xyloptopus	0	0	0
Lepidostoma	0	0	0					Zavrelimyia	0.19	0.17	0.17

Biodiversity for LRTB 203B:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.1	0	0.01
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.61	0.67	0.62
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.05	0	0
Cambaridae	0	0	0
Capniidae	0.12	0.22	0.09
Carabidae?	0	0	0
Ceratopogonidae	0.05	0	0
Chironomidae	1	1	1
Chloroperlidae	0.04	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corticulidae	0	0	0
Cordulegastriidae	0	0	0
Cordulidae	0	0	0
Corydalidae	0.19	0.11	0.12
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0	0	0
Dolichopodidae	0	0	0
Dryopidae	0.09	0	0
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.86	0.78	0.9
Empididae	0.63	0.78	0.77
Enchytraeidae	0	0	0
Ephemereidae	0.87	1	0.97
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ergasilidae	0	0	0
Gammaridae	0.2	0	0.09
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.44	0.45	0.44
Gomphidae	0.05	0	0
Gordidae	0.06	0	0.01
Gyrinidae	0	0	0
Haliplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.63	0.56	0.63
Hydrobiidae	0	0	0
Hydrophilidae	0	0	0
Hydropsychidae	0.79	0.89	0.86
Hydroptilidae	0.1	0	0.01
Lepidoptera	0	0	0
Lepidostomatidae	0	0	0
Leploceridae	0	0	0
Leptophlebiidae	0.15	0.11	0.23
Lestidae	0	0	0
Leuctridae	0.04	0.11	0.03
Libellulidae	0	0	0
Limnephilidae	0.42	0.55	0.38
Lumbriculidae	0.13	0.22	0.11
Lymnaeidae	0.05	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.31	0.33	0.27
Nemouridae	0.75	0.66	0.55
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.22	0.11	0.18
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.04	0	0
Perlidae	0.36	0.23	0.32
Periodidae	0.28	0.22	0.15
Philopotamidae	0.34	0.33	0.33
Phryganeidae	0	0	0
Physidae	0	0	0
Piscicolidae	0	0	0
Pisidiidae	0.11	0.11	0.15
Planariidae	0.06	0	0.01
Planorbidae	0	0	0
Pleuroceridae	0.05	0	0
Polycentropodidae	0.16	0.23	0.23
Potamanthidae?	0	0	0
Psephenidae	0.06	0	0.01
Psychodidae	0	0	0
Psychomyiidae	0.1	0	0.01
Pteronarcidae	0.04	0	0
Ptilodactylidae?	0.06	0.11	0.15
Ptychopteridae	0	0	0
Rhyacophilidae	0.42	0.44	0.39
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.89	0.88
Siphonuridae	0.38	0.67	0.53
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.17	0.11	0.07
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.81	0.78	0.77
Tubificidae	0.15	0.23	0.3

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Ulenidae	0	0	0
Unionidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Ablabesmyia	0.06	0.04	0
Acetrotella	0	0.13	0
Acerpenna	0.12	0.17	0
Acroneria	0.08	0.09	0.29
Agapetus	0	0.04	0
Allocapnia	0.17	0.17	0
Alotrypys	0	0.04	0
Ametetus	0.4	0.35	0.18
Amphinemura	0.67	0.69	0.44
Anchytarsus	0.15	0.13	0.34
Ancyidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.26	0.22	0.29
Aspsectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0	0.04	0
Bezzia	0	0	0
Bitacomorpha	0.06	0.04	0
Blepharicera	0	0	0
Boyeria	0.12	0.09	0
Brachycentrus	0	0.04	0
Branchiobdellidae	0	0	0
Brillia	0.05	0.13	0
Caesiodotea	0.06	0.13	0
Caenis	0.06	0.09	0
Callibaetis	0.06	0.04	0
Calopteryx	0.06	0.04	0
Cambaridae	0.06	0.04	0
Centroptilum	0.06	0.04	0
Ceratopogon	0.19	0.13	0.18
Chaetocladius	0	0.04	0
Chellera	0.07	0.05	0.22
Cheumatopsyche	0.41	0.44	0.74
Chimarra	0.12	0.09	0
Chironomus	0	0.04	0
Chloroperlidae	0.07	0.09	0.22
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.24	0.22	0.26
Clinotanytarsus	0	0	0
Cloperia	0	0	0
Cnephia	0	0	0
Collembola	0.06	0.04	0
Conchapelopia	0.57	0.65	0.69
Cordulegaster	0.06	0.04	0
Cordulidae	0.06	0.04	0
Corydalis	0.06	0.04	0
Corydatus	0	0	0
Corynoneura	0.25	0.26	0.18
Crangonyx	0.14	0.09	0.26
Cricotopus/Orthocladius	0.64	0.65	0.82
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.06	0.04	0
Diamesa	0.22	0.18	0.53
Dicranota	0.05	0.13	0
Dicrolenidipes	0	0.04	0
Dineutus	0	0	0
Dipheter	0.06	0.04	0
Diplectrona	0.26	0.3	0.18
Diplocladius	0	0.04	0
Diploperla	0.06	0.09	0
Dixa	0	0.04	0
Dolichopodidae	0.06	0.04	0
Dolophilodes	0.06	0.09	0
Drunella	0.11	0.09	0
Dubiraphia	0.05	0.04	0
Dugesia	0	0.04	0
Ecocoptura	0.12	0.09	0
Ecotopia	0	0	0
Enallagma	0.06	0.04	0
Enchytraeidae	0.06	0.08	0
Endochironomus	0.06	0.09	0
Epeorus	0.24	0.22	0.18
Ephemera	0	0	0
Ephemerella	0.65	0.61	1
Erioptera	0	0	0
Eukiefferiella	0.26	0.3	0.48
Eurylophella	0.2	0.14	0.44
Fallicon	0	0	0
Gammarus	0.13	0.13	0.26
Glossosoma	0.19	0.13	0.29
Gomphidae	0.21	0.18	0.49
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleliella	0	0	0
Helichus	0	0.08	0
Helius	0	0	0
Hemerodromia	0.34	0.27	0.56
Heptagenia	0	0	0
Heterotrissocladius	0.12	0.09	0
Hexagenia	0	0	0
Hexatoma	0.13	0.13	0
Hyalella	0	0	0
Hydrobaenus	0.13	0.09	0.22
Hydrobius	0.06	0.08	0
Hydroporus	0	0	0
Hydropsyche	0.31	0.3	0.29
Ironoquia	0	0.04	0
Isonychia	0.05	0.04	0
Isoperla	0.06	0.13	0
Kiefferulus	0.06	0.04	0
Krenopelopia	0.06	0.04	0
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0.06	0.04	0
Lepidostoma	0	0.04	0

Leptophlebia	0.11	0.09	0
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.12	0.17	0
Limnophila	0.06	0.04	0
Limnophyes	0.06	0.04	0
Lopescladius	0	0	0
Lumbriculidae	0.22	0.21	0
Lype	0.06	0.04	0
Macronychus	0	0.04	0
Menetus	0	0.04	0
Menopelopia	0.12	0.09	0
Microasema	0	0	0
Micropsocra	0.39	0.35	0.44
Microtendipes	0.07	0.09	0.22
Microvelia	0	0	0
Molophilus	0	0	0
Muscillum	0	0.04	0
Naididae	0.27	0.35	0.47
Nanocladius	0	0.04	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.51	0.56	0.74
Nigronia	0.24	0.26	0
Nixa	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.34	0.39	0.6
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostracerca	0	0	0
Oulimnius	0.48	0.52	0.95
Pagastia	0.08	0.05	0.29
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaelocladus	0	0	0
Parachironomus	0.06	0.04	0
Paragnetia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.14	0.13	0.29
Paramerina	0.06	0.04	0
Parametrocnemus	0.69	0.65	0.77
Paraphaenocladus	0	0.04	0
Paratanytarsus	0.13	0.22	0.22
Pedicia	0	0.04	0
Pelodytes	0.06	0.09	0
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsocra	0.07	0.05	0.22
Phaenaglyphora	0	0	0
Phycocentropus	0	0	0
Physidae	0.11	0.13	0
Pisidium	0.08	0.09	0.29
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0.04	0
Polycentropus	0.28	0.22	0.56
Polypedium	0.31	0.35	0.29
Pothastia	0.06	0.04	0
Probozia	0	0	0
Procladius	0	0.04	0
Promesia	0	0	0
Prosimulium	0.57	0.56	0.71
Prostia	0.26	0.34	0.44
Prostoma	0	0	0
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.12	0.09	0
Pseudorthocladius	0.05	0.04	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0	0.04	0
Ptilostomis	0	0.09	0
Pycnopsyche	0.12	0.09	0
Radix	0	0.04	0
Rheocricotopus	0.06	0.09	0
Rheotanytarsus	0.2	0.17	0.29
Rhyacophila	0.52	0.52	0.71
Serratella	0.14	0.21	0.29
Sialis	0.06	0.04	0
Simulium	0.25	0.3	0.29
Siphonurus	0	0	0
Siphloplecton	0	0	0
Sphaeriidae	0.06	0.04	0
Sphaerium	0	0.04	0
Sphaeromias	0	0	0
Stegopterna	0.32	0.22	0.26
Stempellina	0	0	0
Stempellinella	0.05	0.04	0
Stenacron	0	0	0
Stenelmis	0.07	0.13	0.18
Stenonema	0.43	0.39	0.29
Stictochironomus	0.06	0.04	0
Stilocladius	0.05	0.04	0
Strophopteryx	0.13	0.13	0.26
Stygobromus	0	0	0
Symposiocladus	0	0	0
Symphothastia	0.06	0.09	0
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanyptus	0	0.04	0
Tanytarsus	0.37	0.35	0.29
Thienemannella	0.12	0.21	0
Thienemannimyia(grp)	0	0.04	0
Tipula	0.36	0	

Biodiversity for NWGT 101:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI	Ulenidae	0	0	0	Leucotrichia	0	0	0
Aeshnidae	0	0	0	Uronidae	0	0	0	Leucrocotia	0.06	0.03	0
Ameletidae	0	0	0	Valvidae	0	0	0	Leuctridae	0.23	0.39	0
Ancylidae	0	0	0	Velidae	0	0	0	Linnophila	0	0	0
Asellidae	0.08	0	0	Viviparidae	0	0	0	Linnophyes	0	0.06	0
Astacidae	0	0	0					Lopescladius	0	0	0
Athericidae	0	0	0					Lumbriculidae	0.17	0.18	0.17
Baetidae	0.62	0.77	0.71					Lypus	0	0	0
Baetiscidae	0	0	0					Macronychus	0	0	0
Belostomatidae	0	0	0					Menetus	0	0	0
Bithyniidae	0	0	0					Meropelopia	0.13	0.09	0.17
Blephariceridae	0	0	0					Micrasema	0	0	0
Brachycentridae	0.04	0	0					Micropectra	0.55	0.55	0.78
Braconidae	0	0	0					Microtendipes	0	0.09	0
Branchiobdellidae	0	0	0					Microvelia	0	0	0
Caenidae	0	0	0					Molophilus	0	0	0
Calamoceratidae	0	0	0					Musculium	0	0	0
Calopterygidae	0.1	0	0					Naididae	0.49	0.28	0.6
Cambaridae	0	0	0					Nanocladius	0.11	0.06	0
Capniidae	0	0	0					Nematoomorpha	0	0	0
Carabidae?	0	0	0					Nemoura	0	0	0
Ceratopogonidae	0.12	0.08	0.13					Neophylax	0.76	0.58	0.6
Chironomidae	1	1	1					Nigrionia	0.23	0.24	0.17
Chironomidae	0.1	0.08	0.11					Nixe	0	0	0
Coenagrionidae	0	0	0					Oecetis	0.05	0.03	0
Collembola	0	0	0					Omnophlyx	0	0.03	0
Corbiculidae	0	0	0					Optiosorus	0.36	0.34	0.23
Coruliogastridae	0	0	0					Ormosia	0	0	0
Coruliidae	0	0	0					Orthoclaudiinae(A)	0	0	0
Corixidae	0	0	0					Ostrocerca	0	0	0
Corydalidae	0.23	0.31	0.21					Oulimnius	0.52	0.49	0.34
Crangonyctidae	0	0	0					Pagastia	0.05	0.03	0
Culicidae	0	0	0					Palaeomonetes	0	0	0
Cumuliionidae	0	0	0					Paracapia	0.05	0.03	0
Diplostroca	0	0	0					Parachaetocladius	0.06	0.03	0
Dixidae	0.1	0.15	0.11					Parachironomus	0	0.03	0
Dolichopodidae	0	0	0					Paragnetina	0	0	0
Dryopidae	0.14	0.15	0.16					Parakiefferlebia	0	0	0
Dugesidae	0	0	0					Paraleptophlebia	0.22	0.3	0
Dytiscidae	0	0	0					Paramerina	0	0.03	0
Elmidae	0.82	0.93	0.99					Parameletrocnemus	0.64	0.61	0.56
Empididae	0.53	0.38	0.38					Paraphaenocladus	0	0	0
Enchytraeidae	0	0	0					Paralatyarsus	0.12	0.06	0.3
Ephemereidae	0.96	1	1					Pedicia	0	0	0
Ephemeridae	0	0	0					Pelodytes	0	0.03	0
Ephydriidae	0	0	0					Peltopera	0	0	0
Eprobolidae	0	0	0					Perlesta	0	0	0
Gammaridae	0.2	0	0					Phaenopsectra	0	0.03	0
Gerridae	0	0	0					Phasganophora	0	0	0
Glossiphoniidae	0	0.08	0.01					Phyllocentropus	0	0	0
Glossosomatidae	0.2	0.08	0.01					Physidae	0.12	0.09	0.17
Gomphidae	0.19	0.15	0.18					Pisidium	0.05	0.03	0
Gordidae	0.04	0	0					Planorbella	0	0	0
Gyrinidae	0	0	0					Platycentropus	0.05	0.03	0
Halplidae	0	0	0					Platysmittia	0.06	0.03	0
Helodidae	0	0	0					Polycentropus	0.24	0.16	0.27
Heptageniidae	0.66	0.77	0.74					Polypedium	0.42	0.22	0.56
Hydrobiidae	0	0	0					Pothastia	0.08	0.06	0.27
Hydrophidae	0.06	0.08	0.11					Probestia	0.07	0.06	0.22
Hydropsychidae	0.78	0.77	0.74					Procladius	0	0.03	0
Hydroptilidae	0.1	0	0					Promoresia	0.05	0.09	0
Lepidoptera	0	0	0					Prosimulium	0.69	0.78	0.51
Lepidostomatidae	0.1	0.15	0.11					Prostia	0.35	0.19	0.34
Leptoceridae	0	0	0					Prostoma	0	0	0
Leptophlebiidae	0.18	0.23	0.18					Psephenus	0.06	0.06	0
Lestidae	0	0	0					Pseudochironomus	0	0	0
Leuctridae	0.15	0.23	0.21					Pseudolimnophila	0.07	0.06	0.17
Libellulidae	0	0	0					Pseudorthocladius	0.05	0.03	0
Limnephilidae	0.58	0.54	0.72					Pseudosuccinea	0	0	0
Lumbriculidae	0.08	0.08	0.11					Psilobrya	0	0	0
Lymnaeidae	0.04	0	0					Pteronarcyis	0.12	0.09	0
Macromiidae	0	0	0					Ptilostomis	0	0	0
Maxillopoda	0	0	0					Pycnopsyche	0	0.03	0
Metretopodidae	0	0	0					Radix	0	0	0
Molannidae	0	0	0					Rheocricotopus	0.07	0.09	0.22
Muscidae	0	0	0					Rhosia	0.18	0.15	0.27
Naididae	0.54	0.31	0.45					Rhyacophila	0.4	0.51	0.04
Nemouridae	0.85	1	1					Serratella	0.22	0.15	0
Nepidae	0	0	0					Sialis	0.07	0.03	0
Notonectidae	0	0	0					Simulium	0.28	0.21	0
Odonotercidae	0	0	0					Siphonurus	0	0	0
Oligoneuridae	0.1	0.08	0.02					Siphonurina	0.18	0.09	0.3
Ostracoda	0	0	0					Sphaeridius	0.07	0.03	0.17
Palaemonidae	0.04	0	0					Sphaerium	0.07	0.03	0.22
Peltoperidae	0.05	0.16	0.18					Sphaeromias	0	0	0
Perliidae	0.25	0.15	0.07					Stegoplerma	0.27	0.3	0.66
Periodidae	0.47	0.47	0.61					Stempellina	0.06	0.06	0
Philopotamidae	0.33	0.46	0.43					Stempellinella	0.05	0.09	0
Phryganeidae	0	0	0					Stenacron	0	0.06	0
Physidae	0.06	0.08	0.11					Stenelmis	0.11	0.09	0.04
Piscicolidae	0	0	0					Stenonema	0.47	0.43	0.44
Pisidiidae	0.08	0	0					Stictochironomus	0	0.03	0
Planariidae	0.09	0.08	0.1					Stitocladius	0.05	0.03	0
Planorbidae	0	0	0					Strophopteryx	0.18	0.09	0.3
Pleuroceridae	0.1	0	0					Stygobromus	0	0.03	0
Polycentropodidae	0.24	0.16	0.22					Symposiocladius	0	0	0
Potamanthidae?	0	0	0					Sympothastia	0.19	0.1	0.27
Psephenidae	0.04	0	0					Synurella	0	0.03	0
Psychodidae	0	0	0					Tabanus	0	0	0
Psychomyiidae	0.04	0	0					Taeniopteryx	0	0.03	0
Pteronarcyidae	0.1	0.15	0.16					Talaperla	0.05	0.09	0
Ptilodactylidae?	0.14	0.15	0.15					Tanypterus	0	0	0
Psychopteryidae	0	0	0					Tanytarsus	0.4	0.22	0.27
Rhyacophilidae	0.55	0.54	0.62					Thienemannella	0.26	0.1	0.56
Saldidae	0	0	0					Thienemannimyia(grp)	0.12	0.09	0.3
Sciomyzidae	0	0	0					Tipula	0.29	0.27	0.17
Sericostomatidae	0	0	0					Tropobates	0	0	0
Sialidae	0	0	0					Trienodes	0.05	0.03	0
Simuliidae	0.95	0.92	0.92					Tribelos	0	0	0
Siphonuridae	0.22	0.23	0.31					Trissopelopia	0.31	0.22	0.56
Sisyridae	0	0	0					Tubificidae	0.08	0.06	0.27
Sphaeriidae	0	0	0					Tvetenia	0.13	0.09	0.27
Stratiomyidae	0	0	0					Umnella	0	0	0
Syrphidae	0	0	0					Umnella	0	0.03	0
Tabanidae	0	0	0					Umnella	0	0.03	0
Taeniopterygidae	0.1	0	0					Umnella	0	0.03	0
Talitridae	0	0	0					Xylotopus	0	0	0
Tetrastemmatidae	0	0	0					Zavrelimyia	0.26	0.16	0.48
Tipulidae	0.78	0.7	0.69								
Tubificidae	0.05	0.08	0.11								

Biodiversity for NWLT 101:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC		Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC					
Aeshnidae	0	0	0	Uenidae	0	0	0	Leptophlebia	0.06	0.09	0.08	
Ameletidae	0	0	0	Unionidae	0	0	0	Leptoxis	0	0	0	
Ancylidae	0	0	0	Valvatidae	0	0	0	Leucotrichia	0	0	0	
Asellidae	0.11	0	0	Velidae	0	0	0	Leucrocota	0.05	0.05	0	
Astacidae	0	0	0	Viviparidae	0	0	0	Leuctridae	0.21	0.23	0	
Athericidae	0	0	0					Limnophila	0	0.05	0	
Baetidae	0.61	0.7	0.61	Genus:				Limnophyes	0.07	0.05	0.12	
Baetiscidae	0	0	0	Ablabesmyia	0.06	0.05	0.08	Lopescladius	0	0	0	
Belostomatidae	0	0	0	Acetrotella	0.11	0.14	0.04	Lumbriculiidae	0.18	0.14	0.08	
Bithyniidae	0	0	0	Acerpenna	0.19	0.15	0.16	Lype	0	0	0	
Blephariceridae	0	0	0	Acroneria	0.09	0.05	0.22	Macronychus	0	0	0	
Brachycentridae	0.05	0	0	Agapetus	0	0.05	0	Menetus	0	0	0	
Braconidae	0	0	0	Allocapnia	0.24	0.19	0.16	Meropelopia	0.18	0.14	0.2	
Branchiobdellidae	0	0	0	Alkotrypys	0	0	0	Microasema	0	0	0	
Caenidae	0	0	0	Ameletus	0.38	0.38	0.44	Microprosectra	0.57	0.44	0.78	
Calamoceratidae	0	0	0	Amphinemura	0.79	0.76	0.78	Microtendipes	0.06	0.09	0	
Calopterygidae	0.05	0	0	Anchytarsus	0.2	0.14	0.22	Microvelia	0	0	0	
Cambaridae	0	0	0	Ancyliidae	0	0	0	Molophilus	0	0	0	
Capniidae	0.08	0.2	0.14	Ancyronyx	0	0	0	Muscillum	0	0	0	
Carabidae?	0	0	0	Antocha	0.27	0.24	0.4	Naisidae	0.46	0.34	0.62	
Ceratopogonidae	0.09	0	0	Apsectrotanypus	0	0	0	Nanocladius	0.05	0.05	0	
Chironomidae	1	1	1	Argia	0	0	0	Nematomorpha	0	0	0	
Chloroperlidae	0.08	0	0	Atherix	0	0	0	Nemoura	0	0	0	
Coenagrionidae	0	0	0	Baetis	0.06	0.05	0.04	Neophylax	0.7	0.58	0.8	
Collembola	0	0	0	Bezzia	0	0	0	Nigrinia	0.24	0.24	0.2	
Corbiculidae	0	0	0	Bitacomorpha	0	0.05	0	Nixa	0	0	0	
Coruliogastridae	0	0	0	Blepharicera	0	0	0	Oecetis	0	0	0	
Cordulidae	0	0	0	Boyeria	0	0.05	0	Oemopteryx	0	0	0	
Corixidae	0	0	0	Brachycentrus	0	0	0	Optioservus	0.46	0.43	0.59	
Corydalidae	0.19	0.1	0.13	Branchiobdellidae	0	0	0	Ormosia	0	0	0	
Crangonyctidae	0	0	0	Brillia	0.06	0.09	0	Orthoclaudiinae(A)	0	0	0	
Culicidae	0	0	0	Caecidotea	0.05	0	0	Ostrocerca	0	0	0	
Curculionidae	0	0	0	Caenis	0	0.05	0	Oulimnius	0.63	0.62	0.74	
Diplostroca	0	0	0	Callibaetis	0	0.05	0	Pagasia	0.09	0.05	0.22	
Dixidae	0	0	0	Calopteryx	0	0.05	0	Palaemonetes	0	0	0	
Dolichopodidae	0	0	0	Cambaridae	0.06	0.05	0.08	Paracapnia	0	0	0	
Dryopidae	0.09	0	0	Centroptilum	0.06	0.05	0.08	Parachaetocladius	0.05	0.05	0	
Dugesidae	0	0	0	Ceratopogon	0.2	0.15	0.38	Parachironomus	0	0.05	0	
Dytiscidae	0	0	0	Chaetocladius	0.05	0.05	0	Parametrioctenium	0.7	0.67	0.74	
Elmidae	0.86	0.81	0.93	Chellifera	0.11	0.09	0	Paraphaenocladus	0	0	0	
Empididae	0.59	0.7	0.69	Cheumatopsyche	0.47	0.43	0.76	Paratanypus	0.12	0.19	0.04	
Enchytraeidae	0	0	0	Chimarra	0.13	0.1	0.18	Pedicia	0	0	0	
Ephemeroidea	0.95	1	0.98	Chironomus	0	0	0	Pelodytes	0	0.05	0	
Ephemeridae	0	0	0	Chloroperlidae	0.11	0.14	0	Peloperla	0	0	0	
Ephyridae	0	0	0	Chrysops	0	0	0	Perlesta	0	0	0	
Epobdellidae	0	0	0	Cinygmula	0	0	0	Phaenopsectra	0.06	0.05	0	
Gammaridae	0.2	0	0.07	Cladopelma	0	0	0	Phaenoglyphura	0	0	0	
Gerridae	0	0	0	Cladotanytarsus	0	0	0	Phycocentropus	0	0	0	
Glossiphoniidae	0	0	0	Clinochera	0.19	0.24	0.18	Physidae	0.06	0.14	0.08	
Glossosomatidae	0.4	0.5	0.48	Clinotanytarsus	0	0	0	Pisidium	0.09	0.05	0.22	
Gomphidae	0.09	0	0	Cloperla	0	0	0	Planorbella	0	0	0	
Gordidae	0.06	0	0	Cnephia	0	0	0	Platycentropus	0	0	0	
Gyrinidae	0	0	0	Collembola	0.06	0.05	0.08	Platysmittia	0.05	0.05	0	
Halipidae	0	0	0	Conchapelopia	0.52	0.57	0.62	Polycentropus	0.29	0.24	0.59	
Helodidae	0	0	0	Cordulegaster	0.06	0.05	0.08	Polypedium	0.39	0.34	0.44	
Heptageniidae	0.6	0.61	0.67	Cordulilidae	0	0.05	0	Pothastia	0.07	0.05	0.18	
Hydrobiidae	0	0	0	Corixidae	0	0.05	0	Probezzia	0	0	0	
Hydrophidae	0.04	0	0	Corydalis	0	0	0	Procladius	0	0	0	
Hydropsychidae	0.84	0.9	0.85	Corynorina	0.27	0.29	0.48	Promesia	0	0	0	
Hydroptilidae	0.11	0	0	Crangonyx	0.07	0.1	0.18	Prosimulium	0.73	0.67	0.6	
Lepidoptera	0	0	0	CricobopuOrthocladus	0.57	0.57	0.52	Prostoma	0.32	0.33	0.4	
Lepidostomatidae	0	0	0	Cryptochironomus	0	0	0	Pseudocricotopus	0	0	0	
Leptoceridae	0	0	0	Culicoides	0	0	0	Psephenus	0.05	0.05	0	
Leptophlebiidae	0.19	0.1	0.2	Cullis	0	0	0	Pseudochironomus	0	0	0	
Lestidae	0	0	0	Cura	0	0	0	Pseudolimnophila	0.13	0.1	0.2	
Leuctridae	0.04	0.1	0.06	Deronectes	0	0.05	0	Pseudorthocladus	0.06	0.05	0	
Libellulidae	0	0	0	Diamesa	0.25	0.19	0.22	Pseudosuccinea	0	0	0	
Limnephilidae	0.5	0.5	0.49	Dicranota	0.16	0.19	0.04	Psilotreta	0	0	0	
Lumbriculiidae	0.13	0.2	0.17	Dicrolendipes	0	0	0	Pteronarcys	0.1	0.09	0	
Lymnaeidae	0.05	0	0	Dineutus	0	0	0	Ptilostomis	0	0	0	
Macromiidae	0	0	0	Diphetor	0.05	0.05	0	Pycnopsyche	0.07	0.09	0.12	
Maxillopoda	0	0	0	Diplectrona	0.38	0.38	0.52	Radix	0	0	0	
Metretopodidae	0	0	0	Diplocladius	0.07	0.1	0.18	Rheocricotopus	0	0.05	0	
Molannidae	0	0	0	Dixa	0.05	0.05	0	Rheotanytarsus	0.22	0.15	0.4	
Muscidae	0	0	0	Dolichopodidae	0.06	0.05	0.08	Rhyacophila	0.65	0.62	0.48	
Naisidae	0.31	0.31	0.31	Dolophilodes	0.16	0.14	0	Serratella	0.25	0.24	0.22	
Nemouridae	0.75	0.7	0.67	Drunella	0.11	0.1	0	Sialis	0.07	0.05	0.12	
Nepidae	0	0	0	Dubiraphia	0.06	0.05	0	Simulium	0.31	0.28	0.22	
Notonectidae	0	0	0	Dugesia	0.05	0.05	0	Siphonurus	0	0	0	
Odontoceridae	0	0	0	Ecocoptura	0.07	0.05	0.12	Siphloplecton	0	0	0	
Oligoneuridae	0.14	0.1	0.14	Ectopria	0	0	0	Sphaeriidae	0.06	0.05	0.08	
Ostracoda	0	0	0	Enallagma	0	0.05	0	Sphaerium	0	0	0	
Palaemonidae	0	0	0	Enchytraeidae	0.06	0.05	0.08	Stegopterna	0.28	0.24	0.56	
Pelopelidae	0.04	0	0	Endochironomus	0	0.09	0	Stempellina	0.05	0.05	0	
Perlidae	0.36	0.3	0.3	Epeorus	0.29	0.28	0.18	Stempellinella	0.06	0.05	0	
Periodidae	0.33	0.2	0.13	Ephemera	0.75	0.71	0.76	Stenacron	0	0	0	
Philopotamidae	0.31	0.4	0.42	Ephemerella	0	0	0	Stenelmis	0.07	0.1	0.18	
Phryganeidae	0	0	0	Erioptera	0	0	0	Stenonema	0.51	0.43	0.6	
Physidae	0.04	0	0	Eukiefferiella	0.25	0.33	0.22	Sitochironomus	0.07	0.05	0.12	
Piscicolidae	0	0	0	Eurylophella	0.21	0.15	0.44	Sitocladius	0.06	0.05	0	
Pisidiidae	0.11	0.1	0.14	Falceon	0	0	0	Strophopteryx	0.19	0.15	0.22	
Planariidae	0.06	0	0	Gammarus	0.07	0.05	0.16	Stygobromus	0	0	0	
Planorbidae	0	0	0	Glossosoma	0.2	0.15	0.22	Symposiocladus	0	0	0	
Pleuroceridae	0.05	0	0	Gomphidae	0.2	0.15	0.36	Symphothasta	0.13	0.14	0.18	
Polycentropodidae	0.17	0.2	0.25	Goniobasis	0	0	0	Synurella	0	0	0	
Potamanthidae?	0	0	0	Gyrinus	0	0	0	Tabanus	0	0	0	
Psephenidae	0.06	0	0	Habrophlebia	0	0	0	Taeniopteryx	0	0	0	
Psychodidae	0	0	0	Helemlia	0	0	0	Tallaperla	0	0	0	
Psychomyiidae	0.06	0	0	Helichus	0.05	0.05	0	Tanytus	0	0	0	
Pteronarcidae	0.04	0	0	Hellus	0	0	0	Tanytarsus	0.27	0.33	0.4	
Ptilodactylidae?	0.06	0.1	0.14	Hemerodromia	0.33	0.29	0.4	Thienemannella	0.19	0.24	0.22	
Ptychopteridae	0	0	0	Heptagenia	0	0	0	Thienemannimyia(grp)	0.06	0.05	0.04	
Rhyacophilidae	0.46	0.4	0.42	Heterotrissocladus	0.06	0.09	0.08	Tipula	0.37	0.38	0.38	
Saldidae	0	0	0	Hexagenia	0	0	0	Trepobates	0	0	0	
Sciomyzidae	0	0	0	Hexatoma	0.19	0.15	0.3	Trienodes	0	0.05	0	
Sericostomatidae	0	0	0	Hyalella	0	0	0	Tribelos	0	0.05	0	
Sialidae	0	0	0	Hydrobaenus	0.06	0.09	0	Trisopelopia	0.32	0.29	0.34	
Simuliidae	0.94	0.9	0.87	Hydrobius	0.06	0.09	0.08	Tubificidae	0.07	0.14	0.18	
Siphonuridae	0.41	0.6	0.59	Hydroponus	0	0	0	Tvetenia	0.07	0.05	0.18	
Sisyridae	0	0	0	Hydropsyche	0.27	0.29	0.4	Unniella	0	0	0	
Sphaeriidae	0	0	0	Ironoquia	0	0.05	0	Wormaldia	0	0.05	0	
Stratiomyidae	0	0	0	Isonychia	0	0.05	0	Xylotopus	0	0	0	
Syrphidae	0	0	0	Isoperla	0.18	0.14	0.12	Zavreimyia	0.26	0.19	0.44	
Tabanidae	0	0	0	Kiefferulus	0	0.05	0					
Taeniopterygidae	0.09	0.1	0.08	Krenopelopia	0.07	0.05	0.12					
Talitridae	0	0	0	Labrundinia	0	0	0					
Tetrastemmatidae	0	0	0	Larsia	0	0	0					
Tipulidae	0.82	0.8	0.87	Lepidoptera	0.06	0.05	0.08					
Tubificidae	0.15	0.2	0.31	Lepidostoma	0.05	0.05	0					

Biodiversity for NWNW 301:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.1	0	0
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.62	0.67	0.63
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.04	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.04	0	0
Cambaridae	0	0	0
Capniidae	0.17	0.22	0.22
Carabidae?	0	0	0
Ceratopogonidae	0.04	0	0
Chironomidae	1	1	1
Chloroperlidae	0	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corticulidae	0	0	0
Cordulegastriidae	0	0	0
Cordulidae	0	0	0
Corydalidae	0.19	0.11	0.09
Crangonyctidae	0.04	0	0.04
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostrota	0	0	0
Dixidae	0	0	0
Dolichopodidae	0	0	0
Dryopidae	0.09	0	0
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.86	0.78	0.92
Empididae	0.63	0.78	0.8
Enchytraeidae	0	0	0
Ephemereilidae	0.82	1	0.9
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ergasilidae	0	0	0
Gammauridae	0.19	0	0.05
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.4	0.45	0.49
Gomphidae	0.04	0	0
Gordidae	0.05	0	0
Gyrinidae	0	0	0
Hallipidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.67	0.56	0.64
Hydrobiidae	0	0	0
Hydrophilidae	0	0	0
Hydropsychidae	0.79	0.89	0.79
Hydroptilidae	0.1	0	0
Lepidoptera	0	0	0
Lepidostomatidae	0	0	0
Leploceridae	0	0	0
Leptophlebiidae	0.15	0.11	0.21
Lestidae	0	0	0
Leuctridae	0.04	0.11	0.06
Libellulidae	0	0	0
Limnephilidae	0.46	0.55	0.43
Lumbriculidae	0.13	0.22	0.16
Lymnaeidae	0.04	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.3	0.33	0.24
Nemouridae	0.78	0.66	0.6
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.27	0.12	0.27
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.04	0	0
Perlidae	0.36	0.23	0.33
Periodidae	0.27	0.22	0.04
Philopotamidae	0.33	0.33	0.27
Phryganeidae	0	0	0
Physidae	0	0	0
Piscicolidae	0	0	0
Pisidiidae	0.11	0.11	0.16
Planariidae	0.05	0	0
Planorbidae	0	0	0
Pleuroceridae	0.04	0	0
Polycentropodidae	0.16	0.22	0.18
Potamanthidae?	0	0	0
Psephenidae	0.05	0	0
Psychodidae	0	0	0
Psychomyiidae	0.1	0	0
Pteronarcidae	0.04	0	0
Ptilodactylidae?	0.06	0.11	0.16
Ptychopteridae	0	0	0
Rhyacophilidae	0.38	0.44	0.42
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.95	0.89	0.91
Siphonuridae	0.38	0.67	0.58
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.22	0.11	0.16
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.8	0.78	0.77
Tubificidae	0.15	0.22	0.24

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Ablabesmyia	0	0.04	0
Acetrotella	0.05	0.11	0
Acerpenna	0.11	0.18	0
Acroneria	0.08	0.08	0.35
Agapetus	0.05	0.04	0
Allocapnia	0.11	0.15	0
Alotrypops	0.05	0.04	0
Ametetus	0.35	0.3	0.03
Amphinemura	0.61	0.66	0.39
Anchytarsus	0.14	0.11	0.35
Ancyliidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.26	0.22	0.35
Aspsectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0	0.07	0
Bezzia	0	0	0
Bitacomorpha	0.06	0.04	0
Blepharicera	0	0	0
Boyeria	0.12	0.08	0
Brachycentrus	0	0.04	0
Branchiobdellidae	0	0	0
Brillia	0.11	0.18	0
Caesiodotea	0.11	0.11	0
Caenis	0.11	0.07	0
Calibaetis	0.06	0.04	0
Calopteryx	0.06	0.04	0
Cambaridae	0	0.04	0
Centroptilum	0	0.07	0
Ceratopogon	0.13	0.11	0.03
Chaetocladius	0	0.04	0
Chellera	0.07	0.07	0.25
Cheumatopsyche	0.4	0.45	0.75
Chimarra	0.11	0.08	0
Chironomus	0.05	0.04	0
Chloroperlidae	0.07	0.07	0.25
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0.03	0
Clinocera	0.3	0.19	0.36
Clinotanytarsus	0	0	0
Cloperia	0	0	0
Cnephia	0	0	0
Collembola	0	0.04	0
Conchapelopia	0.68	0.6	0.64
Cordulegaster	0	0.04	0
Cordulidae	0.06	0.04	0
Corydalis	0.06	0.04	0
Corydatus	0	0.03	0
Corynoneura	0.24	0.26	0.03
Crangonyx	0.14	0.11	0.36
Cricolopus/Orthocladius	0.75	0.67	0.97
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.06	0.04	0
Diamesa	0.22	0.19	0.61
Dicranota	0.06	0.11	0
Dicrolenidipes	0.05	0.04	0
Dineutus	0	0	0
Dipheter	0.06	0.04	0
Diplectrona	0.25	0.26	0.03
Diplocladius	0.05	0.04	0
Diploperia	0.06	0.07	0
Dixa	0	0.04	0
Dolichopodidae	0	0.04	0
Dolophilodes	0.06	0.07	0
Drunella	0.11	0.07	0
Dubiraphia	0.05	0.04	0
Dugesia	0	0.04	0
Ecopoptura	0.12	0.08	0
Ectopria	0	0	0
Enallagma	0.06	0.04	0
Enchytraeidae	0.05	0.07	0
Endochironomus	0.11	0.07	0
Epeorus	0.24	0.19	0.03
Ephemera	0	0	0
Ephemerella	0.71	0.63	1
Erioptera	0	0	0
Eukiefferiella	0.32	0.33	0.61
Eurylophella	0.15	0.15	0.39
Falleon	0	0	0
Gammarus	0.14	0.11	0.36
Glossosoma	0.13	0.15	0.35
Gomphidae	0.21	0.15	0.39
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleliella	0	0	0
Helichus	0	0.07	0
Helius	0	0.03	0
Hemerodromia	0.33	0.3	0.61
Heptagenia	0	0	0
Heterotrissocladius	0.06	0.08	0
Hexagenia	0	0	0
Hexatoma	0.12	0.11	0
Hyalella	0	0	0
Hydrobaenus	0.13	0.15	0.25
Hydrobius	0	0.07	0
Hydroporus	0	0	0
Hydropsyche	0.37	0.33	0.35
Ironoquia	0.03	0	0
Isonychia	0.06	0.07	0
Isoperla	0	0.11	0
Kiefferulus	0.06	0.04	0
Krenopelopia	0.06	0.04	0
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0	0.04	0
Lepidostoma	0	0.04	0

Leptophlebia	0.06	0.07	0
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.12	0.15	0
Limnophila	0.06	0.04	0
Limnophyes	0.06	0.04	0
Lopescladius	0	0	0
Lumbriculidae	0.22	0.22	0
Lype	0.06	0.07	0
Macronychus	0	0.04	0
Menetus	0.05	0.04	0
Menopelopia	0.06	0.07	0
Microasema	0	0	0
Microsectra	0.33	0.3	0.39
Microtendipes	0.13	0.08	0.25
Microvelia	0	0	0
Molophilus	0	0	0
Muscillum	0.05	0.04	0
Naididae	0.31	0.33	0.39
Nanocladius	0	0.04	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.45	0.48	0.75
Nigronia	0.18	0.22	0
Nixa	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0.03	0
Optioservus	0.34	0.37	0.72
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0.03	0
Ostracerca	0	0	0
Oulimmus	0.53	0.49	0
Pagastia	0.08	0.04	0.35
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaeocladius	0	0	0
Parachironomus	0.06	0.04	0
Paragnetia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.14	0.15	0.35
Paramerina	0.06	0.04	0
Parametrichonemus	0.69	0.63	0.97
Paraphaenocladus	0.05	0.04	0
Paratanytarsus	0.24	0.19	0.25
Pedicia	0.05	0.04	0
Pelodytes	0.11	0.07	0
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.07	0.04	0.25
Phaenoglyphura	0	0	0
Phycocentropus	0	0	0
Physidae	0.06	0.11	0
Pisidium	0.08	0.08	0.35
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0.04	0
Polycentropus	0.28	0.19	0.72
Polypedium	0.36	0.33	0.35
Pothastia	0.06	0.04	0
Probozia	0	0	0
Procladius	0.05	0.04	0
Pronesia	0	0	0
Prosimulium	0.45	0.58	0.65
Prostoma	0.26	0.37	0.39
Pseudocricotopus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.06	0.07	0
Pseudorthocladius	0.05	0.04	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0	0.04	0
Ptilostomis	0	0	0
Pycnopsyche	0.12	0.11	0
Radix	0	0.04	0
Rheocricotopus	0.06	0.07	0
Rheotanytarsus	0.19	0.18	0.35
Rhyacophila	0.57	0.49	0.65
Serratella	0.14	0.18	0.35
Sialis	0.06	0.04	0
Simulium	0.25	0.36	0.35
Siphonurus	0	0	0
Siphonoclepton	0	0	0
Sphaeriidae	0	0.04	0
Sphaerium	0	0.04	0
Sphaeromias	0	0	0
Stegopterna	0.26	0.19	0.36
Stempellina	0	0	0
Stempellinella	0.05	0.04	0
Stenacron	0	0	0
Stenelmis	0.07	0.14	0.03
Stenonema	0.37	0.37	0.35
Stictochironomus	0.06	0.04	0
Stilocladius	0	0.04	0
Strophopteryx	0.08	0.15	0.36
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphothastia	0.11	0.11	0
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytarsus	0.05	0.04	0
Tanytarsus	0.37	0.4	0.35
Thienemannella	0.17	0.25	0
Thienemannimyia(gp)	0	0.04	0
Tipula	0.26	0.3	0.36
Trepobates	0	0	0
Trienodes	0.06	0.04	0
Tribelos	0.06	0.04	0
Trisopelopia	0.29	0.22	0
Tubificidae	0.29	0.22	0

Biodiversity for NWNW 304:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.11	0.1	0.03
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.61	0.6	0.63
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.05	0	0
Cambardae	0	0	0
Capniidae	0.08	0.19	0.09
Carabidae?	0	0	0
Ceratopogonidae	0.09	0	0
Chironomidae	1	1	1
Chironomidae	0.08	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corbiculidae	0	0	0
Contullegastridae	0	0	0
Corduidae	0	0	0
Corixidae	0	0	0
Corydidae	0.19	0.1	0.15
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0	0	0
Dolichopodidae	0	0	0
Dryopidae	0.09	0	0
Dugesidae	0	0	0
Dytiscidae	0	0	0
Eimidae	0.86	0.8	0.9
Empididae	0.59	0.8	0.76
Enchytraeidae	0	0	0
Ephemereididae	0.95	1	1
Ephemeridae	0	0	0
Ephyridae	0	0	0
Eprobolidae	0	0	0
Gammaridae	0.2	0.1	0.07
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.39	0.5	0.45
Gomphidae	0.08	0	0
Gonidae	0.06	0.1	0.03
Gyrinidae	0	0	0
Halplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.6	0.6	0.68
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.84	0.9	0.92
Hydroptilidae	0.11	0.1	0.03
Lepidoptera	0	0	0
Lepidostomatidae	0	0	0
Leptoceridae	0	0	0
Leptophlebiidae	0.19	0.1	0.19
Lestidae	0	0	0
Leuctridae	0.04	0.1	0.04
Libellulidae	0	0	0
Limnephilidae	0.5	0.5	0.46
Lumbriculidae	0.12	0.2	0.14
Lymnaeidae	0.05	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naicidae	0.32	0.4	0.34
Nemouridae	0.76	0.7	0.62
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.14	0.1	0.12
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperidae	0.04	0	0
Perliidae	0.36	0.3	0.3
Periodidae	0.33	0.2	0.19
Philopotamidae	0.32	0.4	0.41
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.11	0.1	0.15
Planariidae	0.06	0.1	0.03
Planorbidae	0	0	0
Pleuroceridae	0.05	0	0
Polycerentopodidae	0.17	0.2	0.28
Potamanthidae?	0	0	0
Psephenidae	0.06	0.1	0.03
Psychodidae	0	0	0
Psychomyiidae	0.06	0.1	0.03
Pteronarcidae	0.04	0	0
Ptilodactylidae?	0.06	0.1	0.15
Ptychopteridae	0	0	0
Rhyacophilidae	0.46	0.4	0.42
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.93	0.9	0.85
Siphonuridae	0.41	0.6	0.57
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.09	0.1	0.05
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.82	0.8	0.81
Tubificidae	0.15	0.2	0.3

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Ablabesmyia	0	0.05	0
Acetrotella	0.16	0.15	0
Acerpenna	0.16	0.15	0
Acroneria	0.09	0.05	0.19
Agapetus	0.05	0.05	0
Allocaenia	0.21	0.2	0
Alotrypops	0	0	0
Ameletus	0.42	0.4	0.29
Amphinemura	0.65	0.74	0.49
Anchytarsus	0.16	0.15	0.33
Ancyliidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.27	0.25	0.39
Assectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.05	0.05	0
Bezzia	0	0	0
Bitacomorpha	0	0.05	0
Blepharicera	0	0	0
Boyeria	0	0.05	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brillia	0.11	0.1	0
Caecidotea	0.06	0.05	0
Caenis	0.06	0.05	0
Calibaetis	0.06	0.05	0
Calopteryx	0.06	0.05	0
Cambaridae	0	0.05	0
Centroptilum	0	0.05	0
Ceratopogon	0.13	0.15	0.18
Chaetocladius	0.05	0.05	0
Chelifera	0.08	0.05	0.17
Cheumatopsyche	0.52	0.45	0.68
Chimarra	0.12	0.1	0.11
Chironomus	0	0	0
Chironomidae	0.08	0.1	0.17
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.28	0.25	0.11
Clinotanytarsus	0	0	0
Cloperia	0	0	0
Cnephia	0	0	0
Collembola	0	0.05	0
Conchapelopia	0.64	0.6	0.65
Cordulegaster	0	0.05	0
Cordulidae	0.06	0.05	0
Corixidae	0.06	0.05	0
Corydatus	0	0	0
Corynoneura	0.25	0.3	0.29
Crangonyx	0.07	0.1	0.11
Cricobolus/Orthocladus	0.63	0.6	0.71
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.06	0.05	0
Diamesa	0.27	0.2	0.42
Dicranota	0.11	0.15	0
Dicrolentipes	0	0	0
Dineutus	0	0	0
Diphetor	0.06	0.05	0.1
Diplectrona	0.3	0.35	0.29
Diplocladius	0	0	0
Diploperia	0.07	0.1	0.11
Dixa	0	0.05	0
Dolichopodidae	0	0.05	0
Dolophilodes	0.12	0.1	0.1
Drunella	0.11	0.1	0
Dubiraphia	0.06	0.05	0
Dugesia	0	0.05	0
Ecoptura	0.05	0.05	0
Ectopria	0	0	0
Enallagma	0.06	0.05	0
Enchytraeidae	0	0.05	0
Endochironomus	0.11	0.1	0
Epeorus	0.31	0.25	0.28
Ephemerella	0.78	0.71	1
Erioptera	0	0	0
Eukiefferiella	0.36	0.35	0.28
Eurylophella	0.15	0.15	0.29
Fallicon	0	0	0
Gammarus	0.07	0.05	0.11
Glossosoma	0.2	0.15	0.19
Gomphidae	0.22	0.16	0.43
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleliella	0	0	0
Helichus	0.05	0.05	0
Hellus	0	0	0
Hemerodromia	0.35	0.31	0.61
Heptagenia	0	0	0
Heterotrissocladius	0	0.1	0
Hexagenia	0	0	0
Hexatoma	0.12	0.1	0.11
Hyalella	0	0	0
Hydrobaenus	0.14	0.1	0.17
Hydrobius	0	0.1	0
Hydroporus	0	0	0
Hydropsyche	0.38	0.3	0.39
Ironoquia	0	0.05	0
Isonychia	0.05	0.05	0
Isoperla	0.05	0.1	0
Kiefferulus	0.06	0.05	0
Krenopelopia	0.05	0.05	0
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0	0.05	0
Lepidostoma	0	0.05	0

Leptophlebia	0.06	0.1	0
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.11	0.2	0
Limnophila	0	0.05	0
Limnophyes	0.05	0.05	0
Lopescladius	0	0	0
Lumbriculidae	0.11	0.15	0
Lype	0	0	0
Macronychus	0	0	0
Menetus	0	0	0
Menopelopia	0.05	0.1	0
Microasema	0	0	0
Microsectra	0.38	0.4	0.49
Microtendipes	0.13	0.1	0.17
Microvelia	0	0	0
Molophilus	0	0	0
Muscillum	0	0	0
Naicidae	0.39	0.35	0.48
Nanocladius	0	0.05	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.57	0.55	0.59
Nigronia	0.12	0.2	0.1
Nixa	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.45	0.45	0.55
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostracerca	0	0	0
Oulimnius	0.64	0.6	0.74
Pagastia	0.09	0.05	0.19
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaetocladius	0	0	0
Parachironomus	0.06	0.05	0
Paragnetia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.14	0.15	0.19
Paramerina	0.06	0.05	0
Parametrichonemus	0.62	0.65	0.58
Paraphaenocladus	0	0	0
Paratanytarsus	0.24	0.2	0.17
Pedicia	0	0	0
Pelodytes	0.06	0.05	0
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.08	0.05	0.17
Phaenoglyphora	0	0	0
Phycocentropus	0	0	0
Physidae	0.06	0.15	0
Pisidium	0.09	0.05	0.19
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0.05	0
Polycentropus	0.22	0.25	0.4
Polypedium	0.37	0.35	0.3
Pothastia	0.07	0.05	0.11
Probezzia	0	0	0
Procladius	0	0	0
Promesia	0	0	0
Prosimulium	0.62	0.65	0.61
Prostoma	0.36	0.35	0.29
Pseudocricotopus	0	0	0
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.05	0.1	0
Pseudorthocladus	0.05	0.05	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0	0.05	0
Ptilostomis	0	0	0
Ptycopsyche	0.05	0.1	0
Radix	0	0	0
Rheocricotopus	0	0.05	0
Rheotanytarsus	0.21	0.15	0.3
Rhyacophila	0.63	0.6	0.7
Serratella	0.2	0.25	0.28
Sialis	0.05	0.05	0
Simulium	0.25	0.3	0.19
Siphonurus	0	0	0
Siphonoclepton	0	0	0
Sphaeriidae	0	0.05	0
Sphaerium	0	0	0
Sphaeromias	0	0	0
Sieglella	0.18	0.25	0.22
Stempellina	0	0	0
Stempellinella	0.05	0.05	0
Stenacron	0	0	0
Stenelmis	0.08	0.1	0.18
Stenonema	0.43	0.4	0.39
Stictochironomus	0.05	0.05	0
Stilocladius	0.05	0.05	0
Strophopteryx	0.17	0.15	0.11
Stygobromus	0	0	0
Symposiocladus	0	0	0
Symphotha	0.12	0.1	0.11
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytus	0	0	0
Tanytarsus	0.26	0.35	0.3
Thienemannella	0.23	0.25	0.11
Thienemannimyia(grp)	0.05	0.05	0
Tipula	0.28	0.39	0.11
Trepobates	0	0	0
Trienodes	0	0.05	0
Tribelos	0	0.05	0
Trisopelopia	0.28	0.3	0.11
Tubificidae	0.18	0.15	0.11
Tvetenia	0.07	0.05	0.11
Unniella			

Biodiversity for NWPT 101:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.1	0.1	0.05
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.6	0.7	0.58
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.06	0	0
Cambardae	0	0	0
Capniidae	0	0	0
Carabidae?	0	0	0
Ceratopogonidae	0.1	0	0.1
Chironomidae	1	1	1
Chloroperlidae	0.08	0	0.03
Coenagrionidae	0	0	0
Collembola	0	0	0
Corbiculidae	0	0	0
Cortulegastriidae	0	0	0
Cordulidae	0	0	0
Corydidae	0	0	0
Corydalidae	0.22	0.2	0.26
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0.08	0	0
Dolichopodidae	0	0	0
Dryopidae	0.13	0	0.01
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.86	0.9	0.96
Empididae	0.54	0.8	0.65
Enchytraeidae	0	0	0
Ephemereidae	0.95	1	1
Ephemeridae	0	0	0
Ephyridae	0	0	0
Ergobdellidae	0	0	0
Gammaridae	0.21	0.1	0.05
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.33	0.5	0.27
Gomphidae	0.13	0	0.13
Gordidae	0.06	0.1	0.05
Gyrinidae	0	0	0
Haliplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.64	0.8	0.82
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.84	1	0.89
Hydroptilidae	0.12	0.1	0.05
Lepidoptera	0	0	0
Lepidostomatidae	0.08	0	0
Lepiceridae	0	0	0
Leptophlebiidae	0.17	0.1	0.13
Lestidae	0	0	0
Leuctridae	0.12	0	0
Libellulidae	0	0	0
Limnephilidae	0.46	0.3	0.55
Lumbriculidae	0.05	0.1	0.08
Lymnaeidae	0.05	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretropodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.39	0.51	0.51
Nemouridae	0.83	0.7	0.79
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.13	0.2	0.02
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.05	0	0.01
Perlidae	0.38	0.4	0.21
Perlodidae	0.38	0.2	0.33
Philopotamidae	0.34	0.51	0.56
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.09	0.1	0.12
Planariidae	0.1	0.1	0.05
Planorbidae	0	0	0
Pleuroceridae	0.06	0	0
Polycentropodidae	0.2	0.21	0.36
Potamanthidae?	0	0	0
Psephenidae	0.06	0.1	0.05
Psychodidae	0	0	0
Psychomyiidae	0.06	0.1	0.05
Pteronarcidae	0.09	0	0.01
Ptilodactylidae?	0.12	0.1	0.16
Ptychopteridae	0	0	0
Rhyacophilidae	0.55	0.3	0.47
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.89	0.76
Siphonuridae	0.28	0.4	0.4
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.06	0	0
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.76	0.8	0.76
Tubificidae	0.09	0.2	0.24

Uenoidae	0	0	0
Uronidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Ragan ANNA BIC Ragan BROI AIC Ragan TROI AIC			
Genus:			
Ablabesmyia	0.06	0.05	0.08
Acetrella	0.11	0.14	0.03
Acerpenna	0.18	0.15	0.15
Acroneria	0.08	0.05	0.21
Agapetus	0	0.05	0
Allocapnia	0.24	0.19	0.15
Allostyopus	0	0	0
Ameletus	0.38	0.38	0.46
Amphinemura	0.74	0.76	0.76
Anchyrtarsus	0.2	0.15	0.21
Ancylidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.27	0.24	0.4
Aspectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.06	0.05	0.03
Bezzia	0	0	0
Bitacomorpha	0	0.05	0
Blepharicera	0	0	0
Boyeria	0	0.05	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brillia	0.06	0.09	0
Caecidotea	0.05	0.05	0
Caenis	0.05	0.05	0
Calibaetis	0.05	0.05	0
Calopteryx	0.05	0.05	0
Cambardae	0.06	0.05	0.08
Centropilum	0.06	0.05	0.08
Ceratopogon	0.21	0.15	0.39
Chaetocladius	0.05	0.05	0
Chelifera	0.06	0.09	0.03
Cheumatopsyche	0.47	0.43	0.75
Chimarra	0.13	0.1	0.19
Chironomus	0	0	0
Chloroperlidae	0.06	0.14	0.03
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.18	0.24	0.15
Clinotanytarsus	0	0	0
Cloperla	0	0	0
Cnephia	0	0	0
Collembola	0.06	0.05	0.08
Conchapelopia	0.57	0.57	0.66
Cordulegaster	0.06	0.05	0.08
Cordulidae	0.05	0.05	0
Coriidae	0.05	0.05	0
Corydalis	0	0	0
Corynetura	0.28	0.29	0.49
Cranonyx	0.07	0.1	0.15
Cricotopus/Orthocladius	0.56	0.57	0.51
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Curia	0	0	0
Deronectes	0.05	0.05	0
Dianesa	0.25	0.2	0.21
Dicranota	0.11	0.19	0.03
Dicrolentipes	0	0	0
Dineutus	0	0	0
Dipheter	0.05	0.05	0
Diplectrona	0.33	0.38	0.53
Diplocladius	0	0	0
Diploperla	0.08	0.1	0.19
Dixa	0.05	0.05	0
Dolichopodidae	0.06	0.05	0.08
Dolophiodes	0.11	0.14	0
Drunella	0.11	0.09	0
Dubiraphia	0.06	0.05	0
Dugesia	0.05	0.05	0
Ecocoptera	0.07	0.05	0.12
Ectopria	0	0	0
Enallagma	0.05	0.05	0
Enchytraeidae	0.06	0.05	0.08
Endochironomus	0.05	0.09	0
Epeorus	0.24	0.28	0.2
Ephemera	0	0	0
Ephemerebella	0.7	0.71	0.78
Erioptera	0	0	0
Eukiefferiella	0.25	0.33	0.21
Eurylophella	0.21	0.15	0.42
Fallcon	0	0	0
Gammarus	0.07	0.05	0.15
Glossocoma	0.2	0.15	0.21
Gomphidae	0.21	0.15	0.35
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleniella	0	0	0
Helichus	0.05	0.05	0
Helius	0	0	0
Hemenodromia	0.33	0.29	0.43
Heptagenia	0	0	0
Heterotriisocladius	0.06	0.09	0.08
Hexagenia	0	0	0
Hexatoma	0.14	0.14	0.31
Hyalella	0	0	0
Hydrobaenus	0.11	0.1	0.03
Hydrobus	0.06	0.09	0.08
Hydroporus	0	0	0
Hydropsyche	0.27	0.29	0.4
Ironoxipia	0	0.05	0
Isonychia	0	0.05	0
Isoptera	0.13	0.14	0.11
Kiefferulus	0.05	0.05	0
Krenopelopia	0.07	0.05	0.12
Labridmia	0	0	0
Larsia	0	0	0
Lepidoptera	0.06	0.05	0.08
Lepidostoma	0.05	0.05	0

Leptophlebia	0.12	0.1	0.08
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0.05	0
Leuctridae	0.16	0.23	0
Limnophila	0	0.05	0
Limnophyes	0.07	0.05	0.12
Lopescladius	0	0	0
Lumbriculidae	0.18	0.14	0.08
Lype	0	0	0
Macronychus	0	0	0
Menetus	0	0	0
Menopelopia	0.13	0.14	0.19
Micrasema	0	0	0
Micropectra	0.52	0.44	0.76
Microtendipes	0.06	0.09	0.03
Microvelia	0	0	0
Molophilus	0	0	0
Musculium	0	0	0
Naididae	0.46	0.34	0.63
Nanocladius	0.05	0.05	0
Nematomorphia	0	0	0
Nemoura	0	0	0
Neophylax	0.64	0.57	0.78
Nigronia	0.19	0.24	0.19
Nixe	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.45	0.43	0.55
Ormosia	0	0	0
Orthocladinae(A)	0	0	0
Ostrocera	0	0	0
Oulimnius	0.58	0.62	0.74
Pagastia	0.08	0.05	0.21
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaetocladius	0	0.05	0
Parachironomus	0.05	0.05	0
Paragragnia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.15	0.19	0.33
Paramerina	0.05	0.05	0
Parametricnemus	0.64	0.67	0.73
Paraphaenocladus	0	0	0
Paratanytarsus	0.12	0.19	0.06
Pedicia	0	0	0
Pelodytes	0.05	0.05	0
Peltopera	0	0	0
Perlesta	0	0	0
Phanopsectra	0.06	0.05	0.03
Phaenoglyptora	0	0	0
Phyllocentropus	0	0	0
Physidae	0.12	0.14	0.08
Pisidium	0.08	0.05	0.21
Planorbella	0	0	0
Platycentropus	0	0	0
Platynittia	0.05	0.05	0
Polycentropus	0.28	0.24	0.55
Polyptelidum	0.38	0.34	0.43
Pothastia	0.08	0.05	0.19
Probezzia	0	0	0
Procladius	0	0	0
Promeseta	0	0	0
Prosimulium	0.68	0.67	0.6
Prostoia	0.32	0.33	0.38
Prostoma	0	0	0
Psephenus	0	0.05	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.13	0.1	0.19
Pseudorthocladius	0.06	0.05	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcyis	0.05	0.09	0
Ptilostomis	0	0	0
Pycnospsyche	0.07	0.1	0.12
Radix	0	0	0
Rheocricotopus	0	0.05	0
Rheotanytarsus	0.21	0.15	0.4
Rhyacophila	0.61	0.62	0.49
Serratella	0.24	0.24	0.21
Sialis	0.07	0.05	0.12
Simulium	0.3	0.28	0.21
Siphonurus	0	0	0
Siphonoplecton	0	0	0
Sphaeriidae	0.06	0.05	0.08
Sphaerium	0	0	0
Sphaeromias	0.28	0.24	0.53
StemPELLINA	0	0.05	0
StemPELLINELLA	0.06	0.05	0
Stenacron	0	0	0
Stenelmis	0.08	0.1	0.2
Stenonema	0.45	0.43	0.59
Stictochironomus	0.07	0.05	0.12
Stilocladius	0.06	0.05	0
Strophopteryx	0.19	0.15	0.18
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphothastia	0.08	0.14	0.19
Symnella	0	0	0
Tabus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytus	0	0	0
Tanytarsus	0.27	0.33	0.4
Thienemannella	0.19	0.24	0.22

Biodiversity for PBGH 108:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.1	0	0
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.61	0.67	0.65
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0	0
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.05	0	0
Cambardae	0	0	0
Capniidae	0.08	0.22	0.16
Carabidae?	0	0	0
Ceratopogonidae	0.09	0	0
Chironomidae	1	1	1
Chloroperlidae	0.08	0	0
Coenagrionidae	0	0	0
Collembola	0	0	0
Corbiculidae	0	0	0
Cordulegastriidae	0	0	0
Cordulidae	0	0	0
Corydidae	0	0	0
Corydalidae	0.2	0.11	0.15
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0	0	0
Dolichopodidae	0	0	0
Dryopidae	0.09	0	0
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.86	0.78	0.9
Empididae	0.58	0.78	0.75
Enchytraeidae	0	0	0
Ephemeroptera	0.95	1	1
Ephemeridae	0	0	0
Ephydriidae	0	0	0
Ergasilidae	0	0	0
Gammaridae	0.2	0	0.01
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.59	0.45	0.48
Gomphidae	0.09	0	0
Gordidae	0.06	0	0
Gyrinidae	0	0	0
Halplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.6	0.56	0.68
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.84	0.89	0.91
Hydroptilidae	0.11	0	0
Lepidoptera	0	0	0
Lepidostomatidae	0	0	0
Lepidoceridae	0	0	0
Leptophlebiidae	0.19	0.11	0.16
Lestidae	0	0	0
Leuctridae	0.04	0.11	0.08
Libellulidae	0	0	0
Limnephilidae	0.5	0.56	0.57
Lumbriculidae	0.12	0.22	0.2
Lymnaeidae	0.05	0	0
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metreopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.31	0.34	0.36
Nemouridae	0.76	0.67	0.7
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.14	0.11	0.12
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.04	0	0
Peridae	0.36	0.23	0.27
Periodidae	0.33	0.22	0.16
Phlopotamidae	0.32	0.34	0.41
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.11	0.11	0.15
Planariidae	0.06	0	0
Planorbidae	0	0	0
Pleuroceridae	0.05	0	0
Polycentropodidae	0.17	0.23	0.28
Potamanthidae?	0	0	0
Psephenidae	0.06	0	0
Psychodidae	0	0	0
Psychomyiidae	0.06	0	0
Pteronarcidae	0.04	0	0
Ptilodactylidae?	0.06	0.11	0.15
Ptychopleridae	0	0	0
Rhyacophilidae	0.47	0.45	0.48
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.93	0.89	0.85
Siphonuridae	0.41	0.67	0.67
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.09	0.11	0.08
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.82	0.78	0.85
Tubificidae	0.15	0.23	0.27

Ulenidae	0	0	0
Unionidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Genus:			
Ablabesmyia	0.06	0.07	0
Acentrella	0.11	0.07	0
Acerpenna	0.18	0.2	0.09
Acroneria	0.08	0.07	0.21
Agapetus	0	0	0
Allocampina	0.23	0.26	0.09
Allostypus	0	0	0
Ameletus	0.37	0.4	0.34
Amphinemura	0.73	0.73	0.58
Anchyrtarsus	0.15	0.14	0.29
Ancylidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.26	0.27	0.36
Apsectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.06	0.07	0
Bezzia	0	0	0
Bitacomorpha	0.06	0.07	0
Blepharicera	0	0	0
Boyeria	0.06	0.07	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brillia	0.05	0.06	0
Caecidotea	0.06	0.06	0
Caenis	0.06	0.06	0
Calibaetis	0.06	0.06	0
Calopteryx	0.06	0.06	0
Cambardae	0	0	0
Centropilum	0.06	0.07	0
Ceratopogon	0.2	0.2	0.29
Chaetocladius	0.05	0	0
Chelifera	0.07	0.07	0.13
Cheumatopsyche	0.46	0.4	0.71
Chimarra	0.12	0.13	0.14
Chironomus	0	0	0
Chloroperlidae	0.07	0.07	0.13
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.18	0.2	0.15
Clinotanytarsus	0	0	0
Cloperla	0	0	0
Cnephia	0	0	0
Collembola	0.06	0.07	0
Conchapelopia	0.57	0.54	0.69
Cordulegaster	0.06	0.07	0
Cordulidae	0.06	0.06	0
Coriidae	0.06	0.06	0
Corydalis	0	0	0
Corynetura	0.32	0.33	0.34
Crangonyx	0.13	0.13	0.15
CricotopusOrthocladius	0.52	0.54	0.66
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Curia	0	0	0
Deronectes	0.06	0.06	0
Dianesa	0.26	0.27	0.35
Dicranota	0.06	0.07	0
Dicrolentipes	0	0	0
Dineutus	0	0	0
Dipheter	0.06	0.07	0
Diplectrona	0.38	0.34	0.43
Diplocladius	0	0	0
Diploperla	0.07	0.07	0.14
Dixa	0	0	0
Dolichopodidae	0.06	0.07	0
Dolophiodes	0.11	0.07	0
Drunella	0.11	0.13	0
Dubiraphia	0.05	0.06	0
Dugesia	0	0	0
Ecoptera	0.07	0.07	0.09
Ectopria	0	0	0
Enallagma	0.06	0.06	0
Enchytraeidae	0.06	0.07	0
Endochironomus	0.06	0.06	0
Epeorus	0.24	0.2	0.2
Ephemerella	0.65	0.61	0.91
Erioptera	0	0	0
Eukiefferiella	0.25	0.27	0.28
Eurylophella	0.21	0.2	0.35
Fallcon	0	0	0
Gammarus	0.07	0.07	0.15
Glossocoma	0.19	0.2	0.21
Gomphidae	0.21	0.2	0.42
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleniella	0	0	0
Helichus	0.05	0	0
Helius	0	0	0
Hemenodromia	0.29	0.27	0.56
Heptagenia	0	0	0
Heterotrisocladius	0.12	0.13	0
Hexagenia	0	0	0
Hexatoma	0.14	0.14	0.23
Hyaella	0	0	0
Hydrobaenus	0.12	0.13	0.13
Hydrobus	0.06	0.07	0
Hydroporus	0	0	0
Hydropsyche	0.26	0.27	0.36
Ironoxipia	0	0	0
Isonychia	0	0	0
Isoptera	0.12	0.13	0
Kiefferulus	0.06	0.06	0
Krenopelopia	0.07	0.07	0.09
Labridmia	0	0	0
Larsia	0	0	0
Lepidoptera	0.06	0.07	0
Lepidostoma	0	0	0

Leptophlebia	0.12	0.13	0
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucocrota	0	0	0
Leuctridae	0.16	0.13	0
Limnophila	0.06	0.07	0
Limnophyes	0.07	0.07	0.09
Lopescladius	0	0	0
Lumbriculidae	0.17	0.2	0
Lype	0	0	0
Macronychus	0	0	0
Menetus	0	0	0
Menopelopia	0.13	0.13	0.09
Micrasema	0	0	0
Micropectra	0.46	0.47	0.58
Microtendipes	0.07	0.07	0.13
Microvelia	0	0	0
Molophilus	0	0	0
Musculium	0	0	0
Naididae	0.39	0.34	0.56
Nanocladius	0	0	0
Nematomorphia	0	0	0
Nemoura	0	0	0
Neophylax	0.57	0.53	0.71
Nigronia	0.18	0.2	0.09
Nixe	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.39	0.34	0.58
Ormosia	0	0	0
Orthocladinae(A)	0	0	0
Ostrocera	0	0	0
Ouilimnius	0.53	0.47	0.79
Pagastia	0.08	0.07	0.21
Palaemonetes	0	0	0
Paracapia	0	0	0
Parachaetocladius	0	0	0
Parachironomus	0.06	0.06	0
Paragragnia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.15	0.14	0.3
Paramerina	0.06	0.06	0
Parametricnemus	0.63	0.67	0.72
Paraphaenocladus	0	0	0
Paratanytarsus	0.13	0.13	0.13
Pedicia	0	0	0
Pelodytes	0.06	0.06	0
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.07	0.07	0.13
Phaenopsectora	0	0	0
Phyllocentropus	0	0	0
Physidae	0.12	0.13	0
Pisidium	0.08	0.07	0.21
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0	0
Polycentropus	0.28	0.27	0.5
Polypedilum	0.32	0.33	0.36
Pothastia	0.07	0.07	0.14
Probezzia	0	0	0
Procladius	0	0	0
Pronotera	0	0	0
Prosimulium	0.63	0.6	0.64
Prostoia	0.32	0.33	0.35
Prostoma	0	0	0
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.13	0.13	0.09
Pseudorthocladius	0.05	0.06	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcyis	0	0	0
Ptilostomis	0	0	0
Pycnosyche	0.12	0.13	0.09
Radix	0	0	0
Rheocricotopus	0.06	0.07	0
Rheotanytarsus	0.2	0.2	0.36
Rhyacophila	0.57	0.54	0.64
Serratella	0.19	0.13	0.21
Sialis	0.07	0.07	0.09
Simulium	0.3	0.26	0.21
Siphonurus	0	0	0
Siphloplecton	0	0	0
Sphaeriidae	0.06	0.07	0
Sphaerium	0	0	0
Sphaeromias	0	0	0
Siegesoptera	0.32	0.34	0.38
Stempellina	0	0	0
Stempellinella	0.05	0.06	0
Stenacron	0	0	0
Stenelmis	0.08	0.07	0.2
Stenonema	0.44	0.4	0.44
Stictochironomus	0.07	0.07	0.09
Stilocladius	0.05	0.06	0
Strophopteryx	0.18	0.2	0.15
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphothaenia	0.07	0.07	0.14
Symulium	0	0	0
Tabus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytarsus	0.26	0.27	0.36
Thienemannella	0.13	0.13	0.14
Thienemannimyia(gp)	0.06	0.07	0
Tipula	0.41	0.4	0.24
Trepobates	0	0	0
Trienodes	0.06	0.07	0
Tribelos	0.06	0.07	0
Trisoptelopia			

Biodiversity for PBGS111:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TRCI AIC	Tubificidae	0.15	0.22	0.27	Lepidostoma	0.06	0.04	0
Aeshnidae	0	0	0	Uenoidae	0	0	0	Lepidostomia	0.06	0.08	0.09
Ameletidae	0	0	0	Unionidae	0	0	0	Leptoxis	0	0	0
Ancylidae	0	0	0	Valvatidae	0	0	0	Leucotrichia	0	0	0
Asellidae	0.09	0	0	Velidae	0	0	0	Leucrocrua	0	0.04	0
Astacidae	0	0	0	Viviparidae	0	0	0	Leuctridae	0.17	0.16	0.01
Athericidae	0	0	0					Limnophila	0	0	0
Baetidae	0.62	0.67	0.63					Limnophyes	0.05	0.04	0
Baetiscidae	0	0	0					Lopescladius	0	0	0
Bekstomatidae	0	0	0					Lumbriculidae	0.17	0.16	0.15
Bithyniidae	0	0	0					Lype	0	0	0
Blephariceridae	0	0	0					Macronychus	0	0	0
Brachycentridae	0.04	0	0					Menetus	0	0	0
Braconidae	0	0	0					Metopeloplia	0.11	0.12	0.09
Branchiobdellidae	0	0	0					Micrasema	0	0	0
Caeniidae	0	0	0					Microsectra	0.54	0.41	0.69
Calamoceratidae	0	0	0					Microtendipes	0.05	0.08	0
Calopterygidae	0.05	0	0					Microvelia	0	0	0
Cambariae	0	0	0					Molophilus	0	0	0
Capniidae	0.08	0.22	0.18					Muscullum	0	0	0
Carabidae?	0	0	0					Naididae	0.51	0.29	0.79
Ceratopogonidae	0.05	0	0					Nanocladius	0.06	0.08	0
Chironomidae	1	1	1					Nematomorpha	0	0	0
Chloroperlidae	0.08	0	0					Nemoura	0	0	0
Coenagrionidae	0	0	0					Neophylax	0.73	0.53	0.89
Collembola	0	0	0					Nigronia	0.21	0.2	0.09
Corbiculidae	0	0	0					Nixe	0	0	0
Condulegastridae	0	0	0					Oecetis	0	0.04	0
Conuliidae	0	0	0					Oemopteryx	0	0	0
Corixidae	0	0	0					Optioservus	0.52	0.48	0.46
Corydidae	0.2	0.11	0.12					Ormosia	0	0	0
Crangonyctidae	0	0	0					Orthoclaudiinae(A)	0	0.04	0
Culiidae	0	0	0					Ostrocerca	0	0	0
Curculionidae	0	0	0					Oulimnius	0.6	0.6	0.58
Diplostroca	0	0	0					Pagastia	0.07	0.04	0.19
Dividae	0	0	0					Palaeomonetes	0	0	0
Dolichopodidae	0	0	0					Paracapnia	0	0	0
Dryopidae	0.09	0	0.01					Parachaeoclaadius	0	0.04	0
Dugesidae	0	0	0					Parachironomus	0	0.04	0
Dytiscidae	0.87	0.89	0.82					Paragnetina	0	0	0
Eimidae	0.87	0.87	0.86					Parakiefferiella	0	0	0
Enchytraeidae	0	0	0					Paraleptophlebia	0.18	0.2	0.19
Ephemereidae	0.96	1	0.99					Paramerina	0	0.04	0
Ephemeridae	0	0	0					Parameletrocnemus	0.69	0.6	0.68
Ephydriidae	0	0	0					Paraphaenoclaadius	0	0.12	0.18
Eriophlebiidae	0	0	0					Paratanytarsus	0.12	0.2	0.18
Gammaridae	0.22	0.05	0.05					Pedicia	0	0	0
Gerridae	0	0	0					Pelodytes	0	0.04	0
Glossiphoniidae	0	0	0					Peltopera	0	0	0
Glossosomatidae	0.39	0.55	0.5					Perlesta	0	0	0
Gomphidae	0.08	0	0					Phaenopspectra	0.05	0.04	0
Gordiiidae	0.05	0	0					Phasganophora	0	0	0
Gyrinidae	0	0	0					Phloeocentropus	0	0	0
Halplidae	0	0	0					Physida	0.11	0.12	0.09
Helodidae	0	0	0					Pisidium	0.07	0.04	0.19
Hepageniidae	0.57	0.67	0.68					Planorbella	0	0	0
Hydrobiidae	0	0	0					Platycentropus	0	0.04	0
Hydrophilidae	0.04	0	0					Platysmittia	0.06	0.04	0
Hydropsychidae	0.84	0.89	0.86					Polycentropus	0.3	0.2	0.4
Hydroptilidae	0.1	0	0					Polypedium	0.43	0.29	0.64
Lepidoptera	0	0	0					Pothastia	0.12	0.04	0.21
Lepidostomatidae	0	0	0					Probezzia	0	0.04	0
Leptoceridae	0	0	0					Procladius	0	0.04	0
Lepidophlebiidae	0.19	0.11	0.19					Promoesia	0	0.04	0
Lesidae	0	0	0					Prosimulium	0.71	0.68	0.6
Leuctridae	0.04	0.11	0.08					Prostoma	0.37	0.32	0.49
Libellulidae	0	0	0					Psephenus	0	0.04	0
Limnephilidae	0.51	0.56	0.54					Pseudochironomus	0	0	0
Lumbriculidae	0.13	0.22	0.2					Pseudolimnophila	0.11	0.08	0.09
Lymnaeidae	0.04	0	0					Pseudorthoclaadius	0.06	0.04	0
Macroniidae	0	0	0					Pseudosuccinea	0	0	0
Maxillopoda	0	0	0					Ptilotreta	0	0	0
Metreopodidae	0	0	0					Pteronarcy	0.06	0.12	0
Molannidae	0	0	0					Ptilostomis	0	0	0
Muscidae	0	0	0					Pycnopsyche	0.05	0.08	0
Naididae	0.27	0.34	0.33					Radix	0	0	0
Nemouridae	0.75	0.78	0.72					Rheocricotopus	0	0.04	0
Nepidae	0	0	0					Rheotanytarsus	0.24	0.2	0.4
Notonectidae	0	0	0					Rhyacophila	0.63	0.56	0.28
Odonoceridae	0	0	0					Serratella	0.28	0.2	0.19
Oligoneuridae	0.14	0.11	0.11					Sialis	0.05	0.04	0
Ostracoda	0	0	0					Simulium	0.3	0.28	0.25
Palaemonidae	0	0	0					Siphonurus	0	0	0
Peltoperidae	0.04	0	0.01					Siphonocleon	0	0	0
Perflidae	0.35	0.33	0.31					Sphaeriidae	0.06	0.04	0.09
Perlodidae	0.32	0.11	0.13					Sphaerium	0	0.04	0
Philopotamidae	0.32	0.44	0.44					Sphaeromias	0	0	0
Phyganeidae	0	0	0					Stagopterna	0.28	0.2	0.31
Physidae	0.04	0	0					Stempellina	0	0.04	0
Pisicidae	0	0	0					Stempellinella	0.06	0.04	0
Pisidiidae	0.11	0.11	0.13					Stenacron	0	0	0
Planariidae	0.05	0	0					Steneimis	0.14	0.12	0.21
Planorbidae	0	0	0					Stenonema	0.46	0.44	0.55
Pleuroceridae	0.05	0	0					Stictochironomus	0.05	0.04	0
Polycentropodidae	0.22	0.23	0.25					Stiloclaadius	0.06	0.04	0.04
Potamanitidae?	0	0	0					Strophopteryx	0.18	0.12	0.22
Psephenidae	0.05	0	0					Stygobromus	0	0	0
Psychodidae	0	0	0					Symposiocladius	0	0	0
Psychomyiidae	0.05	0	0					Symphylastia	0.12	0.16	0.21
Pteronarcidae	0.04	0	0.01					Synurella	0	0	0
Ptilodactylidae?	0.07	0.11	0.13					Tabanus	0	0	0
Psychopteridae	0	0	0					Taeniopteryx	0	0	0
Rhacophyllidae	0.47	0.44	0.44					Tallaperla	0	0.04	0
Saldidae	0	0	0					Tanytarsus	0	0	0
Sciomyzidae	0	0	0					Tanytarsus	0.35	0.32	0.46
Sericostomatidae	0	0	0					Thienemannella	0.24	0.24	0.4
Sialidae	0	0	0					Thienemannimyia(grp)	0.07	0.04	0.18
Simuliidae	0.83	0.89	0.88					Tipula	0.39	0.32	0.2
Siphonuridae	0.38	0.67	0.62					Trepobates	0	0	0
Sisyridae	0	0	0					Trianaodes	0	0.04	0
Sphaeriidae	0	0	0					Tribelios	0	0	0
Stratiomyidae	0	0	0					Trisopelopia	0.35	0.21	0.45
Syrphidae	0	0	0					Tubificidae	0.12	0.12	0.21
Tabanidae	0	0	0					Tvetenia	0.12	0.04	0.21
Taeniopterygidae	0.13	0.11	0.1					Unniella	0	0	0
Taliridae	0	0	0					Wormaldia	0	0	0
Tetrastemmatidae	0	0	0					Xyotopus	0	0	0
Tipulidae	0.86	0.89	0.9					Zavreimyia	0.31	0.17	0.52

Biodiversity for PBPB 305D:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.1	0.16	0.16
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.58	0.62	0.58
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0.08	0.09
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.06	0	0
Cambaridae	0	0	0
Capniidae	0	0	0
Carabidae?	0	0	0
Ceratopogonidae	0.1	0.07	0.08
Chironomidae	1	1	1
Chloroperlidae	0.08	0.07	0.05
Coenagrionidae	0	0	0
Collembola	0	0	0
Corticulidae	0	0	0
Cordulegastriidae	0	0	0
Cordulidae	0	0	0
Corixidae	0	0	0
Corydalidae	0.19	0.23	0.34
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostroca	0	0	0
Dixidae	0.04	0	0
Dolichopodidae	0	0	0
Dryopidae	0.14	0.08	0.09
Dugesidae	0	0	0
Dytiscidae	0	0	0
Eimidae	0.9	0.92	0.9
Empididae	0.58	0.77	0.58
Enchytraeidae	0	0	0
Ephemereidae	0.95	0.92	0.91
Ephemeridae	0	0	0
Ephyridae	0	0	0
Erboddelidae	0	0	0
Gammaridae	0.23	0.16	0.16
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.31	0.38	0.21
Gomphidae	0.17	0.23	0.22
Gordidae	0.05	0.08	0.07
Gyrinidae	0	0	0
Haliplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.62	0.7	0.81
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.83	0.85	0.9
Hydroptilidae	0.11	0.08	0.07
Lepidoptera	0	0	0
Lepidostomatidae	0.04	0	0
Lepidocnemeidae	0	0	0
Leptophlebiidae	0.13	0.08	0.09
Lestidae	0	0	0
Leuctridae	0.08	0	0
Libellulidae	0	0	0
Limnephilidae	0.51	0.39	0.48
Lumbriculidae	0.05	0.08	0
Lymnaeidae	0.05	0.08	0.09
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.38	0.31	0.47
Nemouridae	0.83	0.69	0.76
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.14	0.15	0
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.05	0	0
Perlidae	0.36	0.31	0.25
Perlodidae	0.39	0.24	0.41
Philopotamidae	0.31	0.32	0.55
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.09	0.16	0.18
Planorbidae	0.1	0.08	0.07
Planorbidae	0	0	0
Pleuroceridae	0.06	0	0
Polycentropodidae	0.24	0.16	0.31
Potamanthidae?	0	0	0
Psephenidae	0.05	0.08	0.07
Psychodidae	0	0	0
Psychomyiidae	0.05	0.08	0.07
Pteronarcidae	0.09	0	0
Ptilodactylidae?	0.13	0.15	0.13
Ptychopteridae	0	0	0
Rhyacophilidae	0.58	0.46	0.43
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.92	0.75
Siphonuridae	0.29	0.38	0.3
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.09	0	0
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.79	0.62	0.73
Tubificidae	0.1	0.16	0.22

Genus:	Ragan ANNA AIC	Ragan BROI AIC	Ragan TROI AIC
Uenoidae	0	0	0
Unionidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Ablabesmyia	0	0.04	0
Acetrotella	0.1	0.13	0
Acerpenna	0.16	0.18	0
Acroneria	0.07	0.09	0.17
Agapetus	0	0.04	0
Allocapnia	0.21	0.18	0
Alotrypys	0	0.04	0
Ametetus	0.32	0.36	0.33
Amphinemura	0.7	0.68	0.37
Anchytarsus	0.21	0.14	0.36
Ancylidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.25	0.23	0.35
Apsectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.05	0.05	0
Bezzia	0	0	0
Bitacocomorpha	0.05	0.04	0
Blepharicera	0	0	0
Boyeria	0.05	0.04	0
Brachycentrus	0	0.04	0
Branchiobdellidae	0	0	0
Brillia	0.05	0.13	0
Caecidotea	0.06	0.13	0.07
Caenis	0.06	0.09	0.07
Callibaetis	0.06	0.05	0.07
Calopteryx	0.06	0.05	0.07
Cambaridae	0	0.04	0
Centroptilum	0	0.04	0
Ceratopogon	0.15	0.14	0.2
Chaetocladius	0.05	0.04	0
Chellifera	0.09	0.05	0.2
Cheumatopsyche	0.46	0.46	0.55
Chimarra	0.12	0.09	0.13
Chironomus	0	0.04	0
Chloroperlidae	0.09	0.09	0.2
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.16	0.22	0
Clinotanytarsus	0	0	0
Cloperia	0	0	0
Cnephia	0	0	0
Collembola	0	0.04	0
Conchapelopia	0.64	0.64	0.77
Cordulegaster	0	0.04	0
Cordulidae	0.06	0.05	0.07
Corixidae	0.06	0.05	0.07
Corydatus	0	0	0
Corynoneura	0.26	0.27	0.33
Crangonyx	0.11	0.09	0
Cricotopus/Orthocladius	0.57	0.63	0.6
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Cura	0	0	0
Deronectes	0.06	0.05	0.07
Diamesa	0.28	0.19	0.38
Dicranota	0.1	0.13	0
Dicrolenidipes	0	0.04	0
Dineutus	0	0	0
Diphetor	0.06	0.05	0.04
Diplectrona	0.37	0.32	0.33
Diplocladius	0	0.04	0
Diploperla	0.07	0.09	0.13
Dixa	0.05	0.04	0
Dolichopodidae	0	0.04	0
Dolophilodes	0.11	0.09	0.04
Drunella	0.1	0.09	0
Dubiraphia	0.05	0.05	0
Dugesia	0.05	0.04	0
Ecopotura	0.05	0.04	0
Ectopria	0	0	0
Enallagma	0.06	0.05	0.07
Enchytraeidae	0	0.09	0
Endochironomus	0.06	0.09	0.07
Epeorus	0.26	0.23	0.24
Ephemera	0.73	0.64	0.93
Ephemerella	0	0	0
Erioptera	0	0	0
Eukiefferiella	0.25	0.32	0.2
Eurylophella	0.15	0.14	0.2
Fallicon	0	0	0
Gammarus	0.06	0.09	0
Glossosoma	0.18	0.14	0.17
Gomphidae	0.23	0.19	0.39
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleliella	0	0	0
Helichus	0.05	0.09	0
Hellus	0	0	0
Hemerodromia	0.36	0.28	0.69
Heptagenia	0	0	0
Heterotrissocladius	0.05	0.09	0
Hexagenia	0	0	0
Hexatoma	0.12	0.14	0.13
Hyalella	0	0	0
Hydrobaenus	0.15	0.09	0.26
Hydrobius	0	0.09	0
Hydroporus	0	0	0
Hydropsyche	0.26	0.32	0.35
Ironoquia	0.04	0.04	0
Isonychia	0	0.04	0
Isoperla	0.05	0.13	0
Kiefferulus	0.06	0.05	0.07
Krenopelopia	0.05	0.04	0
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0	0.04	0
Lepidostoma	0.05	0.04	0

Leptophlebia	0.06	0.09	0.07
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.21	0.18	0
Limnophila	0.05	0.04	0
Limnophyes	0.05	0.04	0
Lopescladius	0	0	0
Lumbriculidae	0.1	0.18	0
Lype	0	0	0
Macronychus	0	0.04	0
Menetus	0	0.04	0
Menopelopia	0.05	0.09	0
Microasema	0	0	0
Micropsocra	0.44	0.37	0.37
Microtendipes	0.09	0.09	0.2
Microvelia	0	0	0
Molophilus	0	0	0
Muscillum	0	0.04	0
Naididae	0.44	0.36	0.51
Nanocladius	0.05	0.04	0
Nematomorpha	0	0	0
Nemoura	0	0	0
Neophylax	0.61	0.55	0.51
Nigronia	0.11	0.22	0.04
Nixa	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.44	0.41	0.49
Ormosia	0	0	0
Orthoclaudiinae(A)	0	0	0
Ostracerca	0	0	0
Oulimnius	0.58	0.55	0.61
Pagastia	0.07	0.05	0.17
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaetocladius	0	0	0
Parachironomus	0.06	0.05	0.07
Paragnetia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.13	0.14	0.17
Paramerina	0.06	0.05	0.07
Parametrocnemus	0.65	0.63	0.5
Paraphaenocladus	0	0.04	0
Paratanytarsus	0.14	0.22	0.2
Pedicia	0	0.04	0
Pelodytes	0.06	0.09	0.07
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsocra	0.09	0.05	0.2
Phaenoglyphora	0	0	0
Phycocentropus	0	0	0
Physidae	0.06	0.14	0.07
Pisidium	0.07	0.09	0.17
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0.05	0.04	0
Polycentropus	0.3	0.23	0.3
Polypedium	0.35	0.32	0.3
Pothastia	0.07	0.05	0.13
Probozia	0	0	0
Procladius	0	0.04	0
Promesia	0	0	0
Prosimulium	0.63	0.58	0.58
Prostia	0.31	0.36	0.2
Prostoma	0	0	0
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.05	0.09	0
Pseudorthocladius	0.05	0.04	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0.05	0.04	0
Ptilostomus	0	0	0
Pycnopsyche	0.1	0.09	0
Radix	0	0.04	0
Rheocricotopus	0.05	0.09	0
Rheotanytarsus	0.19	0.18	0.3
Rhyacophila	0.64	0.55	0.63
Serratella	0.24	0.23	0.22
Sialis	0.05	0.04	0
Simulium	0.33	0.31	0.17
Siphonurus	0	0	0
Siphloplecton	0	0	0
Sphaeriidae	0	0.04	0
Sphaerium	0	0.04	0
Sphaeromias	0	0	0
Stegopterna	0.23	0.23	0.13
Stempellina	0	0	0
Stempellinella	0.05	0.04	0
Stenacron	0	0	0
Stenelmis	0.09	0.14	0.22
Stenonema	0.36	0.41	0.35
Stictochironomus	0.05	0.04	0
Stilocladius	0.05	0.04	0
Strophopteryx	0.16	0.14	0
Stygobromus	0	0	0
Symposiocladus	0	0	0
Symphothasta	0.07	0.09	0.13
Synurella	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytarsus	0	0.04	0
Tanytarsus	0.3	0.32	0

Biodiversity for PBPB 305E:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.1	0.16	0.16
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.58	0.62	0.58
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0.08	0.09
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.06	0	0
Cambidae	0	0	0
Capniidae	0	0	0
Carabidae?	0	0	0
Ceratopogonidae	0.1	0.07	0.08
Chironomidae	1	1	1
Chloroperlidae	0.08	0.07	0.05
Coenagrionidae	0	0	0
Collembola	0	0	0
Corbiculidae	0	0	0
Cortulegastriidae	0	0	0
Cordulidae	0	0	0
Cortixidae	0	0	0
Corydalidae	0.19	0.23	0.24
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostrota	0	0	0
Dixidae	0.04	0	0
Dolichopodidae	0	0	0
Dryopidae	0.14	0.08	0.09
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.9	0.92	0.9
Empididae	0.59	0.77	0.58
Enchytraeidae	0	0	0
Ephemeroptera	0.95	0.92	0.91
Ephemeridae	0	0	0
Ephydriidae	0	0	0
Ergasilidae	0	0	0
Gammaridae	0.23	0.16	0.16
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.31	0.38	0.21
Gomphidae	0.17	0.23	0.22
Gordidae	0.05	0.08	0.07
Gyrinidae	0	0	0
Haliplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.62	0.7	0.81
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.83	0.85	0.91
Hydroptilidae	0.11	0.08	0.07
Lepidoptera	0	0	0
Lepidostomatidae	0.04	0	0
Lepidostomatidae	0.13	0.08	0.09
Lestidae	0	0	0
Leuctridae	0.08	0	0
Libellulidae	0	0	0
Limnephilidae	0.51	0.39	0.48
Lumbriculidae	0.05	0.08	0
Lymnaeidae	0.05	0.08	0.09
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.38	0.31	0.47
Nemouridae	0.83	0.69	0.76
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.14	0.15	0
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.05	0	0
Perlidae	0.36	0.31	0.25
Perlodidae	0.39	0.24	0.41
Philopotamidae	0.32	0.32	0.55
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.09	0.16	0.18
Planariidae	0.1	0.08	0.07
Planorbidae	0	0	0
Pleuroceridae	0.06	0	0
Polycentropodidae	0.24	0.16	0.31
Potamanthidae?	0	0	0
Psephenidae	0.05	0.08	0.07
Psychodidae	0	0	0
Psychomyiidae	0.05	0.08	0.07
Pteronarcidae	0.09	0	0
Pilodactylidae?	0.13	0.15	0.13
Ptychopteridae	0	0	0
Rhycophilidae	0.58	0.46	0.43
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.92	0.75
Siphonuridae	0.29	0.38	0.31
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.09	0	0
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.79	0.62	0.73
Tubificidae	0.1	0.16	0.22

Ulenidae	0	0	0
Umonidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Genus:			
Ablabesmyia	0	0.04	0
Acentrella	0.1	0.13	0
Acerpenna	0.16	0.18	0
Acroneria	0.07	0.09	0.18
Agapetus	0	0.04	0
Allocapnia	0.21	0.18	0
Alotanytus	0	0.04	0
Ameletus	0.32	0.36	0.33
Amphinemura	0.7	0.68	0.38
Anchyrtarus	0.21	0.14	0.36
Ancylidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.25	0.23	0.35
Aspectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.05	0.05	0
Bezzia	0	0	0
Bitacomorpha	0.05	0.04	0
Blepharicera	0	0	0
Boyeria	0.05	0.04	0
Brachycentrus	0	0.04	0
Branchiobdellidae	0	0	0
Brillia	0.05	0.13	0
Caecidotea	0.06	0.13	0.06
Caenis	0.06	0.09	0.06
Calibaetis	0.06	0.05	0.06
Calopteryx	0.06	0.05	0.06
Cambariae	0	0.04	0
Centropilum	0	0.04	0
Ceratopogon	0.15	0.14	0.2
Chaetocladius	0.05	0.04	0
Chelifera	0.09	0.05	0.2
Cheumatopsyche	0.46	0.46	0.55
Chimarra	0.12	0.09	0.13
Chironomus	0	0.04	0
Chloroperlidae	0.09	0.09	0.2
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.16	0.22	0
Clinotanytarsus	0	0	0
Cloperla	0	0	0
Cnephia	0	0	0
Collembola	0	0.04	0
Conchapelopia	0.64	0.64	0.77
Cordulegaster	0	0.04	0
Cordulidae	0.06	0.05	0.06
Cortixidae	0.06	0.05	0.06
Corydalus	0	0	0
Corynoneura	0.26	0.27	0.33
Crangonyx	0.11	0.09	0
CricotopusOrthocladius	0.57	0.63	0.6
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Curia	0	0	0
Deronectes	0.06	0.05	0.06
Dianesa	0.28	0.19	0.38
Dicranota	0.1	0.13	0
Dicrolentipes	0	0.04	0
Dineutus	0	0	0
Diphotor	0.06	0.05	0.04
Diplectrona	0.37	0.32	0.33
Diplocladius	0	0.04	0
Diploperla	0.07	0.09	0.13
Dixa	0.05	0.04	0
Dolichopodidae	0	0.04	0
Dolophilodes	0.11	0.09	0.04
Drunella	0.11	0.09	0
Dubiraphia	0.05	0.05	0
Dugesia	0.05	0.04	0
Ecocoptera	0.05	0.04	0
Ectoptera	0	0	0
Enallagma	0.06	0.05	0.06
Enchytraeidae	0	0.09	0
Endochironomus	0.06	0.09	0.06
Epeorus	0.26	0.23	0.24
Ephemerella	0.73	0.64	0.94
Erioptera	0	0	0
Eukiefferiella	0.25	0.32	0.2
Euryophella	0.15	0.14	0.2
Fallcon	0	0	0
Gammarus	0.06	0.09	0
Glossocoma	0.18	0.14	0.18
Gomphidae	0.23	0.19	0.39
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleniella	0	0	0
Helichus	0.05	0.09	0
Helius	0	0	0
Hemenodromia	0.36	0.28	0.69
Heptagenia	0	0	0
Heterotrissocladius	0.05	0.09	0
Hexagenia	0	0	0
Hexatoma	0.12	0.14	0.13
Hyaella	0	0	0
Hydrobaenus	0.15	0.09	0.26
Hydrobius	0	0.09	0
Hydroporus	0	0	0
Hydropsyche	0.26	0.32	0.35
Ironoxipia	0	0.04	0
Isonychia	0	0.04	0
Isoptera	0.05	0.13	0
Kiefferulus	0.06	0.05	0.06
Krenopelopia	0.05	0.04	0
Labridmia	0	0	0
Larsia	0	0	0
Lepidoptera	0	0.04	0
Lepidostoma	0.05	0.04	0

Leptophlebia	0.06	0.09	0.06
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.21	0.18	0
Limnophila	0.05	0.04	0
Limnophyes	0.05	0.04	0
Lopescladius	0	0	0
Lumbriculidae	0.11	0.18	0
Lype	0	0	0
Macronychus	0	0.04	0
Menetus	0	0.04	0
Menopelopia	0.05	0.09	0
Micrasema	0	0	0
Micropectra	0.44	0.37	0.38
Microtendipes	0.09	0.09	0.2
Microvelia	0	0	0
Molophilus	0	0	0
Musculium	0	0.04	0
Naididae	0.44	0.36	0.51
Nanocladius	0.05	0.04	0
Nematomorphia	0	0	0
Nemoura	0	0	0
Neophylax	0.61	0.55	0.51
Nigronia	0.11	0.22	0.04
Nixe	0	0	0
Oecetis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.44	0.41	0.49
Ormosia	0	0	0
Orthocladinae(A)	0	0	0
Ostrocera	0	0	0
Oulimnius	0.58	0.55	0.62
Pagastia	0.07	0.05	0.18
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaetocladius	0	0	0
Parachironomus	0.06	0.05	0.06
Paragrania	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.13	0.14	0.18
Paramerina	0.06	0.05	0.06
Parametricnemus	0.65	0.63	0.5
Paraphaenocladus	0	0.04	0
Paratanytarsus	0.14	0.22	0.2
Pedicia	0	0.04	0
Pelodytes	0.06	0.09	0.06
Peltopera	0	0	0
Perlesta	0	0	0
Phanopspectra	0.09	0.05	0.2
Phaenoglyptera	0	0	0
Phyllocentropus	0	0	0
Physidae	0.06	0.14	0.06
Pisidium	0.07	0.09	0.18
Planorbella	0	0	0
Platycentropus	0	0	0
Platymitria	0.05	0.04	0
Polycentropus	0.31	0.23	0.31
Polyptelidum	0.35	0.32	0.31
Pothastia	0.07	0.05	0.13
Probezzia	0	0	0
Procladius	0	0.04	0
Prometesia	0	0	0
Prosimulium	0.63	0.59	0.58
Prostoma	0.31	0.36	0.2
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.05	0.09	0
Pseudorthocladius	0.05	0.04	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0.05	0.04	0
Ptilostomis	0	0	0
Ptychopsyche	0.1	0.09	0
Radix	0	0.04	0
Rheocricotopus	0.05	0.09	0
Rheotanytarsus	0.19	0.18	0.31
Rhyacophila	0.64	0.55	0.63
Serratella	0.24	0.23	0.22
Sialis	0.05	0.04	0
Simulium	0.34	0.31	0.18
Siphonurus	0	0	0
Siphonoplecton	0	0	0
Sphaeriidae	0	0.04	0
Sphaerium	0	0.04	0
Sphaeromias	0	0	0
Siegesoptera	0.23	0.23	0.13
Stempellina	0	0	0
Stempellinella	0.05	0.04	0
Stenacron	0	0	0
Stenelmis	0.09	0.14	0.2
Stenonema	0.36	0.41	0.35
Stictochironomus	0.05	0.04	0
Stilocladius	0.05	0.04	0
Strophopteryx	0.16	0.14	0
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphothastia	0.07	0.09	0.13
Symrella	0	0	0
Tabus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytarsus	0	0.04	0
Tanytarsus	0.3	0.32	0.31
Thienemannella	0.17	0.23	0.13
Thienemannimyia(gp)	0.05	0.05	0

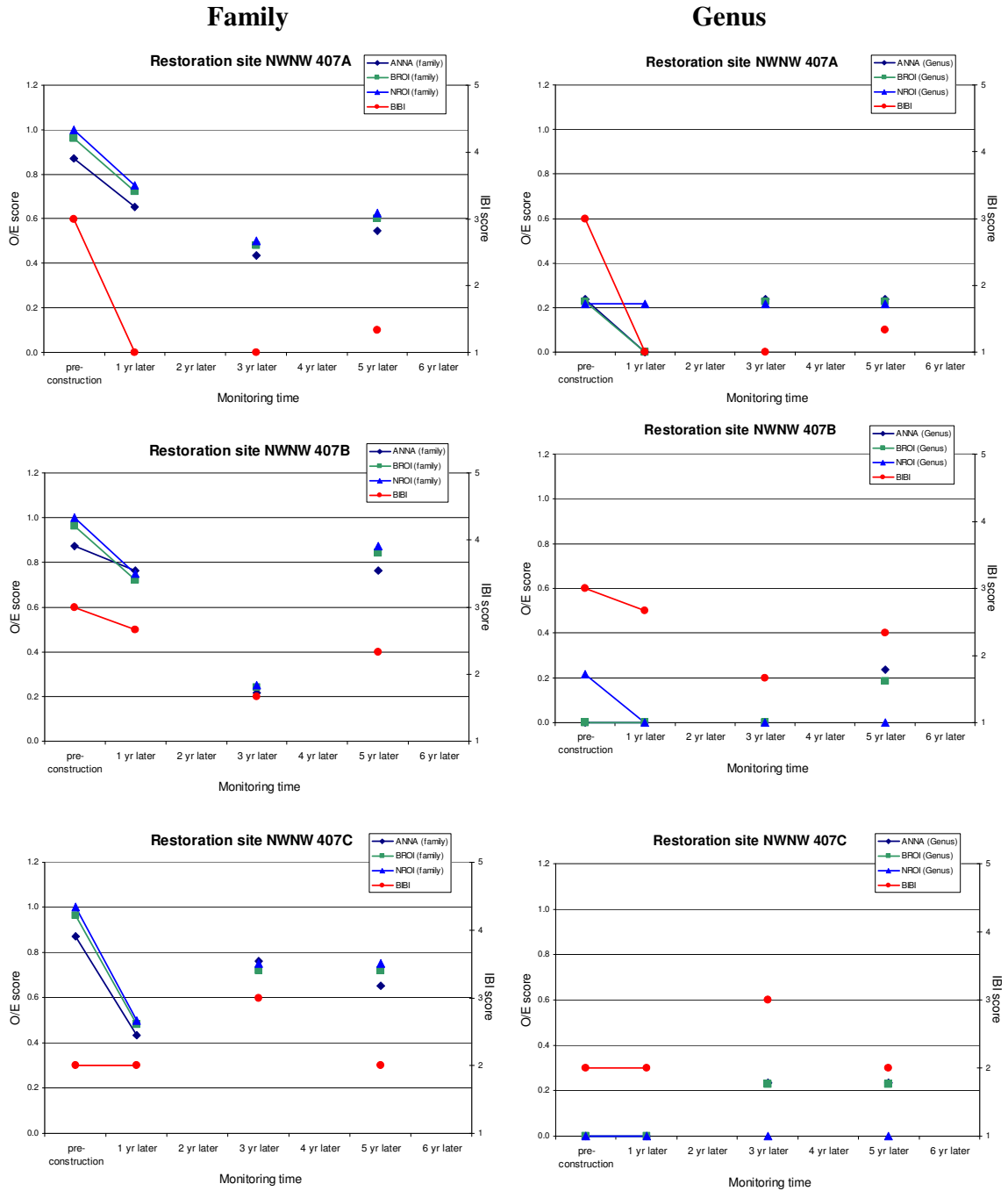
Biodiversity for PBPB 305F:

Family:	Ragan ANNA BIC	Ragan BROI StepwiseAIC	Ragan TROI AIC
Aeshnidae	0	0	0
Ameletidae	0	0	0
Ancylidae	0	0	0
Asellidae	0.11	0.22	0.16
Astacidae	0	0	0
Athericidae	0	0	0
Baetidae	0.63	0.67	0.67
Baetiscidae	0	0	0
Belostomatidae	0	0	0
Bithyniidae	0	0	0
Blephariceridae	0	0	0
Brachycentridae	0.05	0.11	0.08
Braconidae	0	0	0
Branchiobdellidae	0	0	0
Caenidae	0	0	0
Calamoceratidae	0	0	0
Calopterygidae	0.05	0	0
Cambardae	0	0	0
Capniidae	0	0	0
Carabidae?	0	0	0
Ceratopogonidae	0.1	0	0.01
Chironomidae	1	1	1
Chloroperlidae	0.08	0	0.03
Coenagrionidae	0	0	0
Collembola	0	0	0
Corbiculidae	0	0	0
Cortulegastriidae	0	0	0
Cordulidae	0	0	0
Corixidae	0	0	0
Corydalidae	0.2	0.22	0.33
Crangonyctidae	0	0	0
Culicidae	0	0	0
Curculionidae	0	0	0
Diplostrota	0	0	0
Dixidae	0.04	0	0
Dolichopodidae	0	0	0
Dryopidae	0.14	0.11	0.08
Dugesidae	0	0	0
Dytiscidae	0	0	0
Embiidae	0.9	0.89	0.88
Empididae	0.6	0.78	0.66
Enchytraeidae	0	0	0
Ephemeroptera	0.95	0.89	0.92
Ephemeridae	0	0	0
Ephydriidae	0	0	0
Ergasilidae	0	0	0
Gammaridae	0.18	0.22	0.16
Gerridae	0	0	0
Glossiphoniidae	0	0	0
Glossosomatidae	0.34	0.44	0.3
Gomphidae	0.13	0.11	0.09
Gortidae	0.06	0.11	0.08
Gyrinidae	0	0	0
Haliplidae	0	0	0
Helodidae	0	0	0
Heptageniidae	0.67	0.78	0.8
Hydrobiidae	0	0	0
Hydrophilidae	0.04	0	0
Hydropsychidae	0.83	1	0.97
Hydroptilidae	0.11	0.11	0.08
Lepidoptera	0	0	0
Lepidostomatidae	0.04	0	0
Lepidostomatidae	0	0	0
Leptophlebiidae	0.14	0.11	0.13
Lestidae	0	0	0
Leuctridae	0.08	0	0
Libellulidae	0	0	0
Limnephilidae	0.51	0.44	0.43
Lumbriculidae	0.08	0.11	0
Lymnaeidae	0.05	0.11	0.08
Macromiidae	0	0	0
Maxillopoda	0	0	0
Metretopodidae	0	0	0
Molannidae	0	0	0
Muscidae	0	0	0
Naididae	0.37	0.34	0.43
Nemouridae	0.82	0.67	0.69
Nepidae	0	0	0
Notonectidae	0	0	0
Odontoceridae	0	0	0
Oligoneuridae	0.14	0.11	0.03
Ostracoda	0	0	0
Palaemonidae	0	0	0
Peltoperlidae	0.05	0	0
Perlidae	0.36	0.44	0.34
Perlotidae	0.39	0.34	0.41
Phlopotamidae	0.36	0.45	0.55
Phryganeidae	0	0	0
Physidae	0.04	0	0
Piscicolidae	0	0	0
Pisidiidae	0.1	0.22	0.21
Planariidae	0.1	0.11	0.08
Planorbidae	0	0	0
Pleuroceridae	0.05	0	0
Polycentropodidae	0.2	0.23	0.34
Potamanthidae?	0	0	0
Psephenidae	0.06	0.11	0.08
Psychodidae	0	0	0
Psychomyiidae	0.06	0.11	0.08
Pteronarcidae	0.09	0	0
Ptilodactylidae?	0.13	0.11	0.13
Ptychopteridae	0	0	0
Rhacophyllidae	0.57	0.33	0.34
Saldidae	0	0	0
Sciomyzidae	0	0	0
Sericostomatidae	0	0	0
Sialidae	0	0	0
Simuliidae	0.94	0.88	0.75
Siphonuridae	0.32	0.44	0.31
Sisyridae	0	0	0
Sphaeriidae	0	0	0
Stratiomyidae	0	0	0
Syrphidae	0	0	0
Tabanidae	0	0	0
Taeniopterygidae	0.05	0	0
Talitridae	0	0	0
Tetrastemmatidae	0	0	0
Tipulidae	0.79	0.78	0.79
Tubificidae	0.1	0.22	0.24

Ulenidae	0	0	0
Uronidae	0	0	0
Valvatidae	0	0	0
Velidae	0	0	0
Viviparidae	0	0	0
Genus:			
Ablabesmyia	0	0.05	0
Acentrella	0.1	0.15	0
Acerpenna	0.16	0.15	0
Acroneria	0.08	0.05	0.17
Agapetus	0.05	0.05	0
Allocapnia	0.21	0.2	0
Alotanytus	0	0	0
Ameletus	0.41	0.4	0.27
Amphinemura	0.64	0.74	0.39
Anchyrtarsus	0.15	0.15	0.35
Ancylidae	0	0	0
Ancyronyx	0	0	0
Antocha	0.26	0.25	0.34
Aspectrotanytus	0	0	0
Argia	0	0	0
Atherix	0	0	0
Baetis	0.05	0.05	0
Bezzia	0	0	0
Bitacomorpha	0.05	0.05	0
Blepharicera	0	0	0
Boyeria	0.05	0.05	0
Brachycentrus	0	0	0
Branchiobdellidae	0	0	0
Brilia	0.1	0.1	0
Caecidotea	0.06	0.05	0.06
Caenis	0.06	0.05	0.06
Calibaetis	0.06	0.05	0.06
Calopteryx	0.06	0.05	0.06
Cambardae	0	0.05	0
Centropilum	0	0.05	0
Ceratopogon	0.14	0.15	0.19
Chaetocladius	0	0.05	0
Chelifera	0.09	0.05	0.19
Cheumatopsyche	0.45	0.45	0.57
Chimarra	0.12	0.1	0.09
Chironomus	0	0	0
Chloroperlidae	0.09	0.1	0.19
Chrysops	0	0	0
Cinygmula	0	0	0
Cladopelma	0	0	0
Cladotanytarsus	0	0	0
Clinocera	0.27	0.25	0.04
Clinotanytus	0	0	0
Cloperla	0	0	0
Cnephia	0	0	0
Collembola	0	0.05	0
Conchapelopia	0.64	0.6	0.71
Cordulegaster	0	0.05	0
Cordulidae	0.06	0.05	0.06
Corixidae	0.06	0.05	0.06
Corydalus	0	0	0
Corynoneura	0.31	0.3	0.27
Crangonyx	0.12	0.1	0.04
CricotopusOrthocladius	0.63	0.6	0.66
Cryptochironomus	0	0	0
Culicoides	0	0	0
Cullus	0	0	0
Curia	0	0	0
Deronectes	0.06	0.05	0.06
Dianesa	0.28	0.2	0.41
Dicranota	0.1	0.15	0
Dicrolentipes	0	0	0
Dineutus	0	0	0
Dipheter	0.06	0.05	0.08
Diplectrona	0.31	0.35	0.27
Diplocladius	0	0	0
Diploperla	0.06	0.1	0.09
Dixa	0	0.05	0
Dolichopodidae	0	0.05	0
Dolophiodes	0.06	0.1	0.08
Drunella	0.11	0.1	0
Dubiraphia	0.05	0.05	0
Dugesia	0	0.05	0
Ecocoptera	0.05	0.05	0
Ectoptera	0	0	0
Enallagma	0.06	0.05	0.06
Enchytraeidae	0	0.05	0
Endochironomus	0.11	0.1	0.06
Epeorus	0.25	0.25	0.27
Ephemerella	0.73	0.7	0.94
Erioptera	0	0	0
Eukiefferiella	0.36	0.35	0.23
Eurylophella	0.15	0.15	0.23
Fallcon	0	0	0
Gammarus	0.06	0.05	0.04
Glossocoma	0.18	0.15	0.17
Gomphidae	0.22	0.16	0.4
Goniobasis	0	0	0
Gyrinus	0	0	0
Habrophlebia	0	0	0
Heleniella	0	0	0
Helichus	0	0.05	0
Helius	0	0	0
Hemenodromia	0.36	0.31	0.63
Heptagenia	0	0	0
Heterotrisocladius	0.05	0.1	0
Hexagenia	0	0	0
Hexatoma	0.12	0.1	0.09
Hyaella	0	0	0
Hydrobaenus	0.15	0.1	0.26
Hydrobius	0	0.1	0
Hydroporus	0	0	0
Hydropsyche	0.36	0.3	0.34
Ironoxipia	0	0.05	0
Isonychia	0.05	0.05	0
Isoptera	0.05	0.1	0
Kiefferulus	0.06	0.05	0.06
Krenopelopia	0.05	0.05	0
Labrundinia	0	0	0
Larsia	0	0	0
Lepidoptera	0	0.05	0
Lepidostoma	0	0.05	0

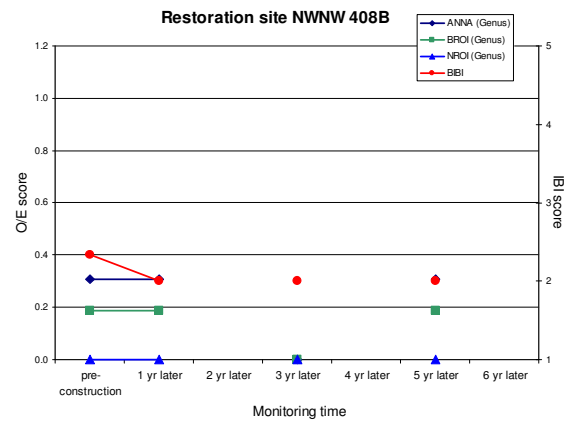
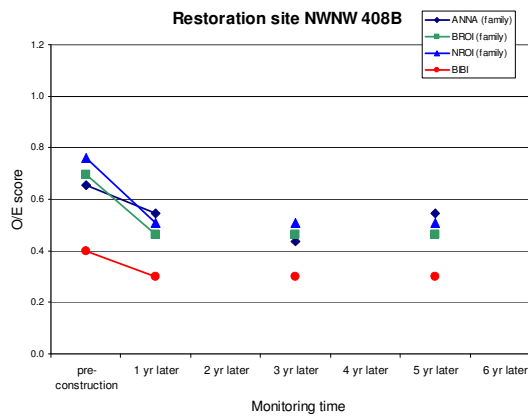
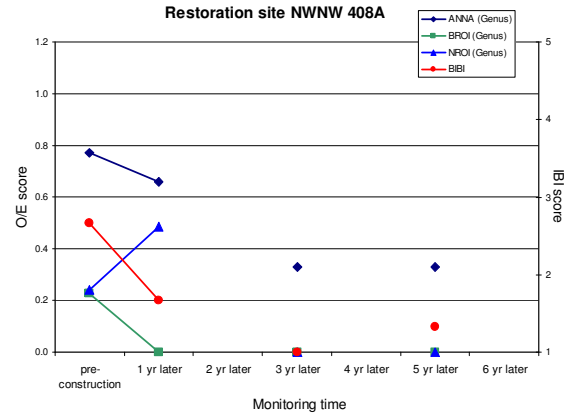
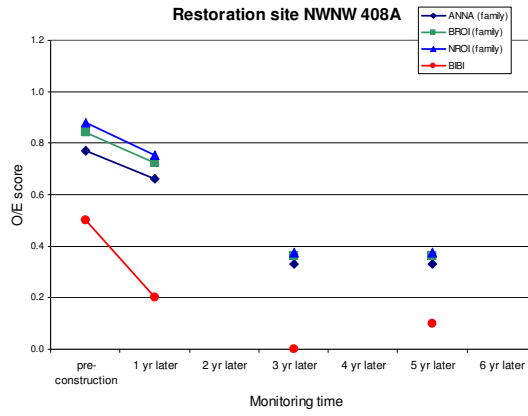
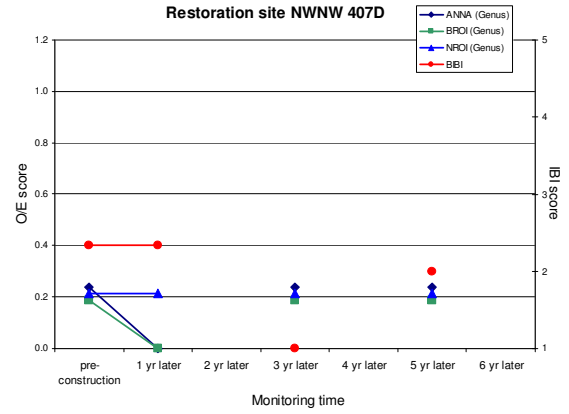
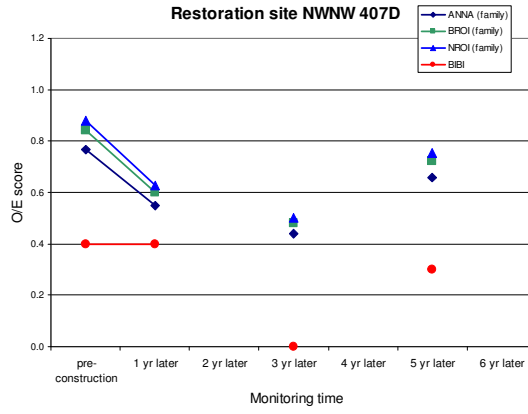
Leptophlebia	0.06	0.1	0.06
Leptoxis	0	0	0
Leucotrichia	0	0	0
Leucrocota	0	0	0
Leuctridae	0.11	0.2	0
Limnophila	0.05	0.05	0
Limnophyes	0.05	0.05	0
Lopescladius	0	0	0
Lumbriculidae	0.11	0.15	0
Lype	0	0	0
Macronychus	0	0	0
Menetus	0	0	0
Menopelopia	0.05	0.1	0
Micrasema	0	0	0
Micropectra	0.38	0.4	0.39
Microtendipes	0.14	0.1	0.19
Microvelia	0	0	0
Molophilus	0	0	0
Musculium	0	0	0
Naididae	0.33	0.35	0.45
Nanocladius	0	0.05	0
Nematomorphia	0	0	0
Nemoura	0	0	0
Neophylax	0.5	0.55	0.49
Nigronia	0.12	0.2	0.08
Nixe	0	0	0
Oceclis	0	0	0
Oemopteryx	0	0	0
Optioservus	0.38	0.45	0.48
Ormosia	0	0	0
Orthocladinae(A)	0	0	0
Ostrocera	0	0	0
Oulimnius	0.58	0.6	0.67
Pagastia	0.08	0.05	0.17
Palaemonetes	0	0	0
Paracapnia	0	0	0
Parachaetocladius	0	0	0
Parachironomus	0.06	0.05	0.06
Paragragnia	0	0	0
Parakiefferiella	0	0	0
Paraleptophlebia	0.13	0.15	0.17
Paramerina	0.06	0.05	0.06
Parametricnemus	0.66	0.65	0.65
Paraphaenocladus	0	0	0
Paratanytarsus	0.24	0.2	0.19
Pedicia	0	0	0
Pelodytes	0.06	0.05	0.06
Peltopera	0	0	0
Perlesta	0	0	0
Phaenopsectra	0.09	0.05	0.19
Phaenopsectora	0	0	0
Phyllocentropus	0	0	0
Physidae	0.06	0.15	0.06
Pisidium	0.08	0.05	0.17
Planorbella	0	0	0
Platycentropus	0	0	0
Platysmittia	0	0.05	0
Polycentropus	0.26	0.25	0.3
Polyptelidum	0.35	0.35	0.26
Pothastia	0.06	0.05	0.09
Probezzia	0	0	0
Procladius	0	0	0
Pronemora	0	0	0
Prosimulium	0.57	0.65	0.6
Prostoia	0.36	0.35	0.23
Prostoma	0	0	0
Psephenus	0	0	0
Pseudochironomus	0	0	0
Pseudolimnophila	0.05	0.1	0
Pseudorthocladius	0.05	0.05	0
Pseudosuccinea	0	0	0
Psilotreta	0	0	0
Pteronarcys	0	0.05	0
Ptilostomis	0	0	0
Pycnosyche	0.11	0.1	0
Radix	0	0	0
Rheocricotopus	0.05	0.05	0
Rheotanytarsus	0.19	0.15	0.26
Rhyacophila	0.58	0.6	0.67
Serratella	0.14	0.25	0.25
Sialis	0.05	0.05	0
Simulium	0.24	0.3	0.17
Siphonurus	0	0	0
Siphonoplecton	0	0	0
Sphaeriidae	0	0.05	0
Sphaerium	0	0	0
Sphaeromias	0.23	0.25	0.13
Stempelella	0	0	0
Stempelella	0.05	0.05	0
Stenacron	0	0	0
Stenelmis	0.08	0.1	0.19
Stenonema	0.36	0.4	0.34
Stictochironomus	0.05	0.05	0
Stilocladius	0.05	0.05	0
Strophopteryx	0.17	0.15	0.04
Stygobromus	0	0	0
Symposiocladius	0	0	0
Symphothaestia	0.12	0.1	0.09
Symulium	0	0	0
Tabanus	0	0	0
Taeniopteryx	0	0	0
Tallaperla	0	0	0
Tanytus	0	0	0
Tanytarsus	0.3	0.35	0.26
Thienemannella	0.22	0.25	0.09
Thienemannimyia(gp)	0.05	0.05	0
Tipula	0.27	0.39	0.04
Trepobates			

Appendix F. The BIBI and O/E Comparison at Restoration Stream Sites



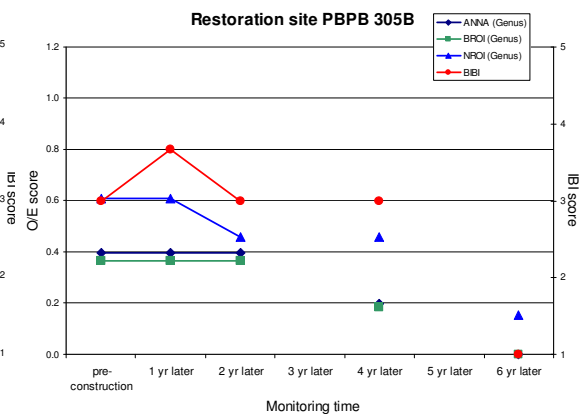
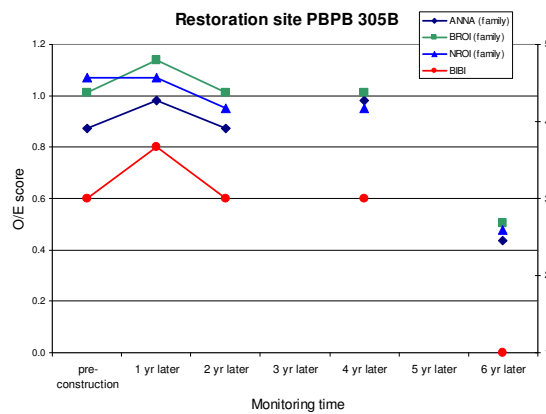
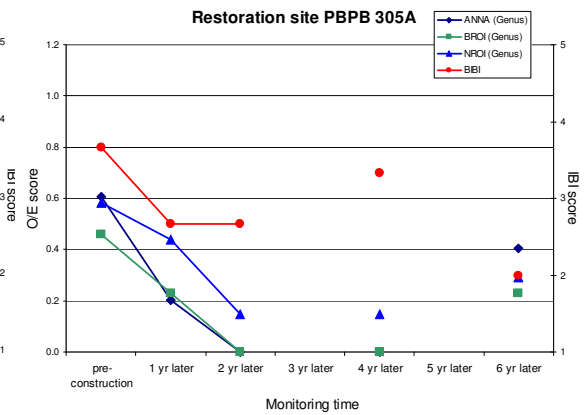
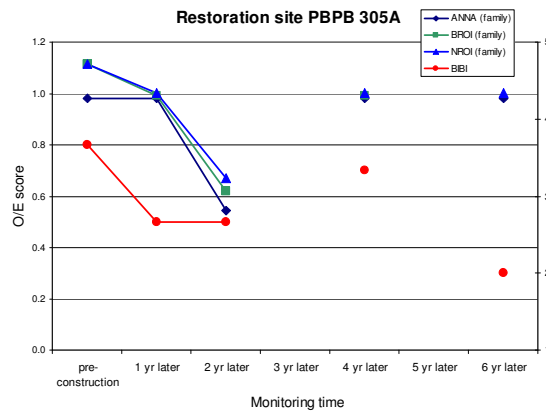
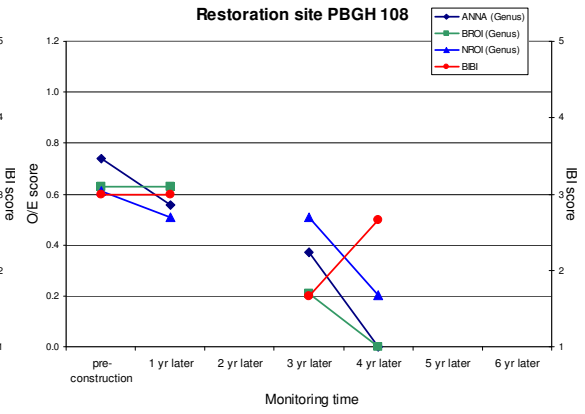
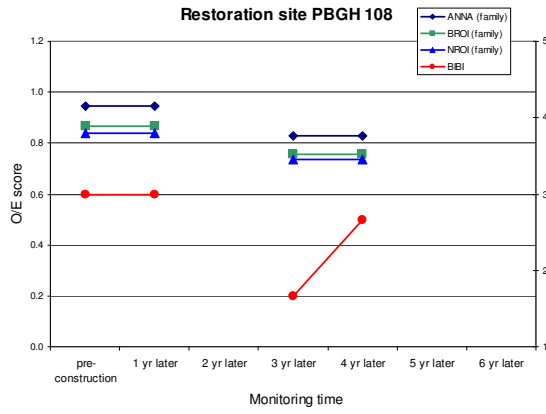
Family

Genus



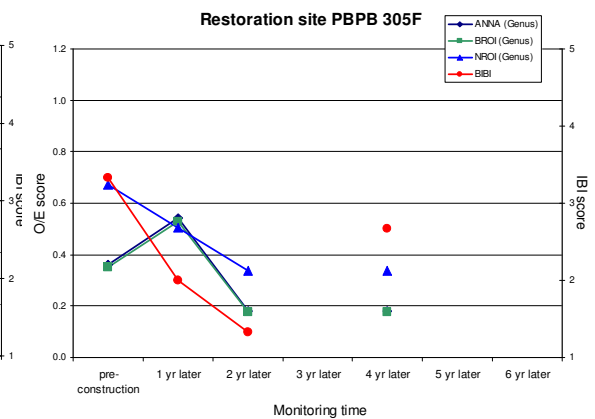
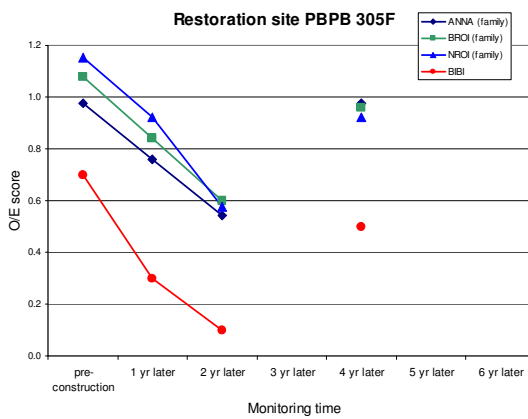
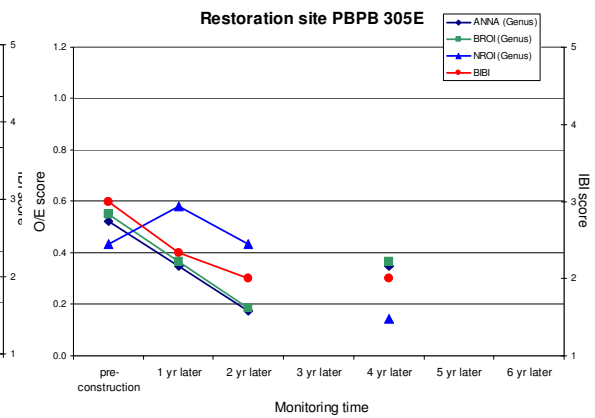
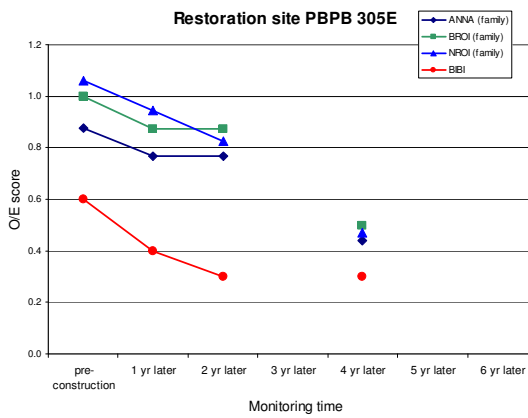
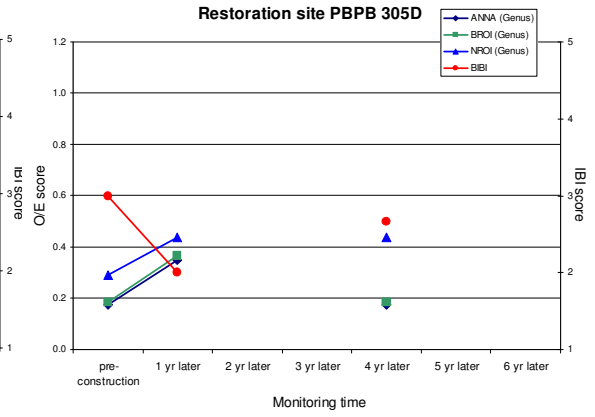
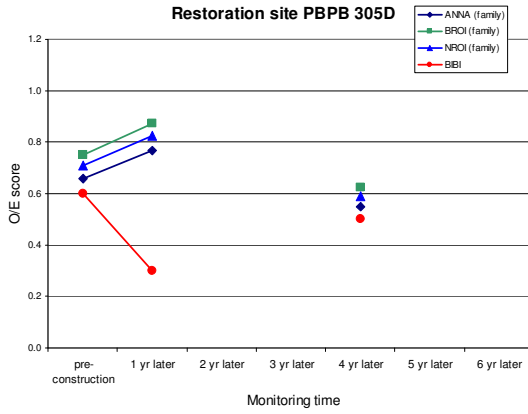
Family

Genus



Family

Genus



References

- Acreman, M. C., and Wiltshire, S. E. (1987). Identification of Regions for Regional Flood Frequency analysis (abstract). *Eos, Transactions, American Geophysical Union*, 68(44), 1262.
- Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. Paper presented at the Second International Symposium on Information Theory, Akademiai Kiadó, Budapest.
- Bernhardt, E. S., Palmer, M. A., Allan, J. D., Alexander, G., Barnas, K., Brooks, S., Carr, J., Clayton, S., Dahm, C., Follstad-Shah, J., Galat, D., S.Gloss, P., Goodwin, D., Hart, B. H., R. Jenkinson, Katz, S., Kondolf, G. M., Lake, P. S., Lave, R., Meyer, J. L., O'Donnell, T. K., Pagano, L., Powell, B., and Sudduth, E. (2005). Synthesizing U.S. River Restoration Efforts. *Science*, 308(5722), 636-637.
- Bras, R. L. (1990). *Hydrology: an introduction to hydrologic science*: Addison-Wesley.
- Bray, J. R., and Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 325-349.
- Burn, D. H. (1990a). An Appraisal of the "Region of Influence" Approach to Flood Frequency Analysis. *Hydrological Sciences*, 35(2), 149-165.
- Burn, D. H. (1990b). Evaluation of Regional Flood Frequency Analysis with a Region of Influence Approach. *Water Resources Research*, 26(10), 2257-2265.
- Burnhan, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference -- A Practical Information-Theoretic Approach* (Second ed.). New York: Springer-Verlag.
- Clarke, K. R. (1993). Non-Parametric Multivariate Analyses of Changes in Community Structure. *Australian Journal of Ecology*, 18, 117-143.
- Clarke, K. R., and Ainsworth, M. (1993). A Method of Linking multivariate Community Structure to Environmental Variables. *Marine Ecology Progress Series*, 92, 205-219.
- Cox, T. F., and Cox, M. A. A. (2001). *Multidimensional Scaling* (Second ed.): Chapman & Hall/CRC.

- Davis, W. S., and Simon, T. P. (Eds.). (1994). *Biological Assessment and Criteria -- Tools for Water Resource Planning and Decision Making*. Boca Raton, Florida: Lewis Publishers.
- Hastie, T. J. (1992). *Generalized Additive Models*. Boca Raton: Chapman & Hall/CRC.
- Hawkins, C. P., Norris, R. H., Hogue, J. N., and Feminella, J. W. (2000). Development and Evaluation of Predictive Models for Measuring the Biological Integrity of Streams. *Ecological Applications*, 10, 1456-1477.
- Hurvich, C. M., and Tsai, C.-L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika*, 76(2), 297-307.
- Jähnig, S. C., Hering, D., Buffagni, A., Erba, S., and Brabec, K. (2008). *Effects of Stream Restoration in European Mountain Streams on Hydromorphological Characteristics and Benthic Macroinvertebrates*. Paper presented at the The NABS 56th Annual Meeting, Salt Lake City, Utah.
- Karr, J. R., and Chu, E. W. (1999). *Restoring Life in Running Waters Better Biological Monitoring*. Washington, DC Covelo, California: Island Press.
- Kruskal, J. B. (1964). Multidimensional Scaling by Optimizing Goodness of Fit for a Nonmetric Hypothesis. *Psychometrika*, 29, 1-27.
- Kruskal, J. B., and Wish, M. (1978). *Multidimensional Scaling*. Murray Hill, N.J.: SAGE.
- Linke, S., Norris, R. H., P.Faith, D., and Stockwell, D. (2005). ANNA: A New Prediction Method for Bioassessment Programs. *Freshwater Biology*, 50, 147-158.
- Lumley, T. (2004). *regression subset selection (2.7)*. Available: <http://cran.r-project.org/web/packages/leaps/index.html>.
- Malakoff, D. (2004). The River Doctor. *Science*, 305(5686), 937-939.
- McCuen, R. H., and Snyder, W. M. (1985). *Hydrologic Modeling Statistical Method and Applications*. Englewood Cliffs, New Jersey: Prentice-Hall, A Division of Simon & Schuster, Inc.
- McQuarrie, A. D. R., and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*: World Scientific.
- Moerke, A. H., Gerard, K. J., Latimore, J. A., Hellenthal, R. A., and Lamberti, G. A. (2004). Restoration of an Indiana, USA, Stream: Bridging the Gap between Basic and Applied Lotic Ecology. *Journal of North American Benthological Society*, 23(3), 647-660.

- Moglen, G. E. (2005, November 2005). *GISHydro2000+ User's Manual*. Available: <http://www.gishydro.umd.edu/document.htm>.
- Moglen, G. E., and Paul, M. J. (2005). *Using GISHydro2000 to Make Pollutant/Nutrient Loading and Stream Biodiversity Estimates*. Available: <http://www.gishydro.umd.edu/document.htm>.
- Moss, D., Furse, M. T., Wright, J. F., and Armitage, P. D. (1987). The Prediction of the Macro-Invertebrate Fauna of Unpolluted Running-Water Sites in Great Britain Using Environmental Data. *Freshwater Biology*, 17, 41-52.
- Moss, D., Wright, J. F., Furse, M. T., and Clarke, R. T. (1999). A Comparison of Alternative Techniques for Prediction of the Fauna of Running-Water Sites in Great Britain. *Freshwater Biology*, 41, 167-181.
- Niemi, G. J., DeVore, P., Detenbeck, N., Taylor, D., Lima, A., and Pastor, J. (1990). Overview of Case Studies on Recovery of Aquatic Systems from Disturbance. *Environmental Management*, 14(5), 571-587.
- Palmer, M. A., Bernhardt, E. S., Allan, J. D., Lake, P. S., G.Alexander, Brooks, S., Carr, J., Clayton, S., Dahm, C. N., Shah, J. F., Galat, D. L., Loss, S. G., Goodwin, P., Hart, D. D., Hassett, B., Jenkinson, R., Kondolf, G. M., Lave, R., Meyer, J. L., O'Donnell, T. K., Pagano, L., and Sudduth, E. (2005). Standards for Ecologically Successful River Restoration. *Journal of Applied Ecology*, 42, 208-217.
- Paul, M. J. (2006). unpublished work
- R Development Core Team. (2007). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available: <http://www.R-project.org>.
- Reice, S. R., Wissmar, R. C., and Naiman, R. J. (1990). Disturbance Regimes, Resilience, and Recovery of Animal Communities and Habitats in Lotic Ecosystems. *Environmental Management*, 14(5), 647-659.
- RIVPACS (River Invertebrate Prediction and Classification System)*. [Verified on Oct.17, 2007]. Available: <http://www.ceh.ac.uk/sections/re/RIVPACS.html>.
- Roth, N. E., Southerland, M. T., Chaillou, J. C., Vølstad, J. H., Weisberg, S. B., Wilson, H. T., Heimbuch, D. G., and Seibel, J. C. (1997). *Maryland Biological Stream Survey: Ecological Status of Non-Tidal Streams in Six Basins Sampled in 1995* (Report No. CBWP-MANTA-EA-97-2. NTIS No. PB98-221310). Annapolis, MD.: Maryland Department of Natural Resources.
- Roth, N. E., Southerland, M. T., Mercurio, G., and Vølstad, J. H. (2001). *Maryland Biological Stream Survey 2000-2004 Volume I: Ecological Assessment of Watersheds sampled in 2000*: Maryland Department of Natural Resources.

- Sarle, W. (2007, Posted on May 13 2007). *What Is Overfitting and How Can I Avoid It?* Available: <http://www.faqs.org/faqs/ai-faq/neural-nets/part3/section-3.html> [Varified on Oct.1, 2007].
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464.
- Shields, F. D., Cooper, C. M., Knight, S. S., and Moore, M. T. (2003). Stream Corridor Restoration Research: A Long and Winding Road. *Ecological Engineering*, 20, 441-454.
- Smith, S., Dawson, F., Lubbers, L., Primrose, N., Smith, K., Yetman, K., Herrmann, M., Wheeler, J., Reger, J., Schwarz, K., Doheny, E., Wheeler, J., Gagnon, A., and Gutierrez, L. A. (2003). *Maryland Streams, Take A Closer Look*.
- Smith, S. M., and Prestegard, K. L. (2005). Hydraulic Performance of a Morphology-ased Stream Channel Design. *Water Resources Research*, 41, w11413.
- Southerland, M. T., Rogers, G. M., Kline, M. J., Morgan, R. P., Boward, D. M., Kazyak, P. F., Klauda, R. J., and Stranko, S. A. (2005). *New Biological Indicators to Better Assess the Condition of Maryland Streams* (CBWP-MANTA-EA-05-13). Annapolis, Maryland: Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division.
- Sugiura, N. (1978). Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections. *Communications in Statistics, Theory and Methods*, A7, 13-26.
- Tanaka, J. S. (1987). "How Big Is Big Enough?": Sample Size and Goodness of Fit in Structural Equitition Models with Latent Variables. *Child Development*, 58(1), 134-146.
- Tasker, G. D., Hodge, S. A., and Barks, C. S. (1996). Region of Influence Regression for Estimating the 50-year Flood at Ungaged Sites. *Water Resources Bulletin*, 32(1), 163-170.
- Tasker, G. D., and Slade, R. M. J. (1994). *An Interactive Regional Regression Approach to Estimating Flood Quantiles* Paper presented at the Water Policy and Management: Solving the Problems: Proceedings of the 21st Annual Conference, Denvor, Colorado.
- U.S.EPA. (1995). *National Water Quality Inventory: 1994 Report to the Congress*. Washington, DC.: U.S. Environmental Protection Agency.
- U.S.EPA. (1999). Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers - Periphyton, Benthic Macroinvertebrates, and Fish. In U.S.EPA. (Ed.),

Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers.
Washington DC.

- Van Sickle, J., Huff, D. D., and Hawkins, C. P. (2006). Selecting Discriminant Function Models for Prediction the Expected Richness of Aquatic Macroinvertebrates. *Freshwater Biology*, 51, 359-372.
- Wallace, J. B. (1990). Recovery of Lotic Macroinvertebrate Communities from Disturbance. *Environmental Management*, 14(5), 605-620.
- Walsh, C. J., Fletcher, T. D., and Ladson, A. R. (2005). Stream Restoration in Urban Catchments Through Redesigning Stormwater Systems: Looking to the Catchment to Save the Stream. *Journal of North American Benthological Society*, 24(3), 690-705.
- Wright, J. F. (1995). Development and Use of a System for Predicting the Macroinvertebrate Fauna in Flowing Waters. *Australian Journal of Ecology*, 20, 181-197.
- Wright, J. F. (2000). An Introduction to RIVPACS. In J. F. Wright and D. W. Sutcliffe and M. T. Furse (Eds.), *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques* (pp. 1-23). Ambleside, England, U.K.: Freshwater Biological Association.
- Wright, J. F., Moss, D., Armitage, P. D., and Furse, M. T. (1984). A Preliminary Classification of Running -Water Sites in Great Britain Based on Macro-Invertebrate Species and the Prediction of Community Type Using Environmental Data. *Freshwater Biology*, 14, 221-256.
- Wright, J. F., Sutcliffe, D. W., and Furse, M. T. (Eds.). (2000). *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques*. Ambleside, Cumbria, UK: Freshwater Biological Association.
- Yount, J. D. (1990). Recovery of Lotic Communities and Ecosystems from Disturbance -- A Narrative Review of Case Studies. *Environmental Management*, 14(5), 547-569.
- Zhu, M., and Chipman, H. A. (2006). Darwinian Evolution in Parallel Universes: A Parallel Genetic Algorithm for Variable Selection. *Technometrics*, 48(4), 491-502.