

A DECISION SUPPORT TOOL FOR PREDICTING WATER QUALITY BASED ON
LAND COVER

by

Brittney Baker

Dr. Patrick N. Halpin, Advisor

May 2011

Masters project submitted in partial fulfillment of the
requirements for the Master of Environmental Management degree in
the Nicholas School of the Environment of
Duke University
2011

ABSTRACT

Managers at Marine Corps Base Camp Lejeune must evaluate the environmental impacts of their proposed development plans. The effect of land cover changes on water quality is an important consideration for these evaluations. An interactive geospatial tool was developed in 2009. The tool allows managers to interactively select their proposed development site and input what the proposed land cover will be for the site. The tool returns the changes in average ammonium concentration in the tributary creeks. The tool incorporates water quality data collected by the DCERP project from 2008-2009 to drive the prediction model.

The purpose of this project was to (1) improve usability of the tool to make it a spatial decision support system, (2) update the water quality data used to drive the statistics of the water quality prediction models, and (3) determine if using the National Land Cover Dataset (NLCD) from 2006 instead of the 2001 NLCD changes the relationship between land cover predictors and water quality response variables.

Tool usability was enhanced by adding in-tool and external help menus, creating a user guide, and adding the ability to name and save outputs. The updated tool allows the user to run multiple land development scenarios for comparison without overwriting the previous results. Adding additional water quality data from 2007-2010 resulted in fewer significant water quality prediction models. The most predictive of these models was for organic nitrogen. The model, predicted by barren (rock/sand/clay), shrub/scrub, and grassland/herbaceous land covers, was incorporated into the decision support tool. Using land cover data from 2001 and 2006 allowed the same two water quality parameters to be predicted: NO_x and ON. Managers at Marine Corps Base Camp Lejeune can use this data driven Spatial Decision Support System to evaluate how different development scenarios will affect the concentration of organic nitrogen in the tributary creeks on base.

Table of Contents

INTRODUCTION	1
BACKGROUND	1
<i>Decision Support Systems</i>	1
<i>Spatial Decision Support Systems</i>	2
DECISION SUPPORT SYSTEM CASE STUDY: WATER QUALITY PREDICTION BY LAND COVER CHANGE TOOL	2
<i>Study Site</i>	3
OBJECTIVE	3
METHODS.....	4
TOOL USABILITY	4
WATER QUALITY DATA & STATISTICAL ANALYSIS	5
<i>Automated Statistical Modeling</i>	5
<i>Manual Statistical Modeling</i>	6
<i>Model Incorporation into the Tool</i>	6
LAND COVER CHARACTERISTICS	6
RESULTS.....	7
TOOL USABILITY	7
WATER QUALITY DATA & STATISTICAL ANALYSIS	7
<i>Automated Statistical Modeling</i>	7
<i>Manual Statistical Modeling</i>	9
<i>Model Incorporation into the Tool</i>	10
LAND COVER CHARACTERISTICS	10
DISCUSSION.....	12
NEXT STEPS	12
CONCLUSION	13
REFERENCES	14
LIST OF TABLES.....	16
LIST OF FIGURES.....	26
APPENDIX 1. USER GUIDE.....	44

INTRODUCTION

BACKGROUND

There is strong evidence for a connection between land use characteristics and the quality of surrounding surface water. There is evidence that anthropogenic change in land cover affects water quality (Tong & Chen 2002; Bolstad & Swank 1997). Being able to model the relationship between land cover characteristics and water quality allows environmental managers to make informed decisions about where and how to develop land in order to maintain or reduce degradation to water quality. Environmental managers can use tools such as Decision Support Systems to aid in the decision making process.

Decision Support Systems

Managers must make decisions about matters that involve uncertainty and risk. They often must make these decisions with a limited timeframe and with incomplete information. Decision Support Systems (DSS) can be used to assist managers in making these often difficult choices. DSS involve a methodology or application that is designed to assist in making choices about actions (Rhodes 1993; Bendoly 2008). DSS provide easier access to data, facilitated analysis, and rich communication of results, often through the effective use of visualization (Bendoly 2008). It is important to note that these applications are intended to support and not replace the decision maker in the decision making process (Bendoly 2008). DSS rely on models and data incorporated into a user-friendly interface that can incorporate the decision-makers' own insights (Matthies et al 2007).

Decision Support Systems must be designed carefully. Best practices have been outlined in the literature. Designers of DSS have the responsibility to pay attention to exactly how applications might be used, in order to ensure that end users will not abuse the results generated (Bendoly 2008). Bendoly suggests that it is critical to guide usage through "built-in assistance in analysis, clear visualization of how characteristics of problems and solutions relate, and formally structured interfaces that deter, if not prohibit, misuse" (2008). It is important for developers to clearly state the objectives of the tools including who will be using the tool and how they will use the tool because these will dictate the strategy of tool design (McIntosh et al 2008).

Although DSS were originally developed for business managers, there has been interest in using DSS for environmental quality management (Matthies et al. 2007). Because

environmental processes take place in three-dimensional space, the spatial component is important in making decisions about environmental issues. Spatial Decision Support Systems can be used to solve environmental problems.

Spatial Decision Support Systems

Spatial DSS (SDSS) integrate Geographic Information System (GIS) tools into the decision making process. SDSS have interfaces available to access and run scientific models (Laudien et al. 2010). SDSS can be used as comprehensive decision support tools for problems in environmental management and research (Laudien et al. 2010).

DECISION SUPPORT SYSTEM CASE STUDY: WATER QUALITY PREDICTION BY LAND COVER CHANGE TOOL

This project focuses on the Water Quality Prediction by Land Cover Change tool. This tool was created as part of the Defense Coastal/Estuarine Research Program. The Defense Coastal/Estuarine Research Program (DCERP) aims to “enhance and sustain the military mission by developing an understanding of coastal and estuarine ecosystem composition, structure, and function within the context of a military training environment” (RTI International 2011). The tool was created for managers at Marine Corps Base Camp Lejeune (MCBCL). The purpose of the tool is to model how changes in land cover affect the water quality. Using the tool in ArcGIS, users are able to interactively select an area of interest. The user indicates the proposed land cover value for the area of interest. The tool, in its original version, calculates how the concentration of ammonium changes in surrounding water bodies.

In 2009, Thomas Minter designed the Water Quality Prediction by Land Cover Change tool (referred to throughout as “the tool”). The purpose of the tool is to predict how the average concentration of ammonium will change in the tributary creeks of the New River when the land cover changes from one cover type to another cover type in a user defined area. The user of this interactive geospatial tool indicates what the new land cover will be based on the National Land Cover Data (NLCD) 2001 Land Cover Class definitions (available at <http://www.epa.gov/mrlc/definitions.html>). The user then interactively draws a polygon or uses a shapefile to indicate where land cover will change. The tool communicates with R statistical software to calculate the change in ammonium concentration based on the statistical relationship that Minter modeled with land use and land cover predictors compared with water nutrient concentration response variables. The tool is able to communicate with R using the Marine

Geospatial Ecology Tools (Roberts et al. 2010). Minter used water quality data collected between January 29, 2008 through July 27, 2009 and the 2001 NLCD land cover types to model the relationship between change in cover type and change in ammonium concentration in surrounding watersheds. The tool is based on water quality data from a limited time period.

This tool helps MCBCL fulfill their environmental protection goals. MCBCL is committed to environmental protection as evidenced from their 2001 Integrated Natural Resources Plan. MCBCL is “committed to environmental protection, continual environmental improvement and pollution prevention” (FINRMP 2006). To meet these goals MCBCL considers up to date science and geospatial analysis. The tool designed could be used by MCBCL to aid in pollution prevention. The tool allows managers to consider impacts on water quality before a land cover change project begins.

Study Site

The tool is applicable for Marine Corps Base Camp Lejeune located in the coastal plains of North Carolina (**Figure 1**). MCBCL is the largest amphibious Marine Corps training base. The base covers an area of 246 square miles (FINRMP 2006). The New River flows through the base. Many tributary creeks on base flow into the New River. Water quality was monitored in ten tributary creeks on the base. The water quality data from the tributary creeks and the base’s land cover were used to model the relationship between land cover and water quality.

OBJECTIVE

The purpose of this project was to (1) improve usability of the decision support system, (2) update the water quality data used to drive the statistics of the model, and (3) determine if using the 2006 NLCD instead of the 2001 NLCD changes the relationship between land cover predictors and water quality response variables.

The purpose of DSS is to streamline, facilitate, and aid decision making; therefore usability is absolutely crucial, because a confusing DSS complicates the decision making process. The user interface was updated to add help menus and tool parameters to enhance usability.

The original tool was based on water quality data from a short period of time. This analysis adds additional water quality data to determine how the relationship between land cover predictors and water quality parameter concentrations change when this additional data is added.

All of the available water quality data from DCERP was used to model statistical relationships. Significant models were then incorporated into the tool. Using the most up-to-date data ensures that decision makers have models that are representative of the past trends.

Since the original tool was designed, the NLCD has been updated through the year 2006. The updated land cover data was used as predictor variables to determine its effect on the water quality prediction models.

METHODS

TOOL USABILITY

Improvements were made to the Water Quality Prediction By Land Cover Change tool in order to enhance the usability of the tool by end users. These adjustments were made based on suggestions of best practices in the literature about decision support systems.

The first set of changes made involved documentation to aid the user in understanding each of the underlying assumptions used in creating the tool. An in-tool help menu was created in ArcGIS 9.3 model builder. This help menu describes each user input into the model. Next, an external help menu was created. This menu is available when the user right clicks on the tool, in the toolbox and selects “Help”. This help menu describes each of the user inputs as well as each processing step in the tool. This external help menu was also created in ArcGIS 9.3 model builder. Finally, a user guide was created to detail the assumptions and methods used in tool creation and a step-by-step guide about how to use the tool.

The next set of changes involved changes to the tool itself. First, the ArcGIS toolbox that the tool resides in was renamed to reflect the contents of the toolbox. Then, the capability for the user to name their area of interest was added. The capability was also added for the user to choose the directory in which to save the area of interest polygon. This allows the user to run the tool multiple times, with different areas of interest, without overwriting the results. This makes it easier for users to compare results from different scenarios and make more informed decisions about land development choices. The capability to save the output file to a text file in the working directory was also added. In the original version of the tool, the results of water quality concentration change were only displayed in the ArcMap results window. Allowing the user to save the results allows the user to run the tool multiple times with each result being saved

automatically. Finally, the file structure of the tool was reorganized to make it easy to provide the tool to users.

WATER QUALITY DATA & STATISTICAL ANALYSIS

The most recent water quality data was obtained from the DCERP website. The data was compiled for all ten monitoring stations that were used in the original analysis (data collected by DCERP project co-PI M. Piehler). The ten monitoring sites are illustrated in **Figure 2** and listed with their locations in **Table 1**. **Figure 2** also shows the delineated watersheds created from each of the sampling points. The land cover in these watersheds represents the land cover that influences the water quality at the associated monitoring station. The six water quality parameters collected are listed in **Table 2**. The number of samples taken of each parameter at each of the monitoring stations are listed in **Table 3** and illustrated in **Figure 3**. Water quality samples were collected between 12/11/2007-12/13/2010. Most of the sampling sites had at least 100 samples taken over the sampling period. Two of the stations, Airport Creek and Southwest Creek, however, had fewer than fifty samples collected. These watersheds are much smaller than the other watersheds where water quality samples were collected. After the data was downloaded, the average concentration of each parameter at each of the ten monitoring sites was calculated. This gives the average concentration of each water quality parameter at each of the ten stations.

Then, the twenty-four land use and land cover predictors, used in the original analysis, were compared with the six nutrient concentration response variables for each of the ten sites. These land use and land cover predictors are listed in **Table 4**. To create water quality prediction models, two statistical methods were used.

Automated Statistical Modeling

The new average concentrations were used as the inputs into a general linear model (GLM) using the R statistical software package. This analysis produces a linear model for each predictor and response variable. The R GLM model code was written by Ben Best and was used in the original analysis. **Table 5** shows the function written for this analysis. Following Minter's methodology, significant predictors ($p < 0.05$) were isolated for each nutrient variable and used to fit the best general linear model. A step-wise model evaluation (StepAIC) was used

to identify the best-fit water quality prediction model (Akaike 1974). The lower the AIC score, the more the model is able to predict the water quality parameter concentration.

Manual Statistical Modeling

The manual method allows the selection of potentially significant land cover predictor variables. There were twenty-four land cover predictors predicting only 10 sample points for each water quality parameter. Because one outlier point could affect the linear models in the automated statistical modeling method, a manual method was also used. The manual method allowed relationships with visual outliers to be excluded as predictors in the model selection process. The linear trend that is predicted with the automated method may not be the strongest model available because it may fit a linear relationship where one is not appropriate. The manual method involves creating pair plots for each predictor and response variable pair. Pair plots were visually evaluated to determine if a linear relationship existed. Then, linear models were constructed by adding together all of the variables that showed a linear relationship with the predictors. StepAIC was used to identify the best fit model. These models were evaluated using p-value of the intercept and the predictors, R-squared value, and residuals.

Model Incorporation into the Tool

Linear models that were found to be strong predictors of a particular water quality parameter were used to drive the outputs of the interactive geospatial tool.

LAND COVER CHARACTERISTICS

For the original analysis, twenty-four land cover characteristics were defined within each of the watersheds to be used as predictors of water quality (**Table 4**). The original analysis included water quality predictors that used the NLCD 2001 data. National Land Cover Data 2006 (NLCD 2006) is an update to the National Land Cover Data 2001 that was used to calculate land cover for the original tool. In ArcMap, the tabulate area tool was used to calculate the area of each land cover type in each watershed. The 2001 values were compared with the 2006 variables. Then, the NLCD 2006 land cover was used as the predictors in a model selection process to see how the updated land cover affected the linear water quality prediction models. To be consistent with the original analysis, land cover types were only used as water quality predictors if the land cover type existed in at least 60 percent of the watersheds. The NLCD 2006 land cover predictors were evaluated in linear relationships using the manual statistics

method (described above). The results were compared to the results of the manual statistics method that used the NLCD 2001 data.

RESULTS

TOOL USABILITY

The first updates made were to the tool documentation. **Figures 4-9** show the updated help menus. If the parameter was a part of the original tool, its help menu is compared with the updated tool help menu in these figures.

Figure 10 shows the external help menu before and after the update. The external help menu defines what each of the tool parameters mean. It also provides an explanation of the tool elements or processes.

A user guide was created to clearly document the assumptions made in tool creation. It also details how to use the tool and how to interpret the results. The user guide is available in **Appendix 1**.

In addition to adding user help and documentation to the tool, changes were made to the tool itself to increase usability. First, the toolbox was renamed from “Toolbox” to “Water Quality Prediction Toolbox”. **Figure 11** shows the toolbox in the ArcGIS toolbox navigation before and after the change. The capability to name and save the user selected area of interest was added (**Figures 7-8**). In the updated version of the tool, the output file is saved to a log text file. The user is able to select the file location to save the output (**Figure 9**). These parameters were added to the tool’s user interface. These parameters allow the user to run the tool multiple times without overwriting their previous output. In this way, the tool can be used as a decision support tool because the user can run multiple scenarios and compare the outputs of the different scenarios. Examples of the tool’s outputs are illustrated in **Figure 12**. Finally, the structure of the tool was changed to make it easy for a new user to navigate. **Figure 13** shows the new file structure.

WATER QUALITY DATA & STATISTICAL ANALYSIS

Automated Statistical Modeling

First, the automated statistical analysis was conducted. **Table 6** shows the linear model results for each predictor (land cover, soil type) and response variable (water quality parameter)

for the updated water quality data. This is the output created when linear models were created for each predictor and response variable. Significant p-values are shown in red. When these linear models were calculated with water quality data from 12/11/2007-12/13/2010, TSS, NO_x, TN and ON each had at least one significant predictor ($p < 0.05$). The water quality parameters predicted with the expanded water quality data differs from the simple linear models modeled for the original set of water quality data. In the original analysis, TSS, NO_x, NH₄, PO₄, and ON each had at least one significant predictor ($p < 0.05$).

The significant predictors found in the last step were added together to calculate a generalized linear model (GLM) for each water quality parameter. Using the StepAIC method, the GLMs with the lowest AIC score were selected. The following General Linear Models were found to have the lowest AIC score for each water quality response variable:

- NO_x ~ Paved + Low Intensity Development + Woody Wetlands + > 30% Impervious surface + High Green Loss + Developed Lawn/Grass
- TSS ~ Light Green Loss
- TN ~ Medium Green Loss
- ON ~ Medium Green Loss + Barren Land (Rock/Sand/Clay) + Deciduous Forest + Mixed Pine/Hardwood Forest

Tables 7-10 show the results for these GLMs. The GLMs for NO_x and TSS did not have any significant predictor variables in the best-fit models ($p < 0.05$).

The GLM for TN had one significant predictor variable, medium green loss. The intercept for this GLM is not significant. This GLM indicates that as the percent area of medium green loss increases in a watershed, the total nitrogen concentration also increases.

The GLM for ON had two significant predictors, deciduous forest and mixed pine/hardwood forest. The intercept for this GLM is also significant. The barren (rock/sand/clay) predictor was significant to the 0.1 level. The final predictor in this best fit equation was medium green loss. Medium green loss had a p-value of 0.21. This GLM indicates that as the percent area of medium green loss, barren (rock/sand/clay), and mixed pine/hardwood forest increases in a watershed, the concentration of organic nitrogen increases. However, as the percent area deciduous forest increases in a watershed the concentration of organic nitrogen decreases.

The automated method used may be fitting linear relationships where one is not appropriate. For instance, **Figure 14** shows a pair-plot relationship between total suspended solids and the percent of area paved in a watershed. Percent paved area was determined to be a significant predictor with the automated method. However, the pair plot reveals that the linear relationship is only being fit due to the presence of an outlier point. The manual statistic method allowed for the exclusion of predictors that were being fit to an outlier point.

Manual Statistical Modeling

Pair plots were plotted for each of the six water quality response variables against each land cover predictor variable. Predictor variables that visually appeared to have a linear relationship with a water quality parameter were selected as possible predictors for the water quality prediction models. **Figures 15-16** show the pair plots of the selected possible land cover predictors. Only NO_x and ON had possible linear land cover predictors.

The possible predictors for each of the water quality parameters were added together and the best fit linear model was selected with the StepAIC method. **Tables 11-12** show the result for the best fit linear models.

The best-fit linear model for NO_x did not have any significant ($p < 0.05$) predictor variables (**Table 11**). Overall, the model was not significant, with a p-value of 0.064. The model had a high R-squared value of 0.9371. The R-squared value indicates explained variance in the model. The model explains about 93% of the variability in NO_x concentration.

The best-fit linear model for ON had two significant predictor variables, barren (rock/clay/sand) and shrub/scrub. The intercept was also significant. The grassland/herbaceous predictor variable was significant to the $p < 0.1$ level. Overall, the model was significant with a p-value of 0.004. The model had a high R-squared value of 0.8722. This indicates that the model explains about 87% of the variability in ON concentration. Because the grassland/herbaceous variable was not significant, the model was tested to see if removing this variable would improve the model. **Table 13** shows the summary of the ON linear model with the grassland/herbaceous predictor variable removed. This model has larger residuals than the ON model with the grassland/herbaceous predictor included.

Model Incorporation into the Tool

The organic nitrogen model chosen by the manual statistic method with the grassland/herbaceous variable included was found to be the most predictive model and was incorporated into the Water Quality Prediction by Land Cover tool. The model was saved as an R model. The tool uses the *Evaluate R Statements in Text File* tool to relate the spatial data about land cover to the water quality prediction model that is predicted by percent area of land cover in a watershed. The graphical view of the model used for the tool and instructions to use the interactive tool are shown in **Appendix 1**.

LAND COVER CHARACTERISTICS

The percent land cover for each watershed was calculated using the NLCD 2001 and NLCD 2006. **Table 14** shows the percent land cover in each watershed based on the NLCD 2001. **Table 15** shows the percent land cover in each watershed based on the NLCD 2006. To stay consistent with the original analysis, a land cover predictor had to be present in at least six of the monitoring stations' watersheds to be used as a water quality parameter predictor. The NLCD 2001 had the following predictor variables: developed, open space; developed, low intensity; developed, medium intensity; barren land (rock/sand/clay); deciduous forest; evergreen forest; mixed forest; shrub/scrub; grassland/herbaceous; woody wetlands; and emergent herbaceous wetlands. Due to change in land cover in the watersheds, deciduous forest and emergent herbaceous wetlands were dropped as predictors when using the NLCD 2006 because these land covers were not present in at least six of the monitoring station watersheds. The deciduous forest land cover made up less than 5% of the land cover in each of the watersheds with the NLCD 2001. The emergent herbaceous wetlands land cover was only present in six of the watersheds and also made up 5% or less of the land cover in each of the watersheds with the NLCD 2001. There were some notable changes in land cover between the two time-periods. There was a general increase in the percent cover of evergreen forest and shrub/scrub land covers from 2001 to 2006. There was also a general decrease in the percent cover of grassland/herbaceous and woody wetlands land covers from 2001 to 2006.

The NLCD 2006 land cover percent area calculations were used as predictors to model the water quality data collected between 12/11/2007-12/13/2010. The manual statistic modeling analysis method was used to determine best-fit linear models. Pair plots were plotted for each of

the six water quality response variables against each land cover predictor variable. Predictor variables that visually appeared to have a linear relationship with a water quality parameter were selected as possible predictors for the water quality prediction models. **Figures 17-18** show the pair plots of the selected possible land cover predictors. Only NO_x and ON had potential linear predictor variables.

The possible predictors for each of the water quality parameters were added together and the best-fit linear model was selected with the StepAIC method. **Tables 16-17** show the result for the best-fit linear models.

The best-fit linear model for NO_x had one significant ($p < 0.05$) predictor, paved land cover. The developed (lawn/grass) predictor was significant to the $p < 0.1$ level, with a p-value 0.06. The intercept was not significant. Overall, the model was significant with a p-value of 0.0003. The model also had a high R-squared value of 0.8971. This indicates that the model is able to explain about 90% of the variability in the NO_x data.

The best-fit linear model for ON had one significant ($p < 0.05$) predictor, barren (rock/sand/clay) land cover. The intercept was also significant. The shrub/scrub predictor variable had a p-value of 0.20. Overall, the model was significant with a p-value of 0.001. The model also had a high R-squared value, of 0.8482. This indicates that the model is able to explain about 85% of the variability in the ON data.

When the NLCD 2001 was used, both NO_x and ON had potential linear land cover predictors based on the pair plots. The linear model for NO_x using the NLCD 2001 had six predictors: high green loss; > 30% impervious surface; paved; developed, lawn/grass; developed, low intensity; and woody wetland. The linear model for NO_x using the NLCD 2006 had two predictors: paved and developed, lawn/grass. The NLCD 2006 model did not include woody wetlands as a possible parameter because it was not present in at least six of the monitoring point watersheds. Neither of the prediction models for NO_x had significant variables ($p < 0.05$).

The linear model for ON using the NLCD 2001 had three predictor variables: barren (rock/sand/clay); shrub/scrub; and grassland/herbaceous. The linear model for ON using the NLCD 2006 had two of those same predictor variables: barren (rock/sand/clay) and shrub/scrub. In both models, the barren (rock/sand/clay) land cover predictor was significant ($p < 0.05$). In the NLCD 2001 model, the shrub/scrub land cover predictor variable was significant, but it was not significant in the NLCD 2006 model, with a p-value of 0.20. Overall, both the NLCD 2001 and

NLCD 2006 water quality prediction models were significant with p-values < 0.05 . Between 2001 and 2006 the percent area of grassland/herbaceous decreased throughout the watersheds. This may be cause of the grassland/herbaceous predictor not appearing in the NLCD 2006 best-fit linear model. In addition, the grassland/herbaceous land cover type was not significant in the NLCD 2001 model.

DISCUSSION

Updating the help menus, documentation, parameters and tool interface allow the Water Quality Prediction by Land Cover Change tool to become a true SDSS. Using this tool, managers are able to explore how their land development decisions will affect the concentration of organic nitrogen in the tributary creeks on base. Managers are able to name and save the outputs of several different scenarios for land development. Because the outputs are saved, without overwriting the previous output, managers can compare the effects of different land development scenarios. Equipping managers with the knowledge about the effects of multiple scenarios allows the manager to make a more informed decision about land development.

Adding additional water quality data reduced the number of water quality parameters that are able to be predicted by percent land cover in a watershed. Because the data from the original analysis was from a one-year period, the water quality samples may not have captured a representative picture of water quality. Adding additional water quality data allows for longer-range trends to be modeled.

When the land cover was updated between 2001 and 2006, the same water quality parameters were able to be predicted, NO_x and ON. In both cases, NO_x did not have any significant predictor variables, however the models were significant overall. In both cases, ON had barren (rock/sand/clay) as a significant predictor. This land cover type may be playing a role in the concentration of organic nitrogen in the water.

NEXT STEPS

There are many improvements that still can be made to make this tool an even more powerful DSS. I recommend the following next steps for tool development: making the tool available online; coordination with water quality modelers to input more predictive water quality models into the tool; create a mechanism that automatically updates the tool based on most up-to-date

data; and add flow rates to the model so that nutrient loads are the output instead of nutrient concentrations.

An online tool would allow a user without GIS on their computer to perform this analysis. This would eliminate the need to have the correct version of ArcGIS and appropriate extensions for tool functionality. The Marine Geospatial Ecology Lab at Duke University is currently working on a GIS server tool prototype, incorporating the work of this project.

The models used in this analysis to predict changes in water nutrient concentration were simple linear models. DCERP has a team of water quality modelers who have been working on more sophisticated water quality prediction models based on more extensive data collection. These models could be incorporated into the interface of the tool to create a more predictive decision support tool.

The DCERP - MARDIS website is a clearinghouse of data that is updated periodically by the researcher who collected the data. A mechanism could be built that would automatically query the most up-to-date water quality data from the MARDIS information system for use in the online SDSS tool. The mechanism would also perform an automated statistical model analysis and incorporate the best fit models into the tool. This mechanism would allow the tool to reflect the most recent water quality data.

Finally, it would be interesting to model loads of nutrients instead of concentrations. Currently, the water quality prediction tool only produces the change in concentration for the water quality parameters. Including the change in load would be a much more useful measure for decision makers. Water quality for streams often must meet Total Maximum Daily Loads (TMDLs). Decision makers would be able to see how their development plans would affect the TMDLs.

CONCLUSION

Currently, Marine Corps Base Camp Lejeune is undergoing rapid construction. As the base continues to develop, it is critical for managers to understand the impact that developing land has on the water quality of the surrounding streams. Decision Support Tools, such as the Water Quality Prediction by Land Cover tool, can help managers begin to quantify the effects of changing land cover on water quality. This tool can be used to determine where development should take place. The manager can run several alternatives scenarios to determine which land

development scenario has the least impact on water quality. Managers at Marine Corps Base Camp Lejeune can use this data driven Spatial Decision Support System to evaluate how different development scenarios will affect the concentration of organic nitrogen in the tributary creeks on base.

The ultimate goal for this project was to aid in the development of a meaningful data driven tool that can help protect water quality. Tools like this one can help decision makers make more informed decisions about how land is used in order to protect our water quality. This tool has practical applications that extend beyond MCBCL. Water quality prediction models, modeled in other geographies, can be input into the tool to predict concentration change in water quality. Broad-scale adoption of decision support tools that link land development to water quality could improve the way decision makers make choices about land development.

REFERENCES

- Akaike, Hirotugu (1974). "A new look at the statistical model identification". *IEEE Transactions on Automatic Control* **19** (6): 716–723.
- Bendoly, Elliot "Excel Basics to Blackbelt: an Accelerated Guide to Decision Support Designs." Cambridge ; New York : Cambridge University Press.
- Bolstad, P.V. and W.T. Swank (1997). "Cumulative impacts of landuse on water quality in a southern Appalachian watershed." Journal of the American Water Resources Association 33:519-533.
- FINRMP (2006). Integrated Natural Resources Management Plan (2007-2011) Marine Corps Base Camp Lejeune, Onslow County, North Carolina, Marine Corps.
- Laudien, R, M. Pofagi, and J Roehrig (2010). "Development and implementation of an interactive Spatial Decision Support System for decision makers in Benin to evaluate agricultural land resources--Case study: AGROLAND, International Journal of Applied Earth Observation and Geoinformation." Supplement Issue on "Remote Sensing for Africa - A Special Collection from the African Association for Remote Sensing of the Environment (AARSE) 12:S38-S44.
- Minter, Thomas (2009). "Tributary creek water quality influenced by land use characteristics at Marine Corps Base Camp Lejeune, NC." <<http://hdl.handle.net/10161/1553>>
- Roberts, J.J., B.D. Best, D.C. Dunn, E.A. Treml, and P.N. Halpin (2010). "Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++." Environmental Modelling & Software 25: 1197-1207.

RTI International (2011). "Defense Coastal/Estuarine Research Program."
<<https://dcerp.rti.org/>>

Tong, S. T. Y. and W. Chen (2002). "Modeling the relationship between land use and surface water quality." Journal of Environmental Management 66:377-393.

LIST OF TABLES

Table 1. Locations of monitoring sites

Site	Lat	Long
Freeman Creek	34.60375	-77.26669
Gillets Creek	34.58889	-77.28497
Cogdel Creek	34.65511	-77.33428
French Cr HW	34.63631	-77.30086
Courthouse Bay	34.583	-77.36517
Trapps Creek	34.58178	-77.33311
Airport Creek	34.70472	-77.45342
Southwest Creek HW	34.71642	-77.46553
Tarawa Terrace	34.73814	-77.38186
Camp Johnson	34.73661	-77.40089

Table 2. Measured nutrient concentrations at monitoring stations throughout Camp Lejeune

Nutrient Concentration	Units
Ammonium (NH ₄)	uM
Nitrate/nitrate (NO _x)	uM
Organic Nitrogen (ON)	uM
Phosphate (PO ₄ ³⁻)	uM
Total Nitrogen (TN)	uM
Total Suspended Solids (TSS)	mg/L

Table 3. Number of Samples Taken at Each Station 12/11/2007-12/13/2010

n samples		Parameter					
		DON	NH4	NO3/NO2	PO4	TDN	TSS
Station	CAJO	108	108	108	108	108	120
	NRAS	39	39	39	39	39	45
	COGD	243	247	247	247	243	234
	COURT	117	118	118	118	117	140
	FREE	136	136	136	136	136	128
	FRHW	236	238	238	238	236	225
	GILL	228	230	230	230	228	221
	SWCR	44	46	46	46	46	52
	TT	116	117	117	117	116	129
	TRAPP	97	98	98	98	97	111

Table 4. List of Land Use/Land Cover predictors defined by NLCD 2001, MCBCL 2004 and others.

Land Use Characteristics		
NLCD 2001	Additional	MCBCL 2004
Watershed with greater than 30% Impervious Surface	High Green Loss	Predominant Pine Forest
Developed, Lawn Grass	Unpaved	Mixed Pine/Hardwood Forest
Developed, Low Intensity	Paved	Bottomland Hardwood Forest
Developed, Med Intensity	BaB (Soil Series)	Business/Commercial
Barren (Rock, Sand, Clay)	MaC (Soil Series)	Barren
Deciduous Forest		
Evergreen Forest		
Mixed Forest		
Shrub/Scrub		
Grassland/Herbaceous		
Woody Wetland		
Emergent Herbaceous Wetland		

Table 5. R model written by Ben Best

```
ys = colnames(dat[,1:7]); n.y =
length(ys)
xs = colnames(dat[,8:ncol(dat)]); n.x =
length(xs)
cols = character(0)
for (y in ys){
  cols = c(cols, sprintf('%s.p',y),
  sprintf('%s.B',y))
}
p.xs = matrix(rep(NA, n.x * n.y * 2), nrow=n.x, ncol=n.y*2,
dimnames=list(xs, cols))
for (y in ys){ # y = ys[1]
  for (x in xs){ # x = xs[1]
    p = sprintf('%s.p',y)
    B = sprintf('%s.B',y)
    dat.lm = dat[,c(y,x)]
    if (sum(dat[,x])==0){
      p.xs[
        x,p] = NA
      p.xs[
        x,B] = NA
    } else {
      dat.lm =
data.frame(y=dat[,y],x=dat[,x])
      lm.xy = summary(lm(y~x,
dat.lm))
      p.xs[x,p] =
lm.xy$coefficients['x', 'Pr(>|t|)']
      p.xs[x,B] =
lm.xy$coefficients['x', 'Estimate']
    }
  }
  print(y)
  o =
order(p.xs[,sprintf('%s.p',y)])
  print(head(p.xs[o, c(p,B)],
10))
}
write.csv(p.xs, csv.out)
```

Table 6. Automated linear model results using average nutrient concentrations. Significant predictors are in red.

	TSS.p	TSS.B	Nox.p	Nox.B	NH4.p	NH4.B	PO3.p	PO3.B	TN.p	TN.B	ON.p	ON.B	Chla.p	Chla.B
X..Light.Green.Loss	0.01	-9.58	0.00	-4.90	0.44	-1.96	0.24	-0.41	0.88	-1.09	0.25	5.75	0.53	-0.21
X..Medium.Green.Loss	0.33	-9.29	0.65	-2.05	0.09	8.57	0.85	0.15	0.04	27.03	0.03	20.48	0.42	-0.58
X..High.Green.Loss	0.20	7.22	0.02	5.43	0.70	1.31	0.99	0.00	0.83	-1.95	0.18	-8.69	0.72	-0.17
X..watershed.with.greater.than.30..Impervious.Surface	0.17	3.13	0.02	2.19	0.61	0.69	0.82	-0.04	0.89	-0.52	0.19	-3.41	0.76	-0.06
X..Unpaved	0.39	113.03	0.40	-51.96	0.22	89.10	0.43	8.26	0.61	104.89	0.65	68.89	0.06	17.38
X.Paved	0.02	217.18	0.00	130.52	0.14	86.89	0.23	10.17	0.45	127.16	0.47	-89.77	0.94	-0.69
X.Developed..Lawn.Grass	0.30	3.48	0.03	3.03	0.11	2.88	0.95	0.02	0.44	4.12	0.65	-1.80	0.97	-0.01
X..Developed..Low.Intensity	0.06	4.09	0.00	2.75	0.07	2.25	0.58	0.11	0.14	5.24	0.93	0.23	0.97	0.01
X.Developed..Med.Intensity	0.82	1.52	0.43	2.41	0.91	-0.41	0.38	-0.46	0.31	-10.27	0.07	-12.29	0.85	-0.09
X.Barren..Rock..Sand..Clay.	0.26	-10.01	0.31	-4.20	0.57	2.95	0.56	-0.43	0.11	21.12	0.01	22.28	0.66	-0.31
X.Deciduous.Forest	0.73	6.35	0.49	5.90	0.27	-11.10	0.23	-1.70	0.09	-43.75	0.03	-38.65	0.99	-0.02
X.Evergreen.Forest	0.93	0.15	0.86	-0.15	0.80	-0.26	0.33	0.14	0.35	-2.61	0.28	-2.17	0.05	0.24
X.Mixed.Forest	0.74	-3.52	0.74	-1.65	0.56	-3.45	0.73	-0.29	0.09	-24.86	0.06	-19.72	0.04	1.43
X.Shrub.Scrub	0.23	-7.28	0.07	-4.80	0.95	-0.23	0.31	-0.50	0.31	9.70	0.01	14.65	0.92	0.05
X.Grassland.Herbaceous	0.28	-3.02	0.12	-1.95	0.55	-0.98	0.51	-0.15	0.55	2.69	0.06	5.59	0.33	-0.21
X.Woody.Wetland	0.13	-3.09	0.01	-2.09	0.24	-1.39	0.79	-0.05	0.70	-1.28	0.37	2.19	0.09	-0.25
X.Emergent.Herbaceous.Wetland	0.76	-6.04	0.50	-6.13	0.34	-10.42	0.95	0.10	0.76	-9.46	0.75	7.10	0.38	-1.27
X.BaB	0.21	-2.03	0.75	-0.26	0.07	1.59	0.32	-0.13	0.39	2.24	0.64	0.90	0.12	0.19
X.MaC	0.45	-5.14	0.54	-2.00	0.44	-2.99	0.15	-0.75	0.07	-17.45	0.08	-12.48	0.40	-0.44
X.Predominant.Pine.Forest	0.65	-0.44	0.19	-0.57	0.54	-0.33	0.65	0.03	0.41	-1.21	0.79	-0.30	0.72	0.03
X.Mixed.Pine.Hardwood.Forest	0.47	-1.35	0.75	-0.28	0.42	0.86	0.57	-0.09	0.09	4.59	0.03	3.99	0.88	0.02
X.Bottomland.Hardwood.Forest	0.40	-3.46	0.82	-0.45	0.39	-2.03	0.67	-0.14	0.13	-9.09	0.14	-6.59	0.38	-0.27
X.Business.Commercial	0.93	-0.11	0.24	0.68	0.10	1.11	0.79	-0.03	0.53	1.24	0.70	-0.56	0.51	0.06
X.Barren	0.51	-1.55	0.35	-1.02	0.34	-1.26	0.70	-0.07	0.81	-0.91	0.62	1.35	0.59	-0.10
X.of.Cells.with...1000.flow.accumulation	0.47	139.73	0.28	95.66	0.01	230.71	0.33	14.81	0.00	777.34	0.01	450.44	0.68	6.16

Table 7. GLM summaries for NOx

```

Call:
glm(formula = Nox ~ X.Paved + X.Developed.Low.Intensity + X.Woody.Wetland +
  X.watershed.with.greater.than.30.Impervious.Surface + X.High.Green.Loss +
  X.Developed.Lawn.Grass, family = gaussian, data = dat)

Deviance Residuals:
    AirStation    Camp Johnson      Codgels Courthouse Bay      Freeman      French
Gilletts          3.687          10.595          -9.324          -14.736          -8.068          3.033
5.402          -10.500          13.687
    Trapps
    6.223

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   21.4842    19.1298    1.123    0.343
X.Paved                       106.4001    51.2654    2.075    0.130
X.Developed.Low.Intensity      1.6093     1.3325    1.208    0.314
X.Woody.Wetland                -0.6184     0.6243   -0.991    0.395
X.watershed.with.greater.than.30.Impervious.Surface -1.5503     1.5678   -0.989    0.396
X.High.Green.Loss              3.5050     3.6858    0.951    0.412
X.Developed.Lawn.Grass        -2.0799     1.6299   -1.276    0.292

(Dispersion parameter for gaussian family taken to be 289.9045)

Null deviance: 13834.33 on 9 degrees of freedom
Residual deviance: 869.71 on 3 degrees of freedom
AIC: 89.035

Number of Fisher Scoring iterations: 2

```

Table 8. GLM summaries for TSS

```

Call:
glm(formula = TSS ~ X.Light.Green.Loss + X.Paved, family = gaussian,
  data = dat)

Deviance Residuals:
    Min      1Q  Median      3Q     Max
-69.805 -45.280  9.969  26.716  99.465

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   166.796    139.029    1.200    0.269
X.Light.Green.Loss             -6.951     4.906   -1.417    0.199
X.Paved                        80.351    120.252    0.668    0.525

(Dispersion parameter for gaussian family taken to be 3440.407)

Null deviance: 62604 on 9 degrees of freedom
Residual deviance: 24083 on 7 degrees of freedom
AIC: 114.25

Number of Fisher Scoring iterations: 2

```

Table 9. GLM summaries for TN

```

Call:
glm(formula = TN ~ X..Medium.Green.Loss, family = gaussian, data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-134.27  -54.74  -10.80   52.70  186.37

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)         54.09    146.96   0.368  0.7224
X..Medium.Green.Loss  27.03     11.20   2.413  0.0423 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 10946.83)

Null deviance: 151329  on 9  degrees of freedom
Residual deviance:  87575  on 8  degrees of freedom
AIC: 125.16

Number of Fisher Scoring iterations: 2

```

Table 10. GLM summaries for ON

```

Call:
glm(formula = ON ~ X..Medium.Green.Loss + X.Barren..Rock..Sand..Clay. +
  X.Deciduous.Forest + X.Mixed.Pine.Hardwood.Forest, family = gaussian,
  data = dat)

Deviance Residuals:
    AirStation   Camp Johnson      Codgels  Courthouse Bay      Freeman      French
Gilletts
    -7.519      23.435     -25.064         -3.807     -19.000     -1.049
30.556
    -21.558      20.522
    Trapps
     3.483

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    203.0691    48.4212   4.194  0.00854 **
X..Medium.Green.Loss      5.2667     3.6455   1.445  0.20815
X.Barren..Rock..Sand..Clay.  8.2253     3.6006   2.284  0.07114 .
X.Deciduous.Forest    -23.4984     6.6898  -3.513  0.01705 *
X.Mixed.Pine.Hardwood.Forest  3.0898     0.5931   5.210  0.00344 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 688.4455)

Null deviance: 81549.5  on 9  degrees of freedom
Residual deviance:  3442.2  on 5  degrees of freedom
AIC: 98.792

Number of Fisher Scoring iterations: 2

```

Table 11. Linear Model summaries for NOx

```

Call:
lm(formula = dat.NoX$Nox ~ dat.NoX$X.High.Green.Loss +
dat.NoX$X.watershed.with.greater.than.30.Impervious.Surface +
dat.NoX$X.Paved + dat.NoX$X.Developed.Lawn.Grass + dat.NoX$X.Developed.Low.Intensity +
dat.NoX$X.Woody.Wetland)

Residuals:
    1     2     3     4     5     6     7     8     9    10
 3.687 10.595 -9.324 -14.736 -8.068  3.033  5.402 -10.500 13.687  6.223

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)          21.4842    19.1298   1.123  0.343
dat.NoX$X.High.Green.Loss      3.5050     3.6858   0.951  0.412
dat.NoX$X.watershed.with.greater.than.30.Impervious.Surface -1.5503     1.5678  -0.989  0.396
dat.NoX$X.Paved             106.4001    51.2654   2.075  0.130
dat.NoX$X.Developed.Lawn.Grass -2.0799     1.6299  -1.276  0.292
dat.NoX$X.Developed.Low.Intensity  1.6093     1.3325   1.208  0.314
dat.NoX$X.Woody.Wetland      -0.6184     0.6243  -0.991  0.395

Residual standard error: 17.03 on 3 degrees of freedom
Multiple R-squared:  0.9371,    Adjusted R-squared:  0.8114
F-statistic: 7.453 on 6 and 3 DF,  p-value: 0.06388

```

Table 12. Linear Model summaries for ON

```

Call:
lm(formula = dat.ON$ON ~ dat.ON$X.Barren.Rock.Sand.Clay.+
dat.ON$X.Shrub.Scrub + dat.ON$X.Grassland.Herbaceous)

Residuals:
    Min       1Q   Median       3Q      Max
-46.977 -25.903  -3.896   28.432   51.378

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)          194.386    25.174   7.722 0.000248 ***
dat.ON$X.Barren.Rock.Sand.Clay.    32.234     9.199   3.504 0.012760 *
dat.ON$X.Shrub.Scrub             11.614     3.435   3.381 0.014841 *
dat.ON$X.Grassland.Herbaceous     -6.363     3.050  -2.086 0.082036 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.68 on 6 degrees of freedom
Multiple R-squared:  0.8722,    Adjusted R-squared:  0.8083
F-statistic: 13.65 on 3 and 6 DF,  p-value: 0.00434383
F-statistic: 13.65 on 3 and 6 DF,  p-value: 0.004343

```

Table 13. Linear Model summaries for ON with grassland term dropped

```

Call:
lm(formula = dat.ON$ON ~ dat.ON$X.Barren.Rock.Sand.Clay.+
dat.ON$X.Shrub.Scrub)

Residuals:
    Min       1Q   Median       3Q      Max
-57.56 -30.67 -16.66   30.71   85.88

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)          191.475    30.567   6.264 0.000418 ***
dat.ON$X.Barren.Rock.Sand.Clay.    15.805     5.780   2.734 0.029153 *
dat.ON$X.Shrub.Scrub             9.517     3.995   2.382 0.048703 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50.69 on 7 degrees of freedom
Multiple R-squared:  0.7795,    Adjusted R-squared:  0.7165
F-statistic: 12.37 on 2 and 7 DF,  p-value: 0.005036

```

Table 14. Percent land cover based on the National Land Cover Dataset 2001

Percent Land Cover	Watershed Site									
	Camp Johnson	Tarawa Terrace	Southwest Creek HW	Airport Creek	Cogdel Creek	French Creek HW	Freeman Creek	Gillets Creek	Trapps Creek	Courthouse Bay
Open Water	0.00%	0.00%	0.00%	0.00%	0.84%	0.00%	0.09%	0.07%	0.00%	0.00%
Developed, Open Space	8.63%	22.63%	0.00%	20.21%	9.30%	2.40%	7.88%	5.58%	5.44%	24.70%
Developed, Low Intensity	15.99%	34.76%	0.80%	20.02%	8.43%	2.69%	4.37%	7.40%	6.83%	33.58%
Developed, Medium Intensity	3.57%	4.56%	0.00%	10.54%	11.68%	0.09%	0.00%	0.00%	0.00%	3.96%
Developed, High Intensity	0.00%	0.98%	0.00%	18.85%	6.34%	0.00%	0.00%	0.00%	0.00%	0.00%
Barren Land (Rock/Sand/Clay)	0.00%	0.00%	0.00%	0.00%	2.57%	9.60%	3.01%	4.11%	3.07%	6.92%
Deciduous Forest	1.71%	1.80%	3.81%	4.24%	2.11%	0.12%	1.72%	1.40%	0.00%	0.00%
Evergreen Forest	37.88%	17.26%	36.79%	21.49%	26.99%	10.99%	24.17%	14.33%	7.96%	11.19%
Mixed Forest	8.51%	0.00%	0.00%	2.68%	5.27%	0.80%	0.28%	0.31%	0.35%	0.00%
Shrub/Scrub	5.33%	2.26%	0.26%	0.08%	4.43%	7.47%	14.18%	12.96%	8.58%	8.40%
Grassland/Herbaceous	2.37%	2.75%	4.03%	0.00%	7.56%	34.66%	16.86%	18.97%	14.05%	11.25%
Pasture/Hay	0.35%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Cultivated Crops	5.65%	12.20%	1.49%	0.00%	3.51%	0.00%	0.00%	0.00%	0.00%	0.00%
Woody Wetlands	10.01%	0.09%	52.82%	1.89%	9.80%	26.06%	27.23%	34.16%	53.29%	0.00%
Emergent Herbaceous Wetlands	0.00%	0.73%	0.00%	0.00%	1.17%	5.12%	0.21%	0.72%	0.44%	0.00%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 15. Percent land cover based on the National Land Cover Dataset 2006

Percent Land Cover	Watershed Site									
	Camp Johnson	Tarawa Terrace	Southwest Creek HW	Airport Creek	Cogdel Creek	French Creek HW	Freeman Creek	Gillets Creek	Trapps Creek	Courthouse Bay
Open Water	0.00%	0.00%	0.00%	0.00%	1.13%	0.19%	0.11%	0.00%	0.00%	0.00%
Developed, Open Space	11.54%	34.58%	0.00%	22.78%	8.78%	2.34%	7.72%	5.68%	4.12%	22.61%
Developed, Low Intensity	16.64%	34.23%	0.86%	20.15%	8.60%	2.76%	4.28%	7.41%	7.72%	34.84%
Developed, Medium Intensity	3.95%	5.98%	0.00%	10.35%	11.99%	0.05%	0.00%	0.00%	0.00%	4.39%
Developed, High Intensity	0.00%	1.56%	0.00%	19.35%	6.08%	0.00%	0.00%	0.00%	0.00%	0.00%
Barren Land (Rock/Sand/Clay)	0.00%	0.00%	0.00%	0.00%	1.96%	4.51%	2.79%	3.87%	3.26%	7.64%
Deciduous Forest	0.00%	0.00%	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%
Evergreen Forest	39.03%	19.84%	55.53%	27.33%	30.80%	9.10%	24.79%	21.01%	13.15%	12.05%
Mixed Forest	7.83%	0.00%	3.65%	0.00%	3.82%	0.88%	0.57%	0.54%	0.00%	0.00%
Shrub/Scrub	5.28%	3.81%	3.20%	0.00%	12.53%	27.43%	28.37%	24.70%	15.50%	18.43%
Grassland/Herbaceous	0.00%	0.00%	2.48%	0.00%	3.74%	21.30%	5.81%	7.16%	5.87%	0.04%
Cultivated Crops	4.45%	0.00%	0.00%	0.00%	0.99%	3.20%	2.66%	1.96%	0.00%	0.00%
Woody Wetlands	11.28%	0.00%	34.28%	0.04%	8.92%	21.75%	22.74%	26.97%	48.39%	0.00%
Emergent Herbaceous Wetlands	0.00%	0.00%	0.00%	0.00%	0.65%	6.44%	0.17%	0.71%	1.99%	0.00%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Table 16. Linear Model summaries for NOx with 2006 NLCD predictors

```

Call:
lm(formula = dat.NoX$Nox ~ dat.NoX$X.Paved + dat.NoX$X.Developed..Lawn.Grass)

Residuals:
    Min       1Q   Median       3Q      Max
-16.6375 -11.8287  0.7884  11.2036  16.0174

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -4.0550     7.4085  -0.547  0.6012
dat.NoX$X.Paved  82.1425    27.8107   2.954  0.0213 *
dat.NoX$X.Developed..Lawn.Grass  1.5239     0.6844   2.227  0.0613 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.21 on 7 degrees of freedom
Multiple R-squared:  0.8978,    Adjusted R-squared:  0.8686
F-statistic: 30.74 on 2 and 7 DF,  p-value: 0.0003416
    
```

Table 17. Linear Model summaries for ON with 2006 NLCD predictors

```

Call:
lm(formula = dat.ON$ON ~ dat.ON$X.Barren..Rock..Sand..Clay. +
  dat.ON$X.Shrub.Scrub)

Residuals:
    Min       1Q   Median       3Q      Max
-53.784 -33.081   9.922  26.720  47.404

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      210.666     22.673   9.291 3.47e-05 ***
dat.ON$X.Barren..Rock..Sand..Clay.    25.146     8.174   3.076 0.0179 *
dat.ON$X.Shrub.Scrub         2.738     1.954   1.401 0.2040
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 42.06 on 7 degrees of freedom
Multiple R-squared: 0.8482,    Adjusted R-squared: 0.8048
F-statistic: 19.55 on 2 and 7 DF,  p-value: 0.001364

```

LIST OF FIGURES

Figure 1. Location of Marine Corps Base Camp Lejeune (FINRMP 2006).

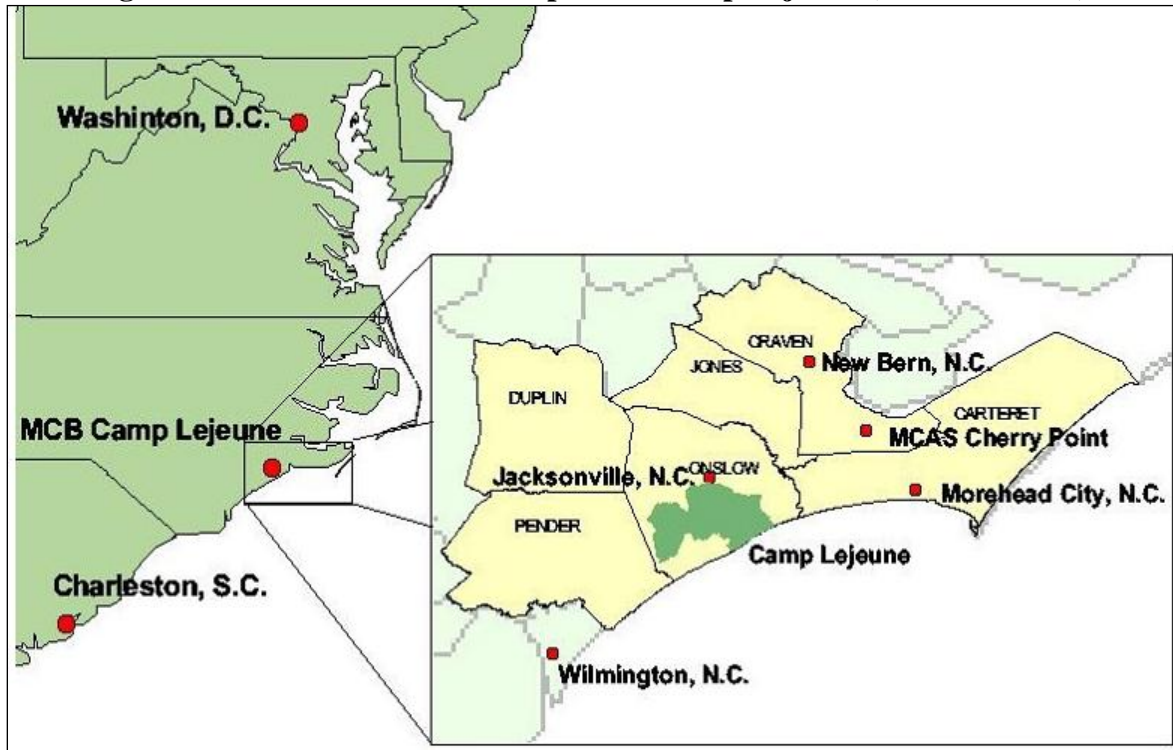


Figure 2. Map of monitoring sites

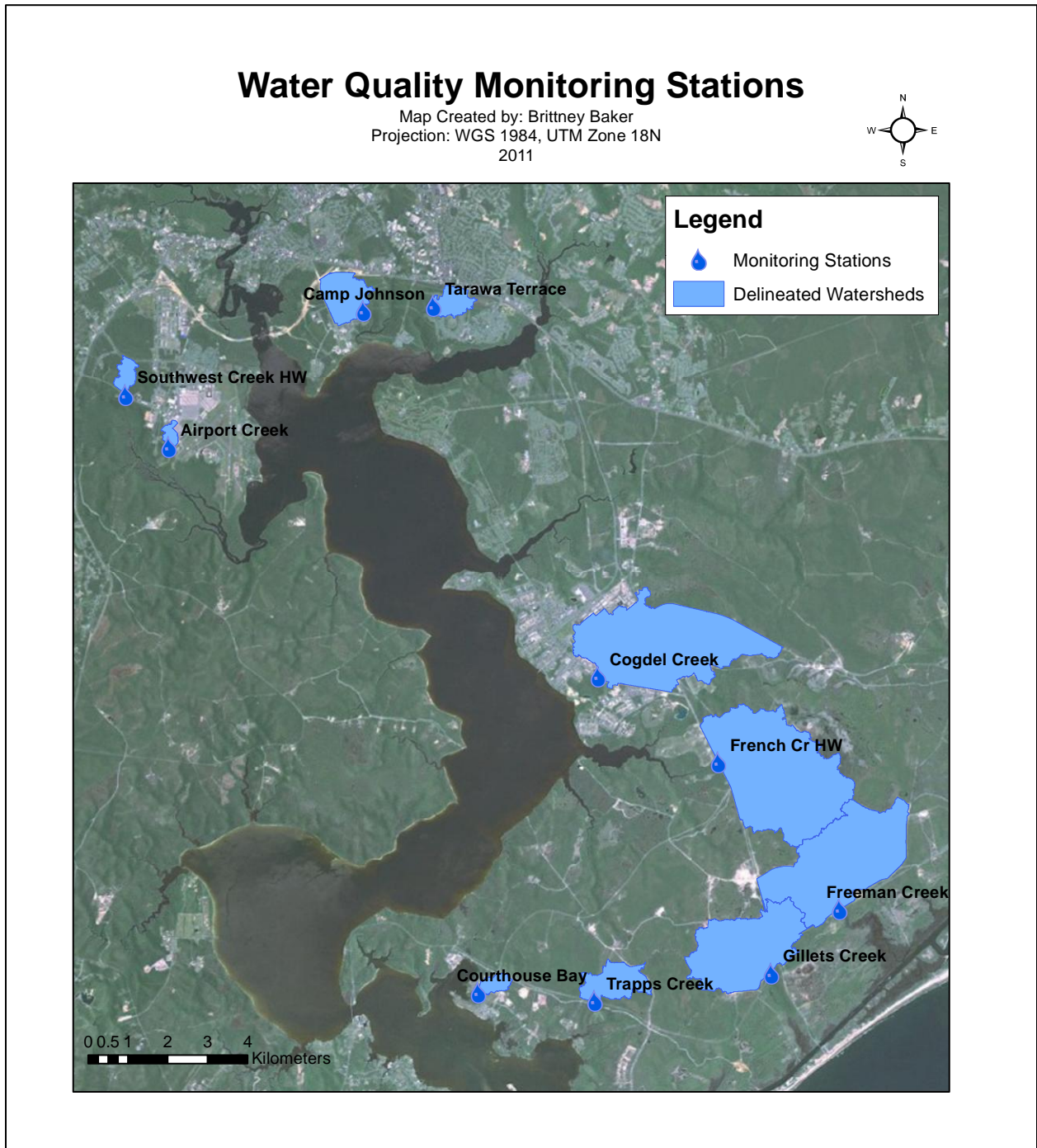


Figure 3. Number of water quality samples collected

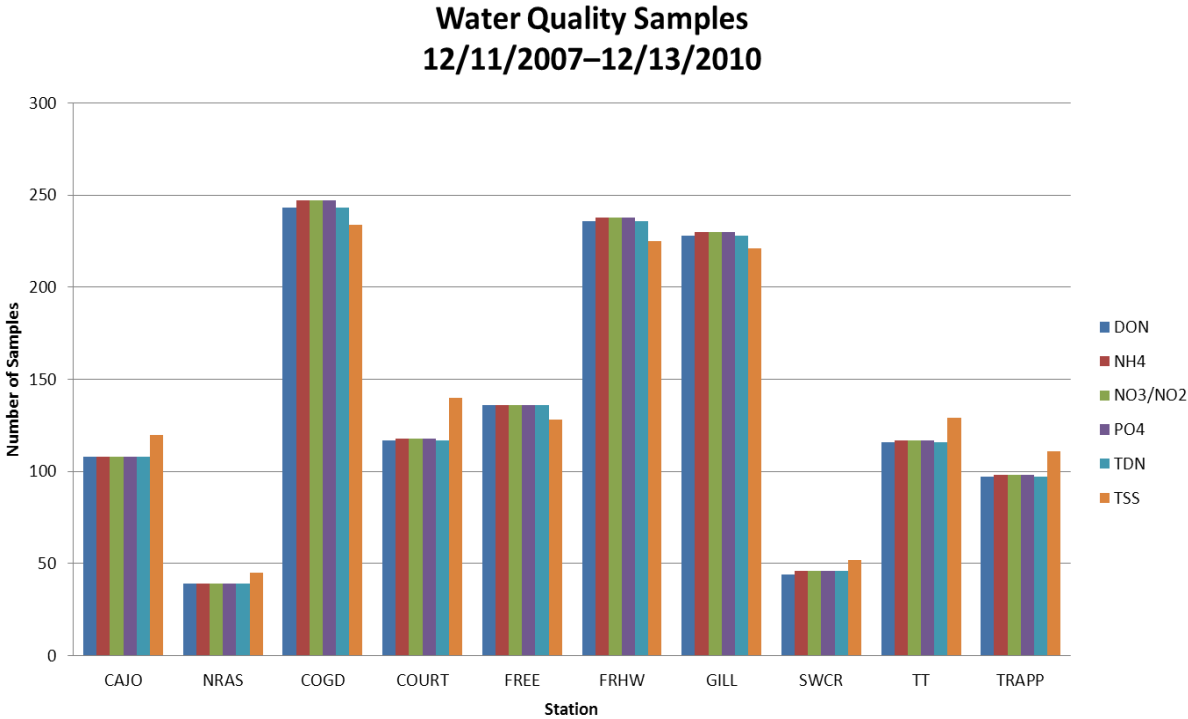


Figure 4. In tool Help update -- Working Directory

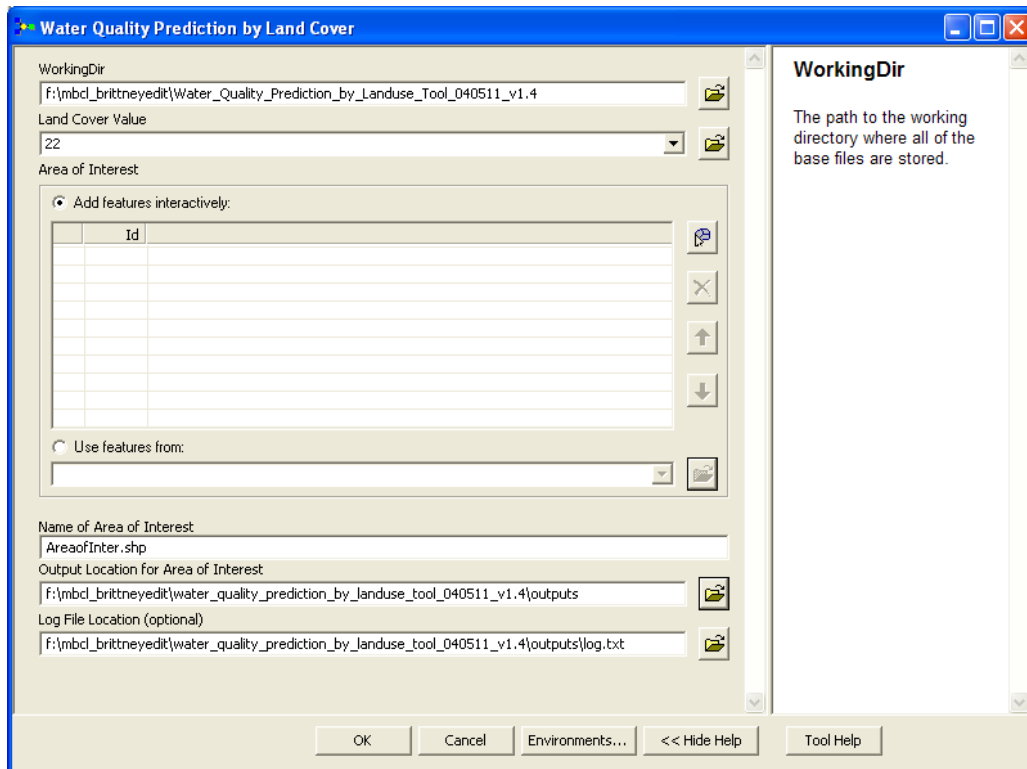
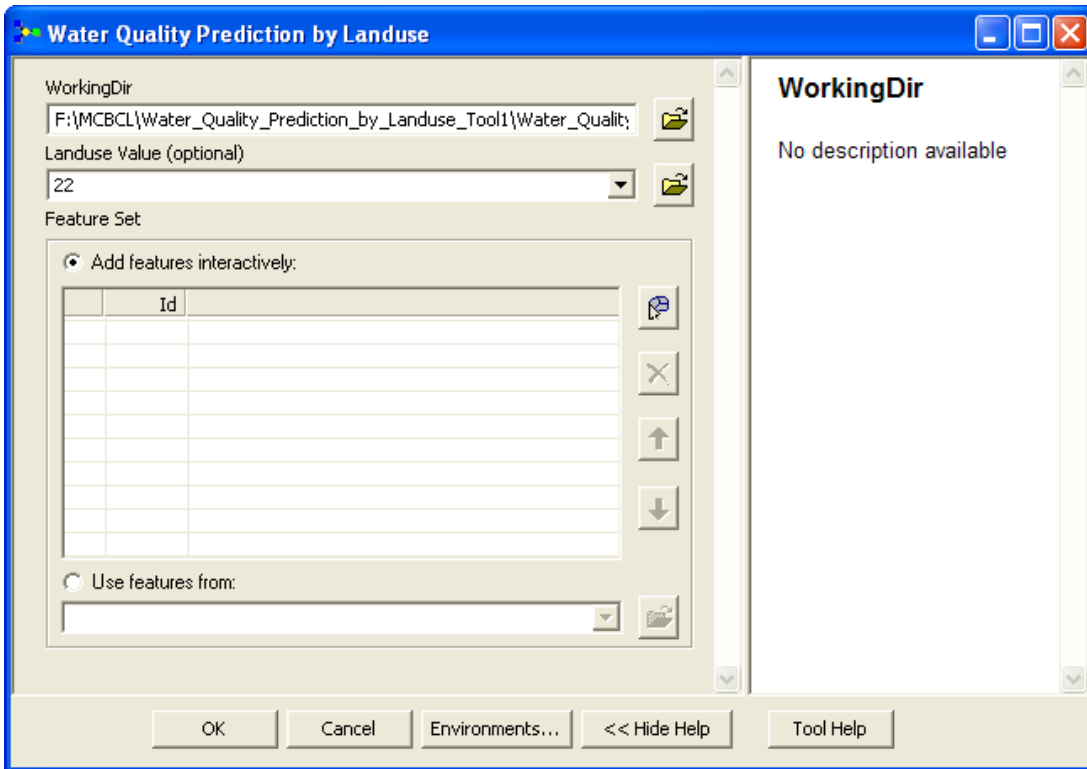


Figure 6. In tool Help update -- Feature Set/Area of Interest

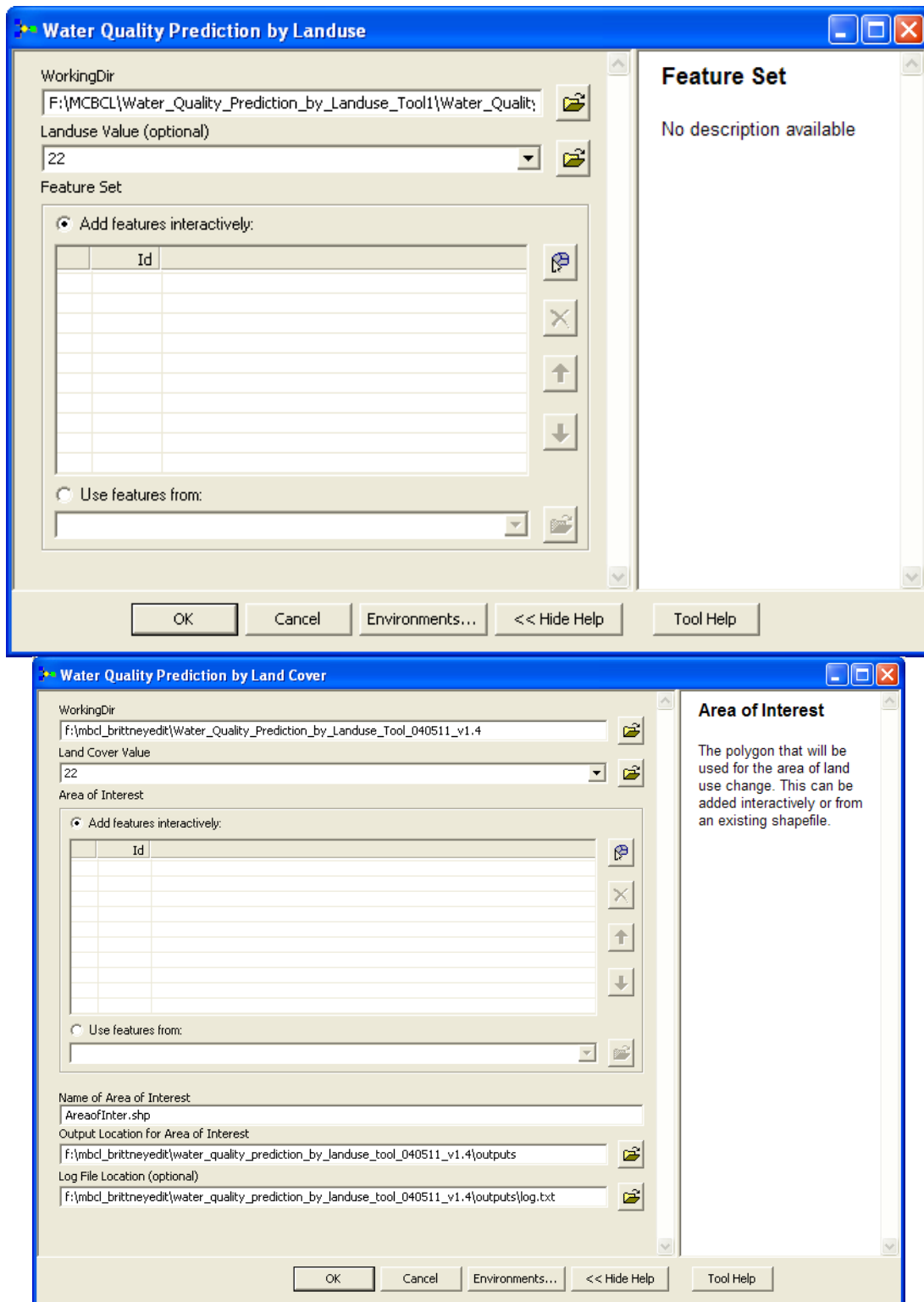


Figure 7. Name of Area of Interest

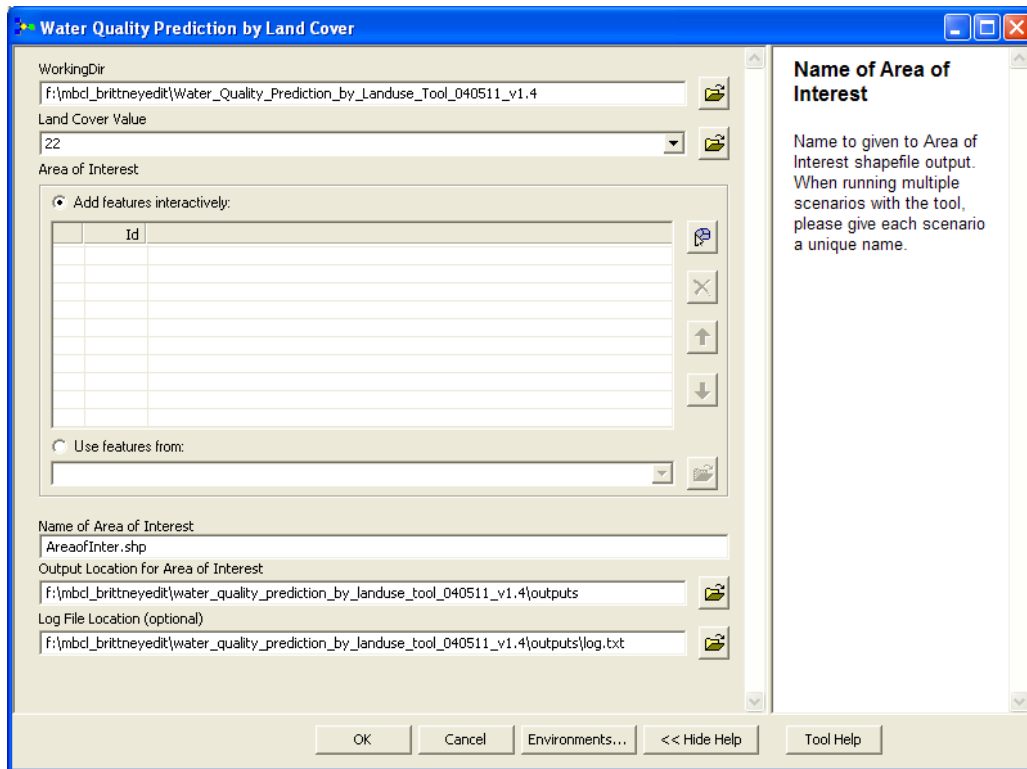


Figure 8. Output location of Area of Interest

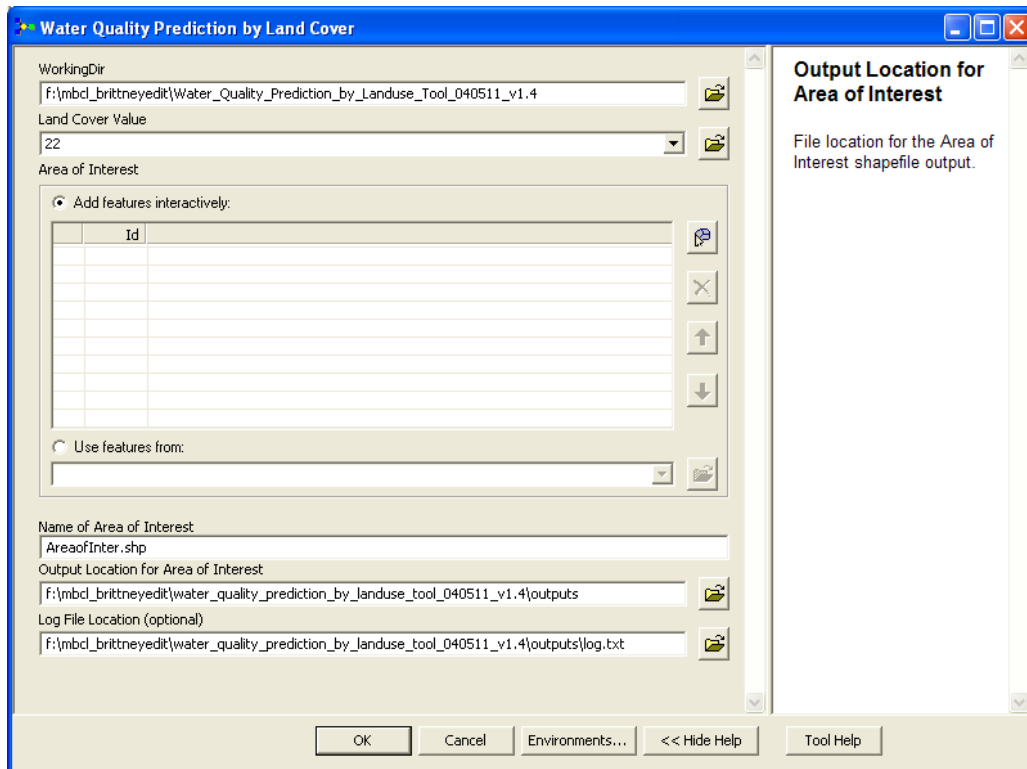
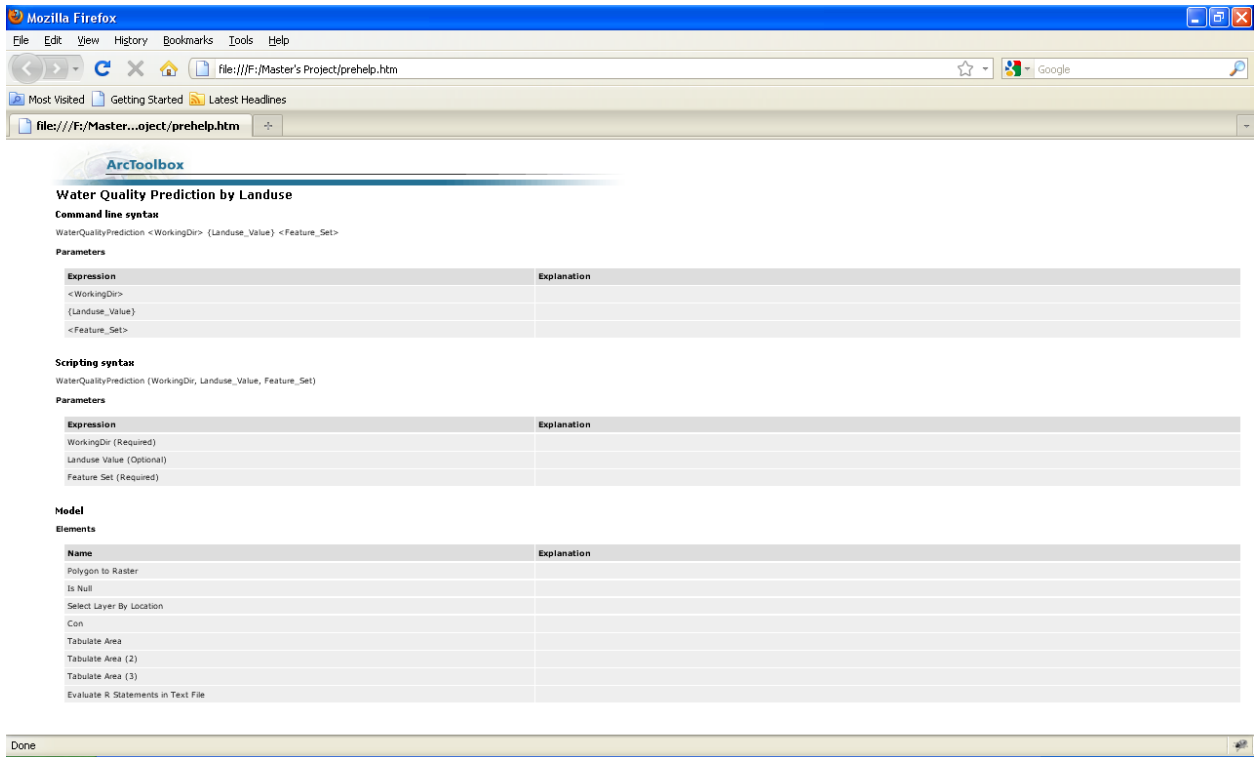


Figure 10. External tool help update



C:\Documents and Settings\bab39\Application Data\ESRI\ArcToolbox\Dig\WdTool\Help.htm - Windows Internet Explorer

C:\Documents and Settings\bab39\Application Data\ESRI\ArcToolbox\Dig\WdTool\Help.htm

Google

C:\Documents and Settings\bab39\Application Data\ESRI\ArcToolbox\Dig\WdTool\Help.htm

Page Safety Tools

ArcToolbox

Water Quality Prediction by Land Cover

The purpose of this tool is to determine how the concentration of organic nitrogen will change in surrounding streams when land use is changed.

Command line syntax

```
WaterQualityPrediction <WorkingDir> <Land_Cover_Value> <Area_of_Interest> <Name_of_Area_of_Interest> <Output_Location_for_Area_of_Interest> {Log_File_Location}
```

Parameters

Expression	Explanation
<WorkingDir>	The path to the working directory where all of the base files are stored.
<Land_Cover_Value>	The proposed Landuse Value based on the 2001 NLCD
<Area_of_Interest>	The polygon that will be used for the area of land use change.
<Name_of_Area_of_Interest>	Name to given to Area of Interest shapefile output.
<Output_Location_for_Area_of_Interest>	File location for the Area of Interest shapefile output.
{Log_File_Location}	

Command Line Example

```
WaterQualityPrediction WorkingDir Land_Cover_Value Area_of_Interest Name_of_Area_of_Interest Output_Location_for_Area_of_Interest Log_File_Location
```

Parameters

Expression	Explanation
WorkingDir (Required)	The path to the working directory where all of the base files are stored.
Land Cover Value (Required)	The proposed Landuse Value based on the 2001 NLCD
Area of Interest (Required)	The polygon that will be used for the area of land use change.
Name of Area of Interest (Required)	Name to given to Area of Interest shapefile output.
Output Location for Area of Interest (Required)	File location for the Area of Interest shapefile output.
Log File Location (Optional)	

Script Example

Model Elements

Name	Explanation
Polygon to Raster	Converts the selected area of interest from a polygon shapefile to a raster file for further analysis.
Is Null	Gives the raster of the area of interest a value of 0 and all other areas a value of 1.
Select Layer By Location	Selects the watersheds that intersect with the user selected area of interest.
Con	Inputs the NLCD values for all areas that equal "1" (every area that is not the area of interest). Inputs the Land Cover Value (user variable) for all areas that do not equal "1" (Area of interest).
Tabulate Area	Calculates the area of NLCD/Land Cover values in the selected watershed. The output is a table.
Tabulate Area (2)	Calculates area of NLCD values (before proposed change). The output is a table.
Tabulate Area (3)	Calculates the area of each soil type. The output is a table.

Done

My Computer 70%

Figure 11. Toolbox Name changed name from “Toolbox” to “Water Quality Prediction Toolbox”

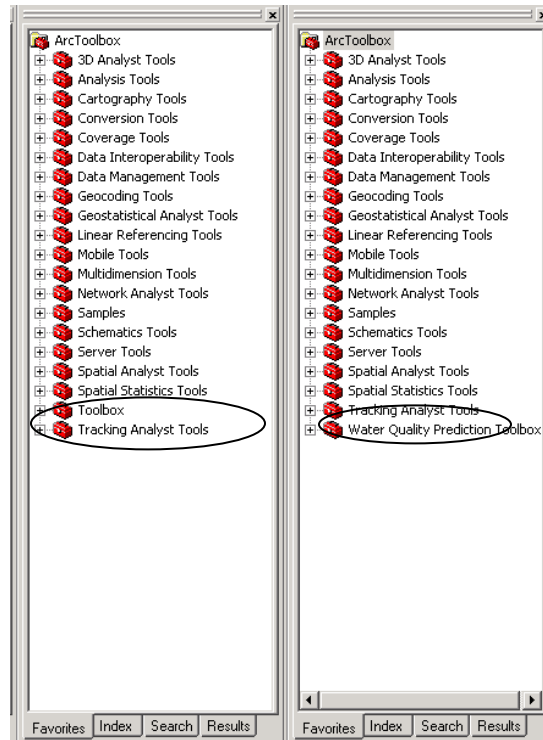


Figure 12. Tool file structure

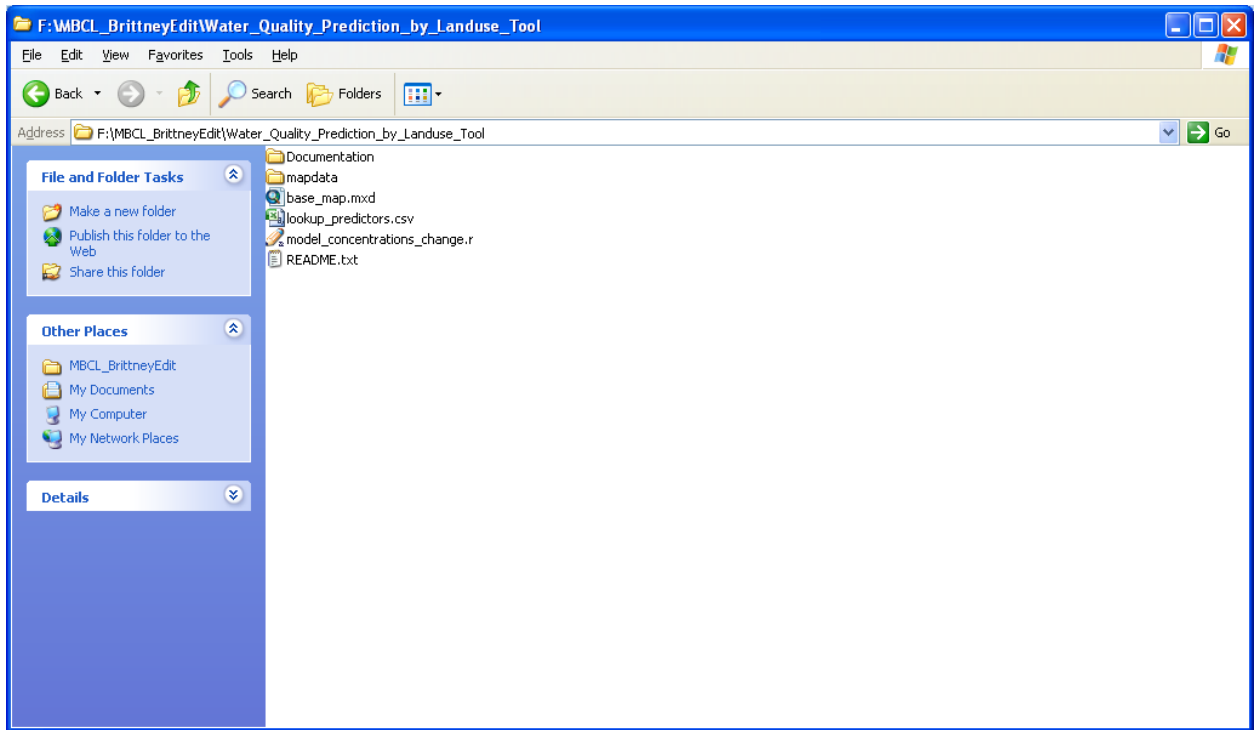


Figure 13. Tool outputs

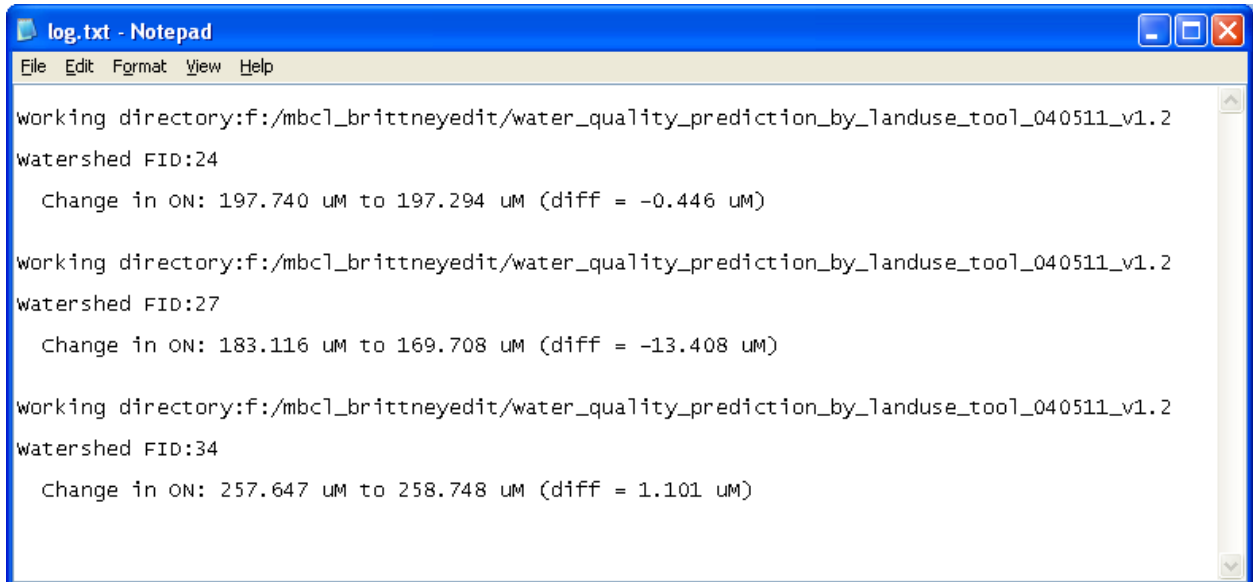
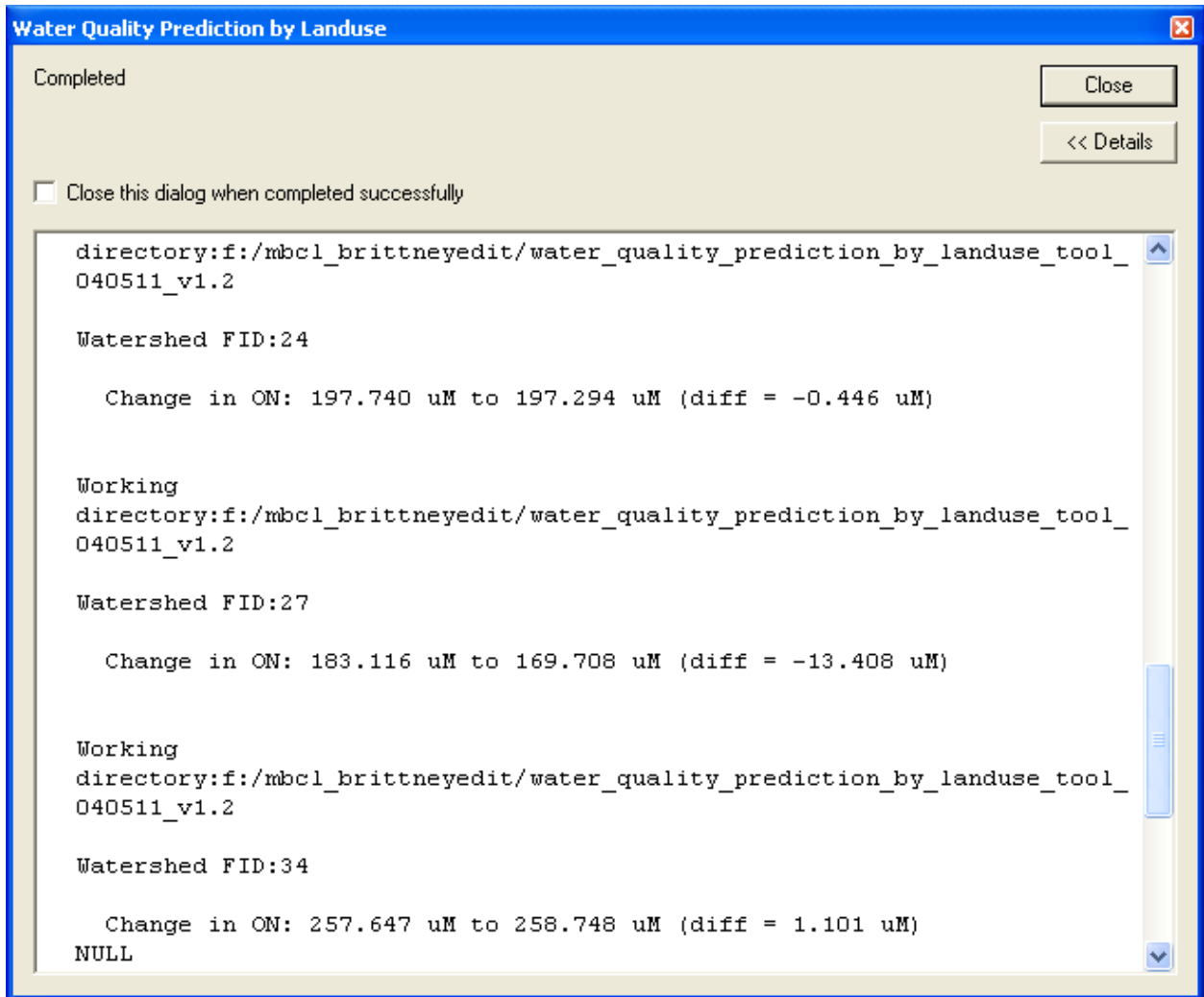


Figure 14. Pair plot illustrating a linear model due to an outlier

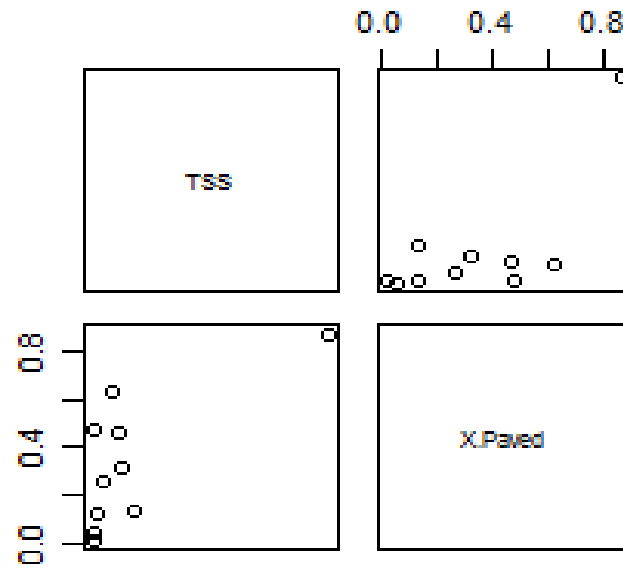


Figure 15. Pair plots for possible NOx predictors

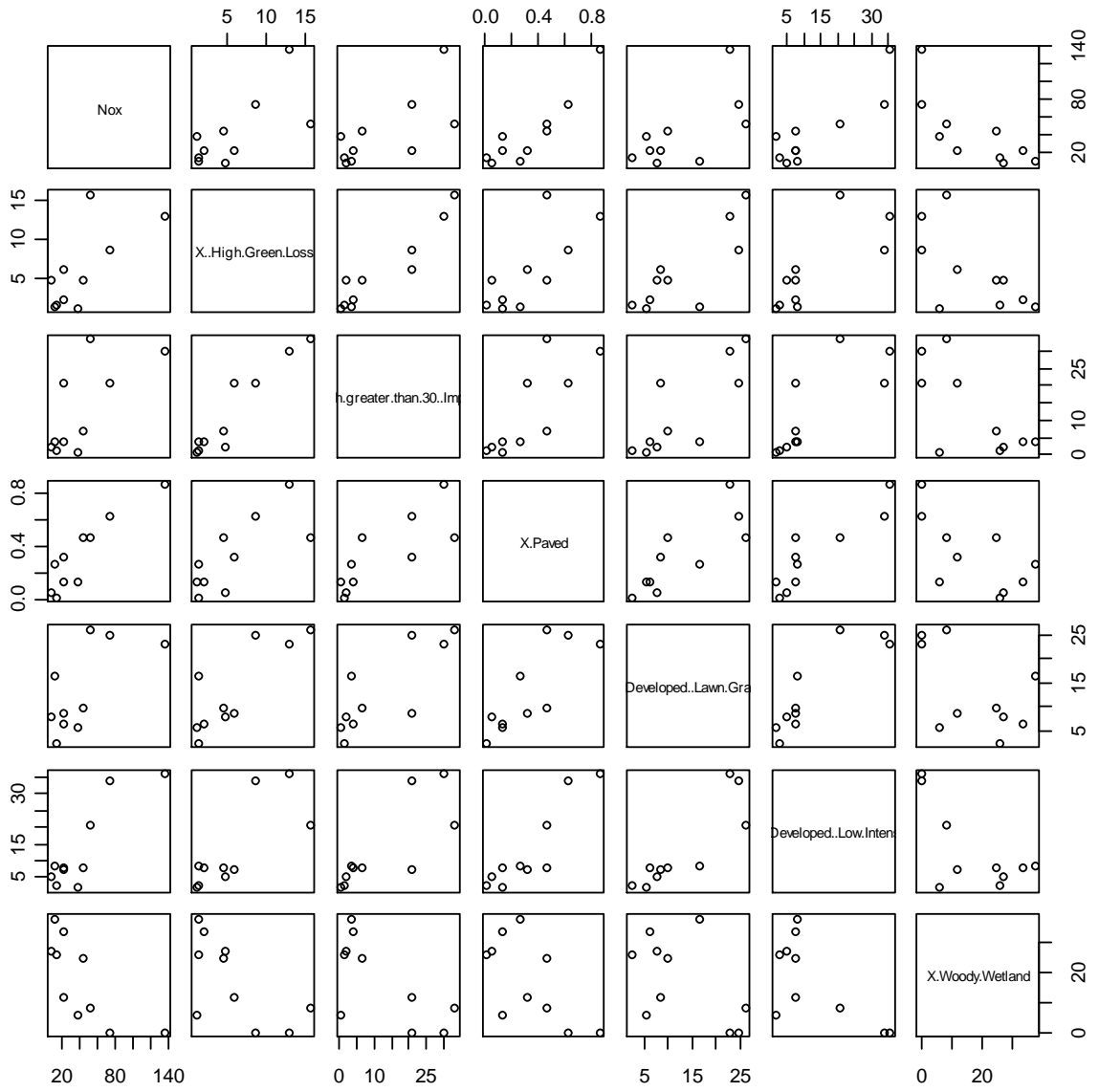


Figure 16. Pair plots for possible ON predictors

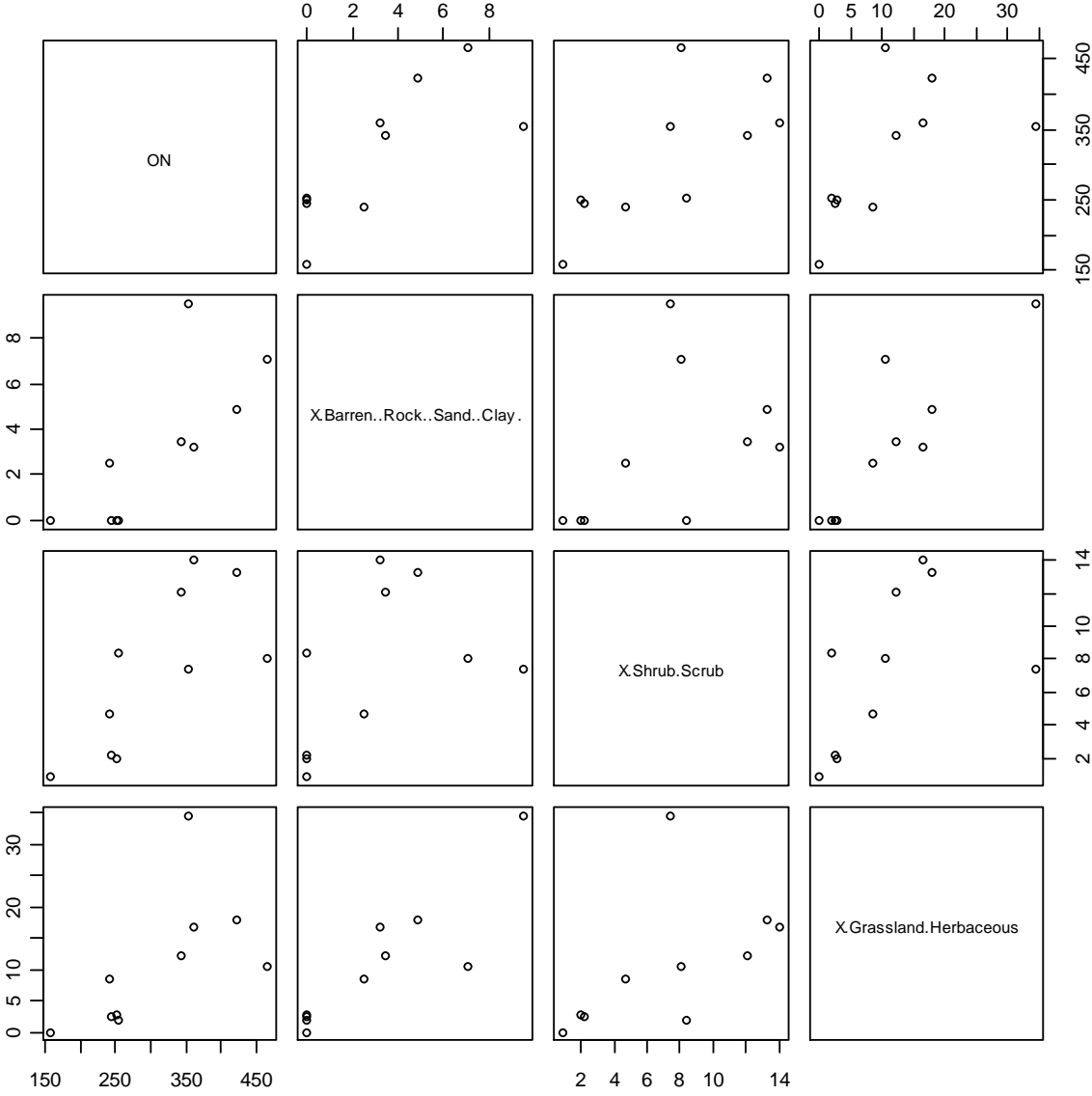


Figure 17. Pair plots for possible NLCD 2006 predictors for NOx

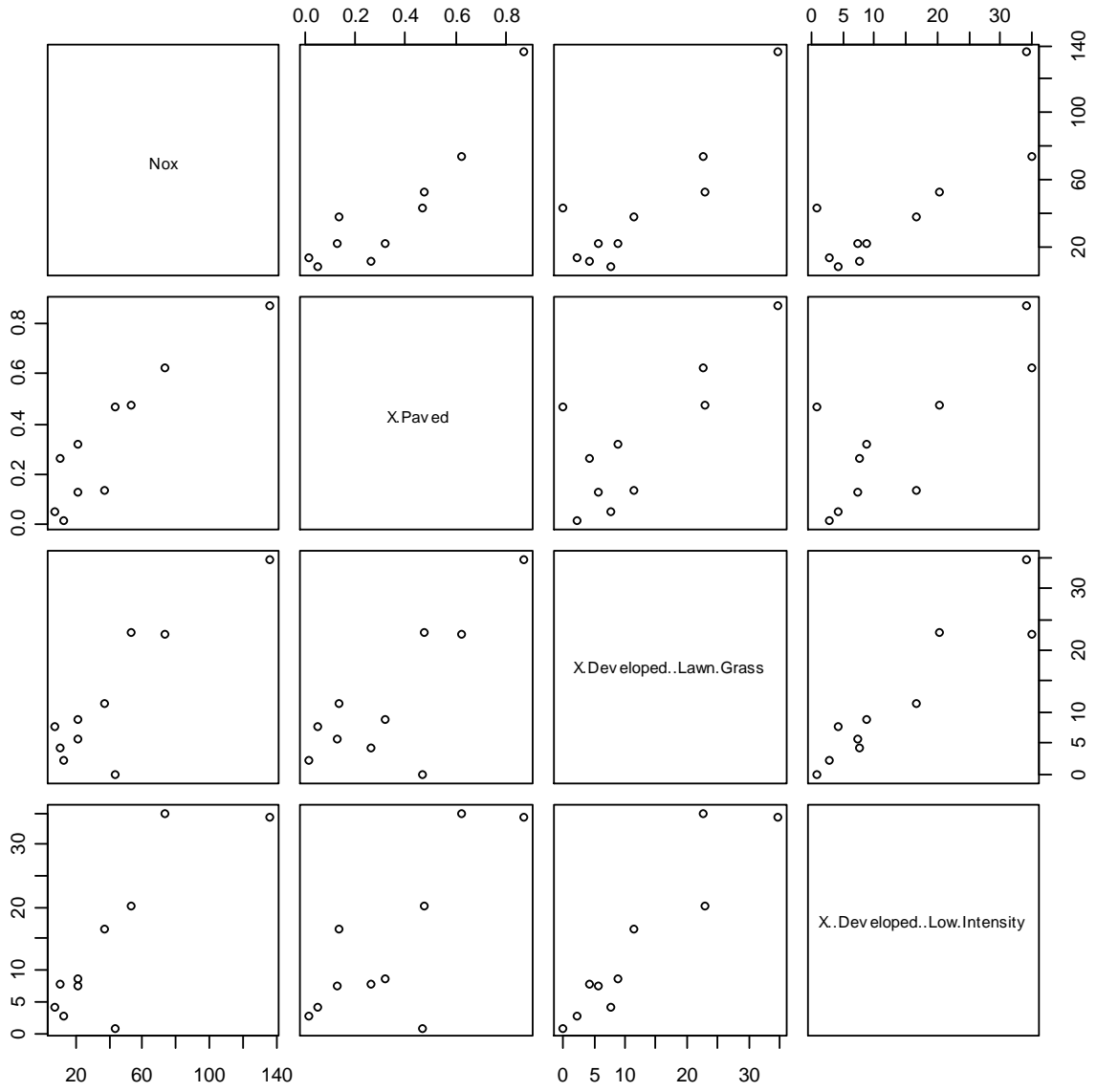
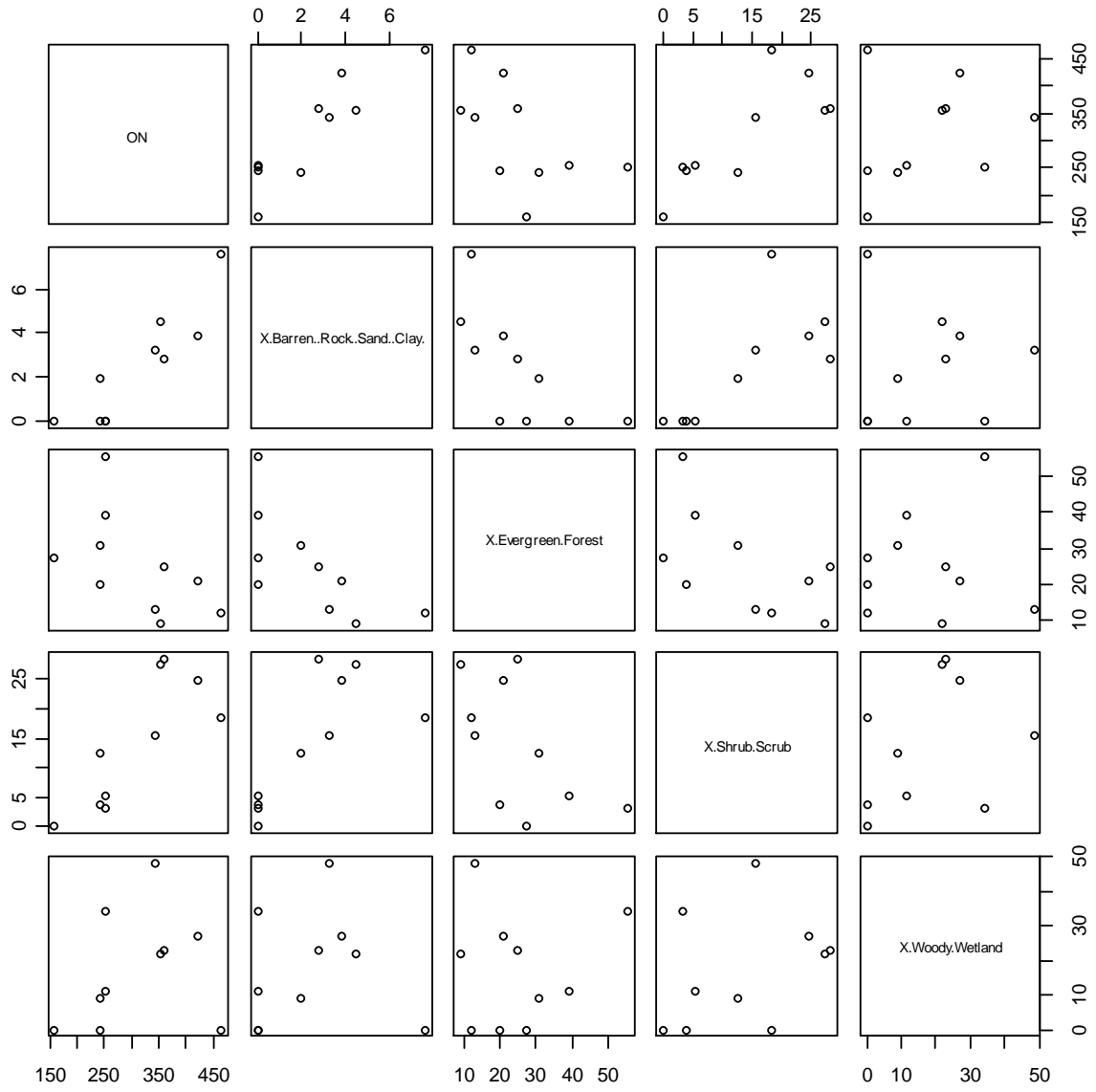


Figure 18. Pair plots for possible NLCD 2006 predictors for ON



Water Quality Prediction by Land Cover Tool Documentation

Contents

PURPOSE	44
SOFTWARE REQUIREMENTS	44
TOOL DEVELOPMENT & ASSUMPTIONS	44
USER'S GUIDE	46
CONTACT INFORMATION	49

Purpose

The purpose of this tool is to predict the change in organic nitrogen concentration when a selected area changes in land cover. This tool allows the user to interactively select a polygon that may undergo land cover change and provide the new land cover type based on the National Land Cover Dataset 2001 definitions. The result of the tool is the change in concentration of organic nitrogen.

This tool may be used as a decision support to tool. Land managers may select several proposed development areas and use the tool to determine how each impacts water quality.

Software Requirements

To use this tool, you must have the following installed on your computer:

1. Esri's ArcMap 9.3.x
2. R 2.9.1 or newer
3. Marine Geospatial Ecology Tools

Tool Development & Assumptions

The tool was created based on water quality data collected in the field. Water quality samples were collected between 12/11/2007-12/13/2010. Statistical relationships were modeled for water quality parameters as the response variables and land cover and soil types as the predictor variables. Organic nitrogen was found to be predicted by barren (rock/sand/clay), shrub/scrub, and grassland/herbaceous land covers. Organic nitrogen concentration in a stream tends to increase with increasing barren (rock/sand/clay) and shrub/scrub land covers. Organic nitrogen concentration in a stream tends to

decrease with increasing grassland/herbaceous land cover. The summary of this linear model is shown below:

```

Call:
lm(formula = ON ~ Barren(Rock/Sand/Clay) + Shrub/Scrub + Grassland/Herbaceous)

Residuals:
    Min       1Q   Median       3Q      Max
-46.977 -25.903  -3.896  28.432  51.378

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      194.386     25.174   7.722 0.000248 ***
Barren(Rock/Sand/Clay)
  32.234      9.199   3.504 0.012760 *
Shrub/Scrub
  11.614      3.435   3.381 0.014841 *
Grassland/Herbaceous
  -6.363      3.050  -2.086 0.082036 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.68 on 6 degrees of freedom
Multiple R-squared:  0.8722,    Adjusted R-squared:  0.8083
F-statistic: 13.65 on 3 and 6 DF,  p-value: 0.00434383
F-statistic: 13.65 on 3 and 6 DF,  p-value: 0.004343

```

Table 1. Linear Model summaries for organic nitrogen

This model is used to drive the results of the interactive tool. The below figure shows the graphical view of the tool model.

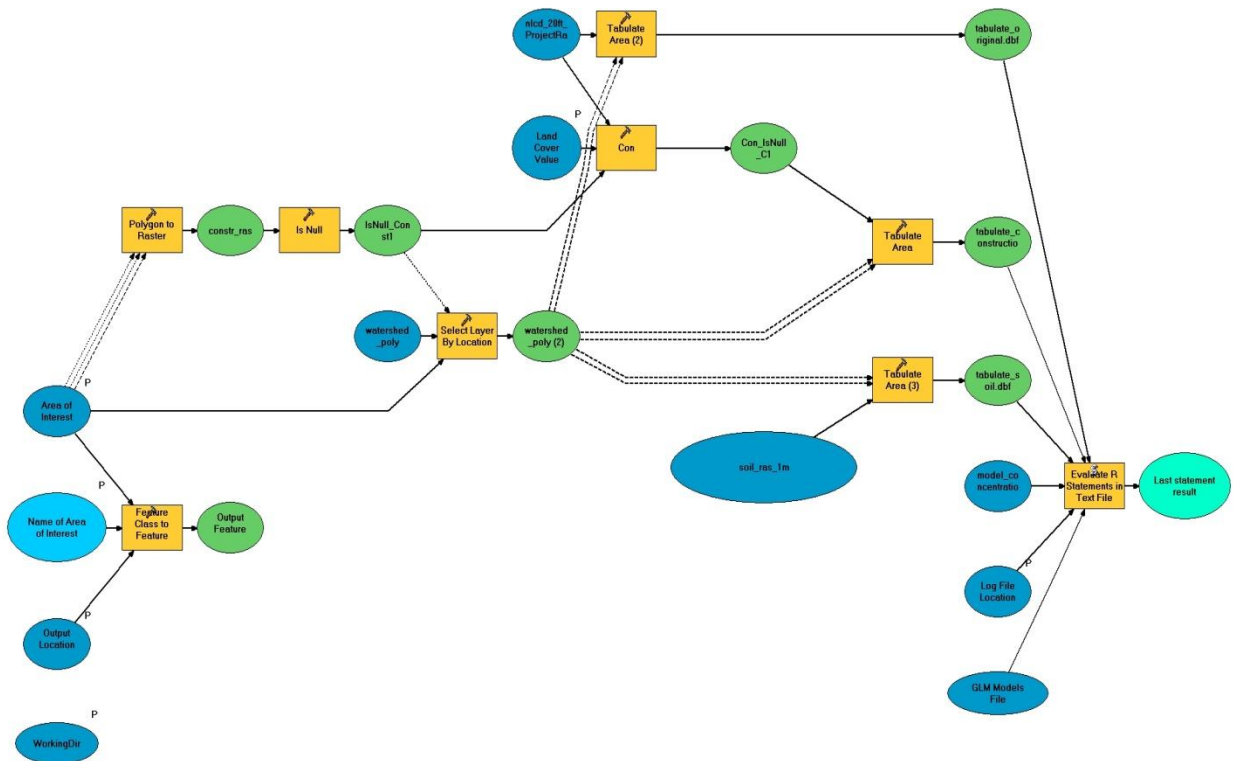


Figure 1. Graphical View of Model

Tool Processes

This section will outline each of the geospatial tools used in the interactive tool.

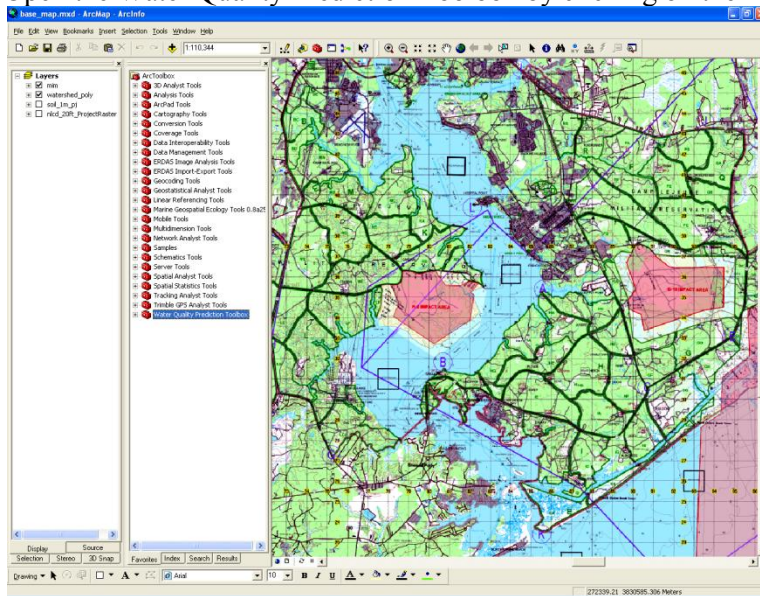
Name	Explanation
Polygon to Raster	Converts the selected area of interest from a polygon shapefile to a raster file for further analysis.
Is Null	Gives the raster of the area of interest a value of 0 and all other areas a value of 1.
Select Layer By Location	Selects the watersheds that intersect with the user selected area of interest.
Con	Inputs the NLCD values for all areas that equal "1" (every area that is not the area of interest). Inputs the Land Cover Value (user variable) for all areas that do not equal "1" (area of interest).
Tabulate Area	Calculates the area of NLCD/Landuse values in the selected watershed. The output is a table.
Tabulate Area (2)	Calculates area of NLCD values (before proposed change). The output is a table.
Tabulate Area (3)	Calculates the area of each soil type. The output is a table.
Evaluate R Statements in Text File	Calculates change in organic nitrogen concentration when the area of interest changes land cover types.
Feature Class to Feature Class	Makes a copy of the Area of Interest shapefile.

User's Guide

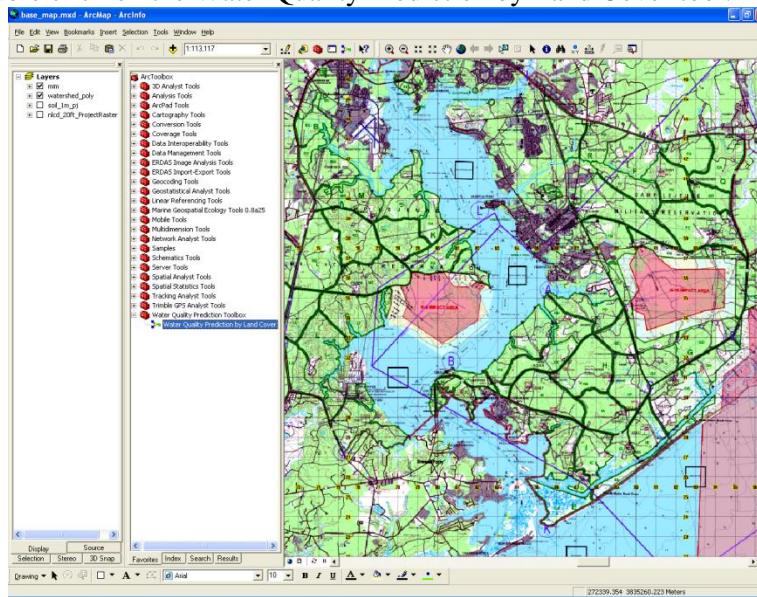
This section outlines the steps for successfully using the tool.

1. Open basemap.mxd located in the tool directory.
2. Turn on the Spatial Analyst extension if it is not turned on already. To do this, click on "Tools", then "Extensions" and check the box next to Spatial Analyst. Some of the underlying processes require the Spatial Analyst extension.

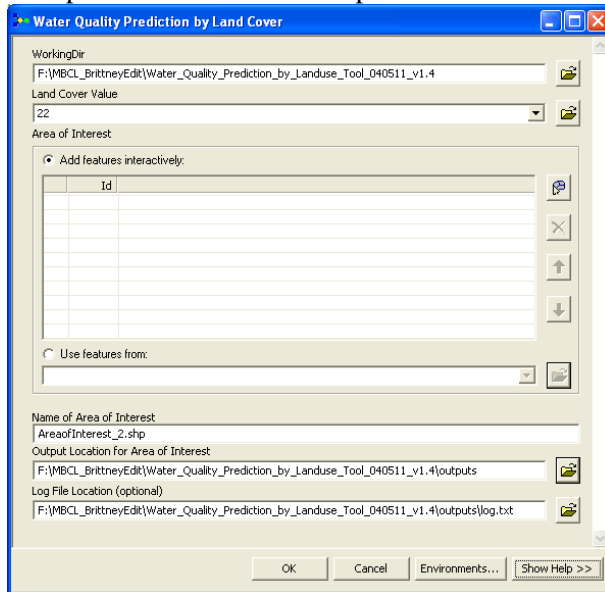
3. Open the Water Quality Prediction Toolbox by clicking on the “+” symbol to the left of the name.



4. Double click on the Water Quality Prediction by Land Cover tool.




5. Fill in the required parameters into the tool. The “Show Help >>” button may be clicked to show an explanation of each model parameter. Each of the parameters is also explained below:



- a. WorkingDir

The path to the working directory where all of the base files are stored. This is also where the outputs of the tool will be written.
 - b. Land Cover Value

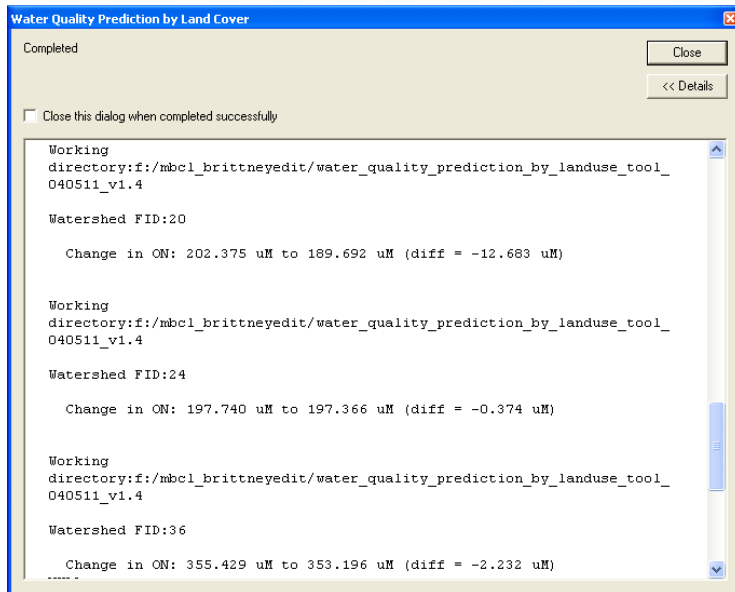
The new land cover value for the Area of Interest based on the NLCD 2001 definitions.
 - c. Area of Interest

The area to undergo the Land Cover Change. This can be selected interactively by drawing a polygon on the map. To select a feature interactively, click on the “Add Feature” button (). Then, draw a polygon on the map by clicking where vertices should be added. Double click to finish drawing the polygon. The Area of Interest may alternatively be loaded from an already existing shapefile. To use an already existing shapefile, use the “Use features from.” option and navigate to the location of the shapefile.
 - d. Name of Area of Interest

The name of the area to undergo the Land Cover Change. The user can run multiple scenarios with different names without overwriting the previous result.
 - e. Output Location for Area of Interest

The file location for the Area of Interest shapefile output.
 - f. Log File Location

The pathway to an existing .txt file to output the concentration results. There are no requirements for the name of the file.
6. Click “OK” and wait for the result.
 7. Interpret the results. The results dialog appears when the tool finishes running. The result gives the Watershed FID and the change in ON. The Watershed FID refers to the watersheds in the watershed_poly shapefile. This shapefile is part of the base map. This indicates how the concentration of ON will change in each watershed. This output is also written to the Log text file.



When interpreting results keep the limitations of the tool in mind. First, the result is given in concentration and not load. This tool cannot be used to calculate how land cover changes will affect Total Maximum Daily Loads (TMDLs). Also, the result only indicates the water quality changes and other changes must be considered for a fully informed decision about future land development, this tool should only influence one part of the decision making process.

Contact Information

For further assistance using this tool, please contact Brittney Baker at brittneyabaker@gmail.com.