FINAL REPORT

Period Covered by the Report: 6/1/2016 – 5/31/2021
Date of Report: Initially submitted 8/31/2021; Revised 11/22/21
EPA Agreement Number: 83616501
Title: Estimating the Benefits of Stream Water Quality Improvements in Urbanizing Watersheds: An Ecological Production Function Approach
Investigators: Roger von Haefen (lead PI); co-PIs: Daniel Obenour, George Van Houtven, Melissa Kenney; Investigators: Michael Gerst
Institution: N.C. State University, Research Triangle Institute (RTI) Univ. of Minnesota; Univ. of Maryland
Research Category: Water
Project Period: 6/1/2016 – 5/31/2021

Objective(s) of the Research: To support EPA's efforts to advance knowledge for conducting economic evaluations of environmental policies, the main objective of our proposed research is to develop and demonstrate methods for valuing the use and nonuse benefits of improving water quality in wadeable streams in urbanizing watersheds. These streams provide valuable ecosystem services and are subject to a combination of anthropogenic stressors that have led to pervasive degradation referred to as "urban stream syndrome." Understanding how the general public perceives and values improvements in stream conditions is necessary to support EPA's efforts to quantify the public's willingness to pay (WTP) for water quality improvements.

Summary of Findings: Our research efforts consisted of five main components: 1) a stated preference survey designed to elicit household's willingness to pay to improve water quality in wadeable streams in urbanizing watersheds; 2) a series of focus groups and cognitive interviews designed to identify the most important ecological endpoints of wadeable streams and test alternative design features of our stated preference survey instrument; 3) a hierarchical water quality modeling framework that leverages existing sparse data for the Upper Neuse River Basin and can be used to forecast water quality indicators under alternative management scenarios; 4) An expert elicitation that links water quality indicators to the ecological endpoints identified in the focus groups; and 5) a case study of the Upper Neuse River Basin that illustrates how the afore mentioned components can be combined to generate policy-relevant benefit measures for alternative water quality improvement scenarios. Below we highlight our key findings from each of these research efforts.

Stated Preference Survey: We developed and fielded a stated preference survey that elicited household's willingness to pay for water quality improvements in wadeable streams in urbanizing watersheds. The survey instrument was developed over a five-year period by the research team with extensive input from focus groups, cognitive interviews, expert advice from Ms. Christy Perrin, Dr. Laura Taylor and Dr. Vic Adamowicz, and two pilot surveys. The survey employed a "push to internet" format whereby 12,500 randomly selected residents of the North Carolina counties of Wake, Mecklenburg and Guilford were invited by mail to participate in the online survey. 2,511 individuals ultimately completed at least one stated preference question (response rate = 21.4%), and several comprehension checks and debriefing questions suggested the data was of high quality. Results from econometric modeling suggested that households value all three ecological endpoints and are willing to pay roughly $16 to improve one stream mile from the lowest categories to the highest categories. To construct county-level values, separate willingness to pay estimates must be calculated for different households in the relevant county and then summed.

Focus Groups and Cognitive Interviews: We conducted thirteen focus groups and eight cognitive interviews over five years with two primary goals in mind: 1) to identify the ecological endpoints associated with streams that households care the most about; and 2) to pretest and refine aspects of our stated preference survey instrument. Most focus groups and interviews were conducted on the N.C. State campus in Raleigh, NC and lasted for roughly two hours. We identified four main ecological endpoints – ecosystem condition, human health risk, murky water days, and trash – and developed graphical and text-based descriptions of each that were helpful to respondents. We also developed three levels for each endpoint that could be used to characterize stream conditions in ways that were understandable to the lay public and consistent with the available science. As we developed and refined our survey instrument, the focus groups helped us to learn that including four ecological endpoints was overwhelming to some respondents. After an initial pilot survey confirmed this concern, we decided to drop trash as one of the ecosystem conditions included in our description of streams. The focus groups and interviews also were instrumental in identifying a credible payment vehicle (a stormwater fee associated with a household's monthly water bill), the structure and number of our choice experiments, bid values, and how to describe the action plans in a way that provided sufficient detail without overwhelming respondents.

Water Quality Modeling: We used publicly available data for six water quality indicators to develop a predictive model that could be used to forecast water quality conditions in the Upper Neuse River Basin (UNRB) where data is often sparse in both temporal and spatial dimensions. The model uses an elegant hierarchical Bayesian approach that efficiently uses all available data and the linkages in water quality across time and space. Sparse data environments like the one found in our application are common in many policy settings, so the modeling framework employed here should be transferable to other applications.

Expert Elicitation: To link the output from the water quality model to the ecological endpoints that individuals care about, we used an expert elicitation. Eight experts participated, and their mappings from six water quality indicators to three ecological endpoints over 100 hypothetical scenarios allowed us to estimate ecological production functions that link measurable indicators to the three levels of ecological endpoints – ecosystem condition, human health risk, and murky water days – that the public cares about.

Upper Neuse River Basin Case Study: To illustrate how the various aspects of our research efforts can work in unison to inform policy, we developed a case study for the Upper Neuse River Basin. The case study leveraged predictions on how six water quality indicators are likely to change in response to interventions that reduce the negative impacts of streambank erosion, impervious cover, and random pollution sources such as leaking pipes. These changes in indicators are mapped into changes in ecological endpoints which are then fed into the water quality valuation model estimated with the stated preference survey data to generate monetary values for two hypothetical clean-up scenarios in two separate sub-watersheds in Wake County, North Carolina. The results suggest, for example, that the average Wake County household is willing to pay roughly $110 per year for the water quality improvements resulting from a 25 percent increase in canopy cover combined with a 25 percent decrease in the negative effects of impervious cover in the Crabtree and Walnut Creek sub-watersheds located in the densely-populated central part of Wake County. We found some evidence of distance decay in these values (i.e., residents living in close proximity to the cleaned-up watersheds relative to those living further away).

A more detailed discussion of each of these components is found on subsequent pages of this report.

Technical Effectiveness / Economic Feasibility / Environmental Benefits: The current study extends previous EPA-funded research by Phaneuf et al. (2013) that estimates the benefits of water quality improvements in lakes and reservoirs in the Southeast by investigating the benefits of water quality improvements in wadeable streams in the same region. In tandem, the Phaneuf et al. study and the current research quantify the total economic benefits of water quality improvements for a significant percentage of urban surface waters. Future research could extend this research by quantifying the economic benefits of water quality improvements in intermittent streams and wetlands as well as rivers. In terms of methodology, the Hierarchical Bayesian methods used in the current study to model water quality under baseline and alternative policy scenarios are attractive for data environments where existing monitoring data is sparse, and expert elicitation is an effective tool for translating water quality indicators into ecological endpoints.

Subaward Monitoring: Consistent with 2 CFR 200.331(d), PI von Haefen worked closely with co-PIs Van Houtven, Kenney and Obenhour to ensure grant-related resources were allocated appropriately and effectively to achieve the grant's objectives and to keep the project on-schedule.

Publications / Presentations: One dissertation (Miller, 2019) has been published, one manuscript has been published in *Freshwater Science* (Miller, Paul and Obenour, 2019), and two additional manuscripts are in preparation. One of the manuscripts in preparation is for the *Proceedings of the National Academy of Sciences* special issue titled, "The Clean Water Act After 50 Years: Innovations in measuring the social benefits of water quality for research and policy." This research was presented at several EPA workshops, the 2021 virtual W4133 Annual Meetings, the 2021 virtual Social Cost of Water Workshop, and the 2021 Agricultural and Applied Economics Association Meetings in Austin, TX.

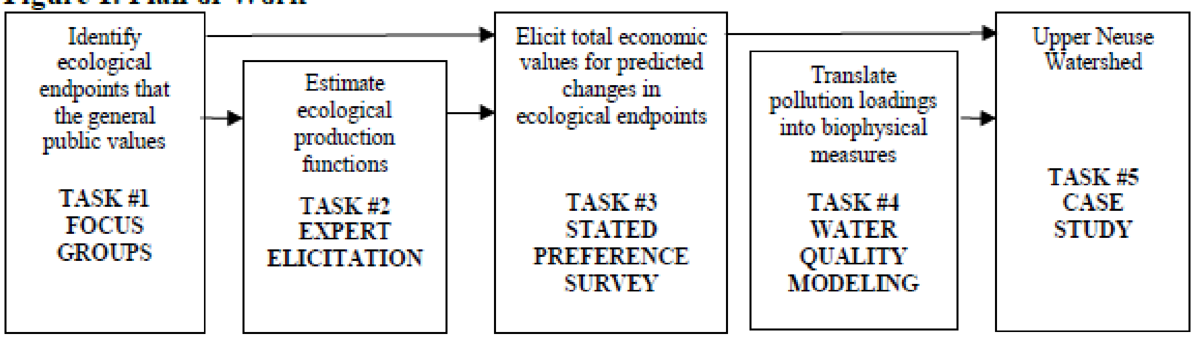
Supplemental Keywords: Water quality, economic benefits, willingness to pay, stated preference methods, expert elicitation, ecological production function, wadeable streams, Upper Neuse River Basin.

Relevant Websites: https://rvhaefen.wordpress.ncsu.edu/epa-water-quality-grant/ (report) & https://dataverse.harvard.edu/dataverse/stream_water_quality_benefits (support files)

Chapter 1. Introduction

This report summarizes research conducted over the past six years to generate benefit estimates for improvements in water quality at wadeable streams in urbanizing watersheds. This research was funded through a grant from the Environmental Protection Agency and conducted by an interdisciplinary team of researchers – economists, decision scientists, and engineers. Figure 1 lays out the research's components and general approach.

Figure 1. Plan of Work



The research began in 2016 with focus groups of water quality experts and lay people in North Carolina's urban counties. The goal of these focus groups was to learn what characteristics of wadeable streams experts and lay people thought were most important and most under stress by human activities. As we describe in detail in Chapter 2, these focus groups identified three key ecological endpoints of wadeable streams that lay people care about – ecosystem condition, human health risk, and murky water days. Chapter 2 also describes the stated preference survey instrument that we developed and tested during the focus groups to measure people's willingness to pay for water quality improvements defined in terms of the three endpoints. The chapter summarizes the data collected, econometric models run, and generic welfare measures for improvements in stream water quality.

A separate thrust of this research – summarized in Chapter 3 - used expert elicitation to map measurable water quality indicators into these ecological endpoints. This research linked output from Hierarchical Bayes water quality modeling – described in Chapter 4 – with the ecological endpoints identified in Chapter 2 by relying on the judgment of eight water quality experts. The elicitation protocols, data collected, and estimated statistical models are all reported in Chapter 3, and the data and methods for the water quality modeling are described in Chapter 4.

Chapter 5 pulls together the three distinct but related lines of research with an Upper Neuse River Basin (UNRB) case study. The case study illustrates how the data and methods summarized in this report can be used for benefit estimation.

Chapter 2. Survey Instrument Design

2.1. Survey Development

The survey was designed to use a "push-to-internet" data collection approach, where all respondents were initially contacted and recruited by mail and directed to a website containing an online data collection instrument.

The web-based survey instrument was systematically designed and pretested through a series of focus groups, one-on-one cognitive interviews, and two pilot surveys. Before drafting and testing specific survey questions and stated preference scenarios, we conducted a workshop involving 11 local environmental and water quality experts from non-profit organizations, local governments, and academic institutions in the Research Triangle Park region of North Carolina. The purpose of this workshop was to elicit participants' views about the most pressing urban stream water quality issues in the region and advice about data and other resources that we could use for the study.

The first set of 9 focus groups with private individuals from the local community were conducted on the NC State campus over a period of 22 months (in 2017 to 2019). Using advertisements posted on Craig's List and Reddit, 10-12 participants were recruited for each meeting, all of whom were residents of Wake County, North Carolina. To ensure a broadly representative sample of participants for each session, the advertisements provided only a general description of the topic to be discussed in the meetings, participants were selected from the list of responders to provide a mix of age, gender, and race characteristics, and all participants were provided with monetary compensation in the form of gift cards for their time.

Wake County was selected as the initial geographic region of interest for the focus groups and survey development because it is among the most densely populated and rapidly urbanizing areas of the Piedmont region of the Southeastern US. As a result, conditions related to urban stream syndrome are commonly observed in the area. County political boundaries, rather than watershed boundaries, were selected for the study region with the expectation that (1) these boundaries would be most familiar to respondents and (2) they correspond best with the areas of influence and responsibility for public sector programs to improve local stream water quality.

The initial focus groups had relatively open-ended formats, where the main objectives were to have participants describe in their own words (1) their understanding, perceptions, and attitudes about the types and importance of different water quality conditions in Wake County streams and (2) the ways in which they observed or interacted with the streams. The facilitated sessions also explored participants' views and attitudes about potential county-level programs to improve water quality conditions. In addition, participants were asked to examine and provide feedback on different map formats showing the location and condition of streams in the region. This process helped to confirm expectations that county boundaries were most familiar and appropriate for defining the geographic area of interest for the study.

The feedback and findings from the initial focus groups were used to develop materials for and preliminary drafts of the survey instrument, which were then pretested in subsequent focus groups. Based on extensive background research and inputs from the expert workshop, focus groups, and our team's water quality experts, we selected four main potential water quality attributes to be included in the survey choice experiments: (1) aquatic ecosystem condition; (2) health risk from pathogens; (3) visual water clarity; and (4) trash. We developed text, graphics, and pictures to describe the main causes of impairments to these water quality attributes (particularly sediment erosion and stormwater runoff) and three discrete quality levels (e.g., poor, fair, good) for each attribute.

The decision to include three levels for each attribute was based on two main factors. First, we knew that the expert elicitation being developed in parallel with the survey would allow us to translate continuously measured water quality parameters, such as turbidity, dissolved oxygen, and fecal coliform bacteria, into multiple discrete levels of the water quality attributes.[1] Second, we needed to limit the cognitive burden of explaining multiple attributes and levels to survey respondents. The selection of three levels for each attribute was, in a sense, a compromise between these two factors.

Through feedback from multiple focus groups, we developed and refined our presentation and description of the causes and levels of the selected water quality attributes. For example, we found that for the most part, simple infographics were more effective than photographs for conveying water quality information. We also tested different formats for presenting risks of gastrointestinal illness for children and iterated towards a colored grid format similar to what is widely used in the risk communication literature that was found to be most understandable for focus group participants.

In addition to conveying the *levels* of each water quality attribute, the survey instrument also needed to present and explain variation in the *spatial extent* of water quality levels and changes. One way to do this was through maps showing the location and network of affected streams in the county. We refined the format and presentation of maps using feedback from focus group participants, and we also developed stated preference scenarios that focused on a specific subset of roughly 100 stream miles within the county. For example, in Wake County we specifically highlighted streams that make up the Crabtree Creek and Walnut Creek sub-watershed located in the center of the county near the urban core. In addition to allowing us to focus the survey scenarios on the most impaired streams in the county, selecting this subset of streams provided a simple way to account for variations in distance between survey respondents' homes and the primarily affected streams (i.e., potential "distance decay" effects).

The second way to convey variations in the spatial extent of water quality levels is by describing the *percent* of streams that can be classified in each level of water quality. Based on focus group pretesting, we concluded that a pie chart format for each of the four water quality attributes was the most effective way to communicate these percentages. Based on evidence from the scientific literature and from focus group feedback, we also concluded that a blue-red-gray color scheme

[1] For example, in a similar previous study of lake water quality, expert elicitation was used to translate six parameters (including nitrogen and phosphorus levels) into five discrete levels of a single attribute (lake eutrophication).

would be most effective, with red used to represent the lowest/worst water quality level for each attribute, blue used to represent the highest/best level, and grey used for the medium level.

Using feedback from focus group participants, we also converged on a choice experiment format and design with several key features. First, we concluded that a county-based water quality improvement program was the most realistic and easy-to-communicate policy mechanism for achieving changes in the water quality attribute levels and percentages. Second, the corresponding payment vehicle that was most believable and acceptable to respondents was a new monthly stormwater fee added to residents' existing water bills. Third, a hypothetical referendum (dichotomous choice) voting format was also most appropriate, whereby respondents are asked to choose between (1) a continuation of current/baseline water quality conditions (i.e., vote AGAINST the new proposed program) and (2) implementation of a new water quality improvement program with defined changes (i.e., vote FOR the new proposed program). Fourth, presenting each respondent with a sequence of four hypothetical program voting referenda was a reasonable compromise between gathering as much preference information as possible and not overburdening respondents. The survey was also designed in a way that presented the first choice experiment as if it were the one and only scenario to be considered (and thus preserve the incentive compatibility of initial responses) followed by three follow up choice experiments where respondents were told that additional choice experiments designed to collect information on additional policies considered. Before the follow up scenarios were presented, respondents were told to treat each vote as if it were the only one on the ballot.

To make the hypothetical water quality improvement programs as believable as possible, we also designed and pretested descriptions of specific county-level program activities that would be used to improve water quality. These descriptions included infographics and text about programs including construction of stormwater capture basins, sewer connection repairs, community information programs, and trash clean-up programs.

As we made progress on developing and refining the survey instrument focused on Wake County, we also developed versions of the instrument that were tailored to residents of two additional counties in North Carolina – Guilford County and Mecklenburg County. When added to Wake County, the areas selected for the study are the most densely populated urban counties in North Carolina, and they are all located in the central Piedmont ecoregion of the state. As many features of the survey instrument as possible were kept identical across the three counties; however, the spatial descriptions, in particular, of existing streams (and which streams would be affected by a water quality improvement program) needed to be tailored to conditions in each county. Most importantly, we developed new county-level maps and descriptions for Guilford and Mecklenburg Counties, which identified specific sub-watersheds of interest for these areas.

The last two focus groups supporting the design of the survey instrument were conducted in Guilford County and Mecklenburg County. In addition to requesting specific feedback on the new county maps and descriptions, these focus groups provided participants with near-final versions (on paper) of the main sections of the survey, including the choice experiment voting questions. Participants were asked to read and answer sections of the survey as if they were taking it at home. They were then asked to discuss areas that needed more clarification or explanation and to describe their reasons for voting for or against the proposed programs. They were also asked to review and comment on a draft of the survey invitation letter that would be mailed to a random sample of county addresses.

Applying the insights gained from the focus groups, the next step was to program and pretest a web-based version of the survey instrument. This online version of the survey was developed using Qualtrics software. To encourage as high response and completion rates as possible, the online instrument was designed to be accessible and easily readable by smartphone, tablet, laptop, or desktop computer screens. Most importantly, the format of the choice experiment questions, including all attributes and options, was designed so that it could be easily viewed and answered even on a small screen.

The web-based version of the survey was then iteratively pretested and refined through a series of one-on-one cognitive "think aloud" interviews. The recruitment process and selection for these interviews was the same as for the focus groups. The interviews were conducted in-person at NCSU with members of the project team. Eight respondents were given instructions to access the survey using different online modes. They were then asked to read all the survey the text and questions out loud and to verbally report and discuss their answers with the interviewers. This version of the survey also included a series of "debriefing" questions to be included in the final survey, which were designed to further explore the respondents' motivations, interpretations, and understanding of the choice experiment questions and their responses.

The final stage of survey pretesting involved implementation and analysis of a pilot version of the survey, which was launched in February 2020. Respondents for the pilot survey were recruited from the Qualtrics Online Panel, with the requirement that they be adult residents of one of the three selected counties. We also specified sampling quotas from the panel to ensure a similar number of respondents from each county, as well as a broad mix of age and race among respondents. For quality control purposes, we also screened out respondents who incorrectly responded to simple quality questions in the survey or who those who completed the survey below a minimum time threshold. The experimental design for the four discrete choice preference elicitation questions, which included four levels for the monthly cost attribute and three levels for the percentage of streams in the best and worst categories of the four water quality attributes, was created using Ngene software.

The final pilot sample was comprised of 730 respondents. Analysis of these survey data, in particular the responses to the discrete choice questions and debriefing questions, indicated that, for the most part, respondents found the choice tasks to be believable and understandable and were able to provide meaningful responses. Analyzing the discrete choice responses using conditional logit models, we found that, although the preference parameters generally had the expected sign, the level of statistical significance was often low for the estimated parameters. One conclusion that we drew from these results (along with observations from the other pretesting activities) was that the choice scenarios and tasks were still cognitively challenging for many respondents and therefore needed to be further simplified and clarified.

Based on these findings we made three main changes to the survey instrument and design. First, we excluded the water quality attribute focused on trash levels in streams. Although focus group discussions had indicated that trash was an important concern for the public, responses to the choice and debriefing questions indicated that it was overall the least important of the four attributes. Excluding this attribute was intended to reduce the cognitive burden on respondents. Second, to help respondents better understand the choice tasks tradeoff, before presenting the first choice task in a single table, we added text and graphics that separately explained the changes that would occur to each attribute (for examples, see pages 40-43 in the Appendix 2A). Third, we revised the bid design to include a wider range of monthly cost (bid) values. A second pilot survey was fielded to confirm that these changes worked as intended. Econometrics results collected from a second pilot survey of 420 Qualtrics respondents were more promising; the water quality parameters had the expected sign and were generally statistically significant. This finding gave the research team confidence that it was ready to field the final survey.

The format of the finalized survey instrument is shown in Appendix 2A. The version shown in the appendix is the one used for Wake County residents. The versions used for Guilford and Mecklenburg County residents are fundamentally the same but modified to include maps and to reference streams and waterbodies specific to these areas. Maps used of the three counties are shown below in Figures 2.1, 2.2, and 2.3.

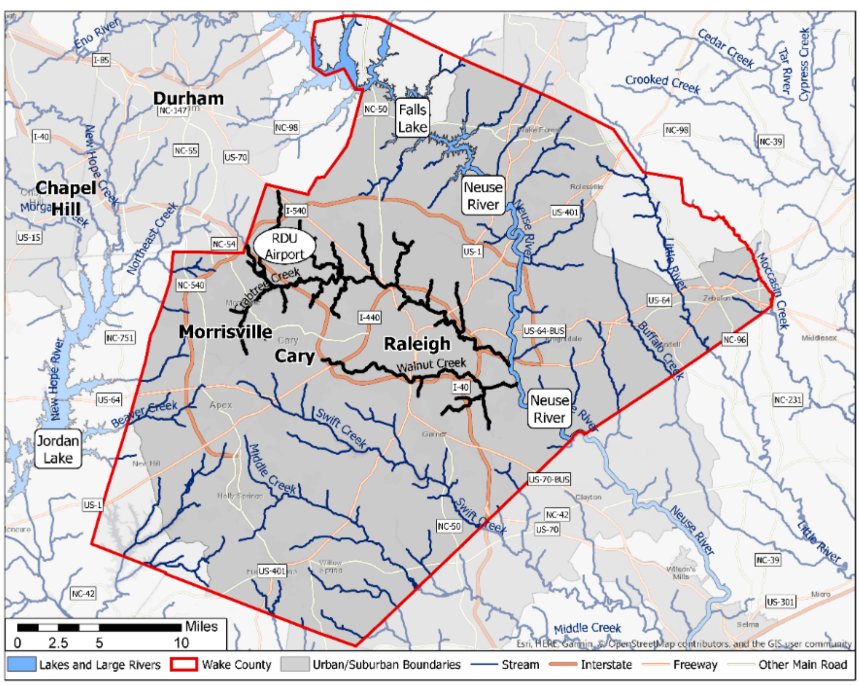


Figure 2.1. Map shown to Wake County residents

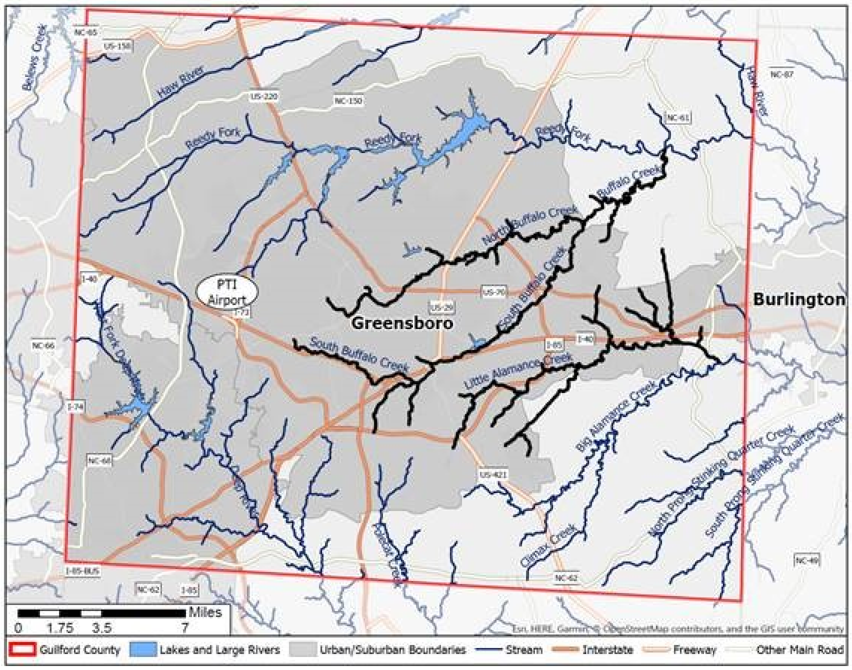


Figure 2.2. Map shown to Guilford County residents

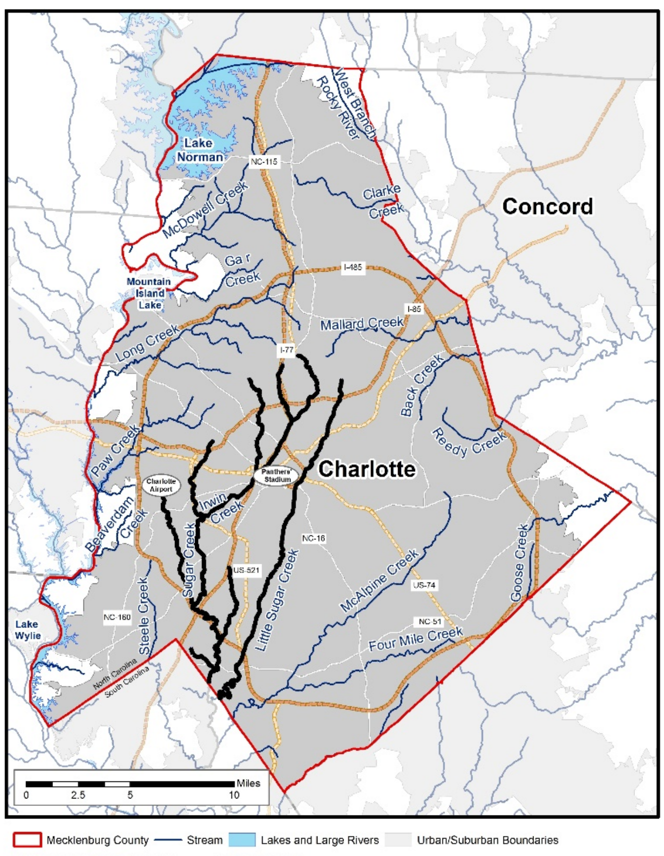


Figure 2.3. Map shown to Mecklenburg County residents

2.2. Final Survey Administration

The final survey was launched in February 2021 and data collection was completed in April 2021. A total sample of 12,500 addresses were randomly drawn (split evenly between the three counties) from an address-based sampling frame developed and maintained by RTI International. Survey invitation letters were sent to each selected address (including a $2 incentive in each envelope) with a general description of the survey and instructions for accessing the questionnaire online (including an individualized code for identification purposes). A follow-up reminder postcard was then also sent to each address one week later. Examples of the letter and postcard (for Wake County residents) are shown in Appendix 2B.

A third reminder postcard (with the same format as the second) was sent three weeks after the initial letter mailing to a subset of the initial addresses. In this mailing, addresses were excluded if (1) the individualized code from that address's invitation letter was entered into the survey website by the recipient/resident or (2) the invitation letter and/or initial postcard were returned by the postal service as undeliverable mail (e.g., no longer valid address).

A final reminder postcard, with a different format and instructions for receiving a $20 completion incentive (see Appendix 2B for an example from Wake County), was sent six weeks after the initial letter mailing. In this mailing, addresses were again excluded if (1) the individualized code from that address's invitation letter was entered into the survey website or (2) the invitation letter and/or initial postcards were returned as undeliverable.

2.3. Summary Statistics an Econometric Results

Of the 12,500 households contacted, 750 had one or more mailings returned undelivered, and 2,511 completed at least one choice experiment. This implied an overall completion rate of 21.4%, and the rate was highest in Wake County (25%) and lowest in Mecklenburg County (18%). To correct for potential nonresponse bias, we used county-level Census data to develop sample weights by county that post-stratify based on income and whether the respondent identifies as white, Latino or Spanish, a college graduate, and retired. In general, item nonresponse was infrequent – for example, over 97% of the 2,511 respondents answered all four choice experiments, 89% reported their household incomes, and most other demographics were missing less than 7% of the time. We used regression-based imputation for missing incomes and imputed sample means by county for the other demographics.

Responses to debriefing questions suggested that respondents believed the survey to be balanced (75%) (i.e., not pushing them to vote for or against the proposed programs), provided enough information for them to make informed choices (83%), and felt the survey was both price and policy consequential (88% and 61%, respectively). 49% and 48% of respondents felt that health risk and ecosystem condition were the most important water quality attributes, respectively, and 86% felt that murky water days was the least important attribute. 41% of respondents did have doubts about their county government's ability to achieve the water quality improvements described in the choice experiments. According to the respondents, the COVID pandemic had only minor effects on their responses; 75% stated the pandemic had no effect on their responses, 19% said it made them more likely to vote in favor of the action plans, and 6% said it made them less likely to vote in favor of the action plans.

Given our focus on streams, an obvious concern is that respondents might perceive that improvements in stream water quality will spill over to downstream and neighboring waterbodies in the watershed. Before the choice experiments, we informed respondents that the action plans under consideration would have no measurable effect on lakes and rivers and only streams that were targeted by the policies would experience improvements in water quality. To confirm that respondents accepted this assertion, we asked respondents whether they agreed with the following statement, "When evaluating the action plans, it was my understanding that water quality would only improve in the streams that are part of or connect to <STREAM NAMES HERE>. There would be no effect on other streams or on lake or river water quality." 87% of respondents agreed or strongly agreed with this statement, and less than 4% disagreed or disagreed with it. Moreover, 90% (3%) of college-educated respondents agreed or strongly agreed (disagreed or strongly disagreed) with this statement, and 92% (3%) of environmentalists agreed or strongly agreed (disagreed or strongly disagreed) with this statement.

Finally, to ensure that we only used high quality data, we dropped from our econometric analysis observations where: 1) respondents took less than eight minutes or more than one week to complete the survey; 2) respondents failed a "trap" question that checked whether they were reading and paying attention to the survey (less than 9% of the sample failed this question); and 3) respondents who stated that the survey was price or policy inconsequential. These quality criteria resulted in 615 (or 24% of) observations being dropped from the final analysis, although their inclusion does not change our results substantively.

Tables 2.1 and 2.2 below report econometric results from a referendum-style fixed coefficient, discrete choice logit model where we model respondents' decisions to vote for or against the action plans as a function of the water quality improvements and the action plan monthly costs. We used the sampling weights in estimation (although not doing so would not change our results qualitatively) and cluster the standard errors by respondent.

Both tables exclude the middle category for ecosystem conditions, health risk, and murky water days to avoid multicollinearity issues. Given that in each county 100 stream miles are affected by the action plans, this specification implies that the water quality variables (in blue and red) should be interpreted as the net utility gain of moving one stream mile from the middle category to either the good/low (in blue) or poor/high (in red) category. Under this specification, our expectation is that moving stream miles from the middle to the good category positively impacts utility, whereas moving stream miles from the middle to the poor categories negatively impacts utility. The results in both tables bear these expectations out, although the percent of streams in the poor category for murky water days is not statistically significant.

Table 2.1. Choice Experiment Results

```
Logistic regression                             Number of obs     =      7,584
                                                Wald chi2(11)     =     512.68
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -3686029.3               Pseudo R2         =     0.1206

                             (Std. Err. adjusted for 1,896 clusters in resp_id)
```

ce	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
cost	-.0561872	.0034198	-16.43	0.000	-.0628898	-.0494846
ec_g	.0162029	.003597	4.50	0.000	.009153	.0232528
ec_p	-.0152104	.0038619	-3.94	0.000	-.0227796	-.0076411
hr_l	.0113868	.0029719	3.83	0.000	.005562	.0172116
hr_h	-.0248146	.0087398	-2.84	0.005	-.0419444	-.0076848
md_l	.0078538	.0022209	3.54	0.000	.003501	.0122066
md_h	-.0021807	.0047154	-0.46	0.644	-.0114227	.0070614
age	-.0162309	.002857	-5.68	0.000	-.0218305	-.0106312
gender	-.1008659	.1026613	-0.98	0.326	-.3020784	.1003466
white	-.2496793	.1215481	-2.05	0.040	-.4879092	-.0114494
college	.2653966	.1025011	2.59	0.010	.0644982	.466295
_cons	1.288001	.2250702	5.72	0.000	.8468716	1.729131

The main difference between the models in Tables 2.1 and 2.2 relates to the specification of the cost variable and their implications for distance decay in willingness to pay, i.e., those who live further from the cleaned-up streams have lower values for improvements. In Table 2.1, the cost variable is assumed have a constant effect across all individuals, implying no distance decay. In Table 2.2, the cost variable is allowed to vary by distance to the centroid of the improved quality streams. In particular, cost enters the model as a separate, independent variable and also as an interaction with $1/(1+dis_i)$, where dis_i is the distance in miles to the centroid of the cleaned-up streams. If this interaction is positive, it implies that individuals who live further from the cleaned-up streams value cost increases more and thus are willing to pay less for stream improvements. Our results suggest that the coefficient on this interaction term is positive and almost significant at the 10 percent level. In other words, our results provide suggestive but not compelling evidence of distance decay in willingness to pay for stream water quality improvements, which may in part be due to our analysis being limited to the specific counties where streams are cleaned up. In ongoing research, we are further investigating this important issue.

The results in both tables allow for observable demographic heterogeneity in responses. On average, these results suggest that older whites without a college degree (college = 0) are more likely to vote against the action plans.

To gain a sense of the welfare implications of these results, we consider a scenario whereby one stream mile is cleaned up from the lowest to the highest category for all three attributes. Since all hypothetical action plans affected 100 stream miles, this scenario corresponds to moving one percent of streams from the lowest to the highest category for all three attributes. In this case, a representative household's annual willingness to pay (WTP_i) is therefore:

$WTP_i = 12\times(\beta_{ec\_g}\times\Delta ec\_g + \beta_{ec\_p}\times\Delta ec\_p + \beta_{hr\_l}\times\Delta hr\_l + \beta_{hr\_h} + \beta_{md\_l}\times\Delta md\_l + \beta_{md\_h}\times\Delta md\_h)/MUI_i$

where in the current example

Δec_g = total stream miles × % change = 100 × .01 = 1,
Δec_p = total stream miles × % change = 100 × -.01 = -1,
Δhr_l = total stream miles × % change = 100 × .01 = 1,
Δhr_h = total stream miles × % change = 100 × -.01 = -1,
Δmd_l = total stream miles × % change = 100 × .01 = 1,
Δmd_h = total stream miles × % change = 100 × -.01 = -1.

and MUI_i is the marginal utility of income, or the negative of the summed cost coefficients evaluated at dis_i. For the first model, the average annual willingness to pay per household is $16.58 (s.e.=2.79). For the second model, the willingness to pay for a respondent living 9.58 miles from the clean-up steams' centroid (i.e., the sample mean) is $16.30 (s.e.=2.73), whereas the willingness to pay for someone living within one mile of the centroid of the cleaned-up streams is $22.76 (s.e.=3.41). To construct county-level willingness to pay for this scenario, separate *WTP* estimates for every household in the relevant county would need to be constructed and then summed.

Although not reported here, we also estimated a variety of discrete choice models within preference and "willingness to pay" space that better account for unobserved preference heterogeneity (e.g., random coefficient, latent class). Qualitatively, results from these alternative models are similar to those reported here.

Table 2.2. Choice Experiment Results Including and Testing for Distance Decay

```
Logistic regression                               Number of obs     =      7,584
                                                  Wald chi2(12)     =     514.22
                                                  Prob > chi2       =     0.0000
Log pseudolikelihood = -3683447.9                 Pseudo R2         =     0.1213

                            (Std. Err. adjusted for 1,896 clusters in resp_id)
```

ce	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
cost	-.0610311	.0046437	-13.14	0.000	-.0701326	-.0519295
cost_dis	.0401019	.0250425	1.60	0.109	-.0089804	.0891843
ec_g	.0162943	.0035968	4.53	0.000	.0092448	.0233438
ec_p	-.0151874	.0038603	-3.93	0.000	-.0227535	-.0076213
hr_l	.0114533	.0029718	3.85	0.000	.0056287	.0172779
hr_h	-.0247109	.0087397	-2.83	0.005	-.0418405	-.0075814
md_l	.0079231	.0022232	3.56	0.000	.0035657	.0122805
md_h	-.0021735	.0047079	-0.46	0.644	-.0114009	.0070539
age	-.0162039	.0028559	-5.67	0.000	-.0218013	-.0106064
gender	-.0959678	.1028224	-0.93	0.351	-.297496	.1055604
white	-.257121	.1215866	-2.11	0.034	-.4954263	-.0188157
college	.2558948	.1027845	2.49	0.013	.0544409	.4573488
_cons	1.293062	.2249336	5.75	0.000	.8522005	1.733924

Chapter 3. Expert Elicitation

The goal of the expert elicitation is to map the relationships between quantifiable, biophysical indicators of water quality and the ecological endpoints identified in the focus groups. The elicitation was developed using best practices (Merkhofer, 1987; Morgan and Henrion, 1990; Meyer and Booker, 1991; Keeney and von Winterfeldt, 1991; Kenney, 2007). These relationships were examined during virtual semi-structured interviews with eight experts who also provided their judgment of the likelihood of being in a particular category for an ecological endpoint using an elicitation survey. Dr. Melissa Kenney led the expert elicitation task in collaboration with Dr. Michael Gerst, Dr. Hillary Waters, and Dr. Alex Venning.

3.1. Methods

The expert elicitation method includes three major components: 1) data identification and development for the elicitation; 2) development of expert elicitation protocols and survey instruments; and 3) analysis of the elicitation data.

3.1.1. Data Identification and Development

To link water quality indicators to ecological endpoints, the expert elicitation survey employed a data set of 100 water quality 'rows', each with six measures that the State of North Carolina regularly uses when doing water testing – Biotic Index (BI), Fecal Coliform (FC), Specific Conductance (SC), Total Nitrogen (TN), Total Phosphorus (TP) and Turbidity (TU). Specific combinations of measurements were developed for the elicitation so that the results would be broadly representative of wadeable streams in urbanizing watersheds throughout the Southeastern U.S. The goal was not for the 100 sample rows to be representative themselves, but to test for the comparative relevance of each water quality indicator. This followed a similar methodological logic to a previous elicitation on eutrophication impairments (Kenney, 2007; Van Houtven et al., 2014). In effect, this meant over-sampling the extremes of each water quality indicator which created tricky problems where different measurements seemed to lead to different endpoint categories.

We started by organizing data from two datasets from the North Carolina Department of Environmental Quality (DEQ), filtering for urban streams, wadeable streams, the six water quality measures, and the growing season. We also ensured that the entire row of water quality measurements was taken at the same stream on the same day (with the exception of BI, which is a more consistent variable anyway). Outliers were identified and excluded - for example, water quality data is often taken downstream of wastewater plants which meant they were over-represented. A lognormal sample of 1000 data rows was generated and Z-values were calculated from this sample data. Rows were then selected based on the magnitude of Z-values for each measurement type with a third of the rows selected to oversample the extremes of each variable. Together, this ensured the elicitation would have enough sample points to test for the relative importance of each measure. This enabled the resulting ecological production function to translate biophysical measures into a probabilistic assessment of endpoint category. Additional details on the data development process is available in Appendix 3A and the 100 data rows used along with their key variables and Z-values is available in Appendix 3B. An example data row is presented in Figure 3.1.

Figure 3.1. Data Row Example

Biotic Index	Fecal Coliform (cfu/100mL)	Specific Conductance (uS/cm)	Total Nitrogen (mg/L)	Total Phosphorus (mg/L)	Turbidity (NTU)
5.72	127.2	205.0	2.56	0.87	15.7

3.1.2. Expert Elicitation Experts and Survey Protocol

The expert elicitation took place in three stages: 1) an initial interview focused on understanding experts' conceptual framing of the topic; 2) survey training; and 3) a follow-up interview to clarify and calibrate experts' final probability judgments. First, the initial interview and training focused on developing an understanding of how experts approached the relationship between water quality data and the more conceptual goals of the endpoints. We also laid out the parameters of the project--that we were focusing on urban (defined as anything not rural), wadeable streams (defined as anything year-round from an ankle-deep trickle to around waist height) in the Piedmont area of North Carolina. We then presented the experts with the publicly identified endpoints and their categories (see Appendix 2A), and experts agreed to use these definitions for the elicitation process (though a few had suggestions for future projects). We also presented experts with the set of water quality variables identified in Section 3.1.1, and asked questions about which water quality data they saw as relevant or irrelevant to particular endpoints and whether there are particular variables that could be strong proxies for a particular category, i.e., so highly correlated with or strongly predictive of an ecological endpoint that a more complex model would not be necessary. In each case, experts concluded that multiple water quality variables were important in determining each endpoint's category and identified particular variables they expected to be significant and why. We also asked the experts to identify data and research gaps that would better allow them to make their judgments. Details about expert recruitment and interview protocol are found in Appendix 3C.

Second, we provided training to set the experts up to consistently and confidently complete the elicitation survey given their judgments. The goal of the elicitation survey is to acquire data that will support the development of ecosystem production function for each endpoint -- allowing us to assess the likelihood of a stream being in a particular endpoint category given the measured water quality variables. Specifically, in the elicitation survey, experts were presented with a row of data - BI, FC, SC, TN, TP, TU - and provided with contextual information to frame their ability to answer the two questions:

> *Context: Suppose this exact data was taken from (coincidentally) 100 different urban, wadeable streams in the Piedmont region. But, you do not know exactly where or when the sample was collected. All you know is the sample was taken during the growing season, and all other factors not listed (e.g., morphological, climatic, watershed land use) vary according to your knowledge of urban, wadeable streams in the Piedmont region - for example, it rains on approximately 30% of the days during the growing season.*

*Question: If 100 urban wadeable stream samples came back with this exact data, what do you think would be the **most likely endpoint category** for the streams? How many streams, out of the 100 streams, would fall into **each endpoint category**?*

The expert would then consider a data row (Appendix 3B) and provide their judgment of the likelihood (using number of streams out of 100) that would fall within the different category levels. This task required two key translations: first, experts had to map the point-in-time sample data we gave them in the data row into the endpoint categories a stream would be placed into on average. And secondly, they had to consider the regional variability in stream size and location to consider how 100 different streams may vary in these categorizations.

During the training, experts were walked through the process and then asked to complete the exercise several times, explaining the reasoning for their answers each time. If there were any errors, confusions, or missteps -- e.g., if experts did not add up probabilities to exactly 100 or if their *most likely* answer did not correspond to the numerical answer that was most likely -- the process was clarified and gently corrected. Experts could ask questions during this time and interviewers were also able to glean more information about probable correlations and variable relevancies. Once both groups were confident in their ability to complete the elicitation, the expert was asked to complete the full elicitation survey with all 100 data rows on their own time via Qualtrics online Experts were able to complete the survey over multiple sittings and could return to previous answers. After completion, the elicitation data went through a preliminary 'soundness check' to view any outliers and choose particularly relevant data rows for further discussion.

Third, once the experts completed the elicitation, we conducted a basic data analysis to structure a follow-up interview to finalize the data set and learn more about how experts prioritized different water quality variables and their associated values to make their judgments. Preliminary data analysis, for calibration discussion purposes, primarily utilized the *most likely* answers. Ordered logistic regressions were run in R to see which variables were most statistically significant and plotted/sorted to search for any extreme outliers that could be errors, or to find 'tipping points' for further discussion. For example, if BI was statistically significant and followed a general trend of worse (higher) BI = poorer ecological condition, but one data row showed a poor BI and an ecological condition categorization as "good", that answer was flagged to review during the follow up interview. Likewise, if streams with a BI of 5.3 and less were "good" and 5.4 and more "fair" - we could ask follow-up questions about thresholds versus linearity that would help us to fully understand and analyze the data.

During the follow up interview, experts were asked to re-answer some of the elicitation questions and were afterwards given the opportunity to see their previous answers. Finally, they were asked to choose their 'best judgement' knowing both their answers. This allowed them to review any potential mistakes or changes in their judgements over the course of the survey itself. In most cases experts were remarkably consistent (spot on or under 5% discrepancy), further solidifying the method, their expertise, and their ability to reliably provide their probability judgments given the elicitation exercise. Follow-up interview questions about the relevance of particular variables were also asked with data rows. For example, we might say, "your results indicate that specific conductance is statistically significant, here are two data rows with SC that appear to contradict the general trend. Can you explain what other factors led to your judgments here?" Or, "generally high fecal coliform levels are associated with high health risk, but here the

FC is over 7,000 and your judgement was that the risk was medium. Can you explain your reasoning?" Their answers to these questions helped form a basis for the quantitative analysis decisions made later on.

3.1.3. Expert Elicitation Data Analysis Methods

The final, certified expert elicitation data is used to create biological production functions that empirically map using stressor-response models from measured biophysical measures (either measured data or modeled scenarios) to the likelihood of ecological endpoint categories. (Kenney, 2007; Van Houtven et al., 2014). As described in Van Houtven et al. (2014), the biological production functions are created using ordered probit regression models with expert fixed effects. An ordered model is chosen to take advantage of the ordinal nature of the discrete endpoints, as opposed to fitting a multinomial model.

For estimating the model, each expert data row is expanded to *n* samples, where *n* equals the number of discrete categories for the endpoint. A weight is assigned to each sample corresponding to the likelihood assigned by the expert. For example, in a three-category endpoint, if an expert assigned likelihoods of 0.70, 0.20, and 0.10, then the category with a likelihood of 0.70 would have the highest weight of 0.70, while the one with a likelihood of 0.10 would have the lowest weight of 0.10. To account for correlation among error terms within experts, robust standard errors are estimated by treating experts as clusters. To facilitate interpretation of the results, data is z-scored so that coefficients on variables with different scales can be more easily compared.

3.2. Results

Tables 3.1, 3.2, and 3.3 report order probit parameter estimates for the ecosystem condition, human health risk, and murky water days expert elicitations. Each model controls for unobserved expert-specific factors through fixed effects. These fixed effects were jointly statistically significant in all models, although our results are largely the same if they are excluded. For compactness, the fixed effects are not reported.

3.2.1. Ecosystem Condition

For ecosystem condition, five experts completed the expert elicitation. All variables are statistically significant, and as expected, biotic index is the strongest predictor. A one standard deviation increase in biotic index leads to a 0.673 ($p < 0.001$) increase in the probit index for ecosystem condition: a lower biotic (increase) is predictive of worse ecosystem condition. Turbidity is the second strongest predictor, with a one standard deviation increase in biotic index leading to a 0.211 ($p < 0.05$) increase in the ordered probit index.

Table 3.1. Ecosystem Condition Ordered Probit Regression Results

Log pseudolikelihood = -40062.104					Pseudo R2 = 0.2358	
	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
bi	.6731241	.1168501	5.76	0.000	.4441021	.9021461
fc	.1691213	.0645956	2.62	0.009	.0425162	.2957263
sc	.157963	.0138884	11.37	0.000	.1307422	.1851839
tn	.1290729	.0213769	6.04	0.000	.0871749	.1709709
tp	.0684721	.0301225	2.27	0.023	.0094331	.1275111
turb	.2098882	.0837005	2.51	0.012	.0458381	.3739383
/cut1	-1.400093	.0603983			-1.518472	-1.281715
/cut2	.3269086	.0370513			.2542894	.3995278

3.2.3. Human Health

For the human health endpoints, three experts completed the expert elicitation. All indicators except for total phosphorus are statistically significant. As expected, fecal coliform is the strongest predictor, with a one standard deviation increase in fecal coliform leading to a 0.621 ($p < 0.01$) increase in the ordered probit index for human health. The second and third strongest predictors are turbidity and specific conductance, at 0.337 ($p < 0.05$) and 0.299 ($p < 0.001$), respectively. As with the other endpoints, there is a significant amount of diversity among experts with respect to the distribution of data rows across human health categories.

Table 3.2. Human Health Risk Ordered Probit Regression Results

Log pseudolikelihood = -25131.389					Pseudo R2 = 0.2301	
cat	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
bi	.1755935	.0443074	3.96	0.000	.0887525	.2624344
fc	.6211938	.2310702	2.69	0.007	.1683046	1.074083
sc	.2985876	.0550163	5.43	0.000	.1907576	.4064175
tn	.0873225	.0166922	5.23	0.000	.0546063	.1200386
tp	-.0231862	.0320875	-0.72	0.470	-.0860765	.0397042
turb	.3368817	.143512	2.35	0.019	.0556034	.61816
/cut1	-.7157802	.0387267			-.7916831	-.6398772
/cut2	.5724916	.0508156			.4728949	.6720883

3.2.2. Murky Water Days

For murky water days, five experts completed the expert elicitation. All indicators except for specific conductance are statistically significant. As expected, turbidity is the strongest predictor, with a one standard deviation increase in turbidity leading to a 0.600 ($p < 0.001$) increase in the ordered probit index for murky water. Biotic index is the second strongest predictor, with a one standard deviation increase in biotic index leading to a 0.234 ($p < 0.01$) increase in the ordered probit index.

Table 3.2. Murky Water Days Ordered Probit Regression Results

Log pseudolikelihood = -37004.628 Pseudo R2 = 0.2241

cat	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
bi	.3228819	.1005827	3.21	0.001	.1257435	.5200203
fc	.1610447	.0527185	3.05	0.002	.0577182	.2643711
sc	.1258448	.0642222	1.96	0.050	-.0000284	.251718
tn	.1209292	.0301164	4.02	0.000	.0619022	.1799562
tp	.1134838	.0406304	2.79	0.005	.0338497	.1931179
turb	.5687407	.1016695	5.59	0.000	.3694721	.7680092
/cut1	-1.141596	.0513119			-1.242165	-1.041026
/cut2	.5606499	.0287737			.5042545	.6170453

Chapter 4. Water Quality Modeling

4.1. Objectives

We developed a hierarchical modeling approach for assessing and predicting water quality in wadeable streams (Miller et al., 2019; Miller 2019; Miller et al., *In Prep.*). Specifically, we assessed how natural and anthropogenic watershed variables are related to a stream's biotic index (BI), turbidity (TDU), specific conductance (SC), fecal coliform (FC), total nitrogen (TN), and total phosphorus (TP) concentrations. These water quality indicators were chosen based on data availability and the capacity of expert elicitation to connect them to ecological endpoints. We demonstrate the utility of this approach for watersheds in the central North Carolina (NC) Piedmont (Figure 4.1). The models were also used to predict water quality conditions under baseline conditions as well as for future management scenarios. These predictions were developed to inform expert elicitation models, which connect the predictions to ecological endpoints.

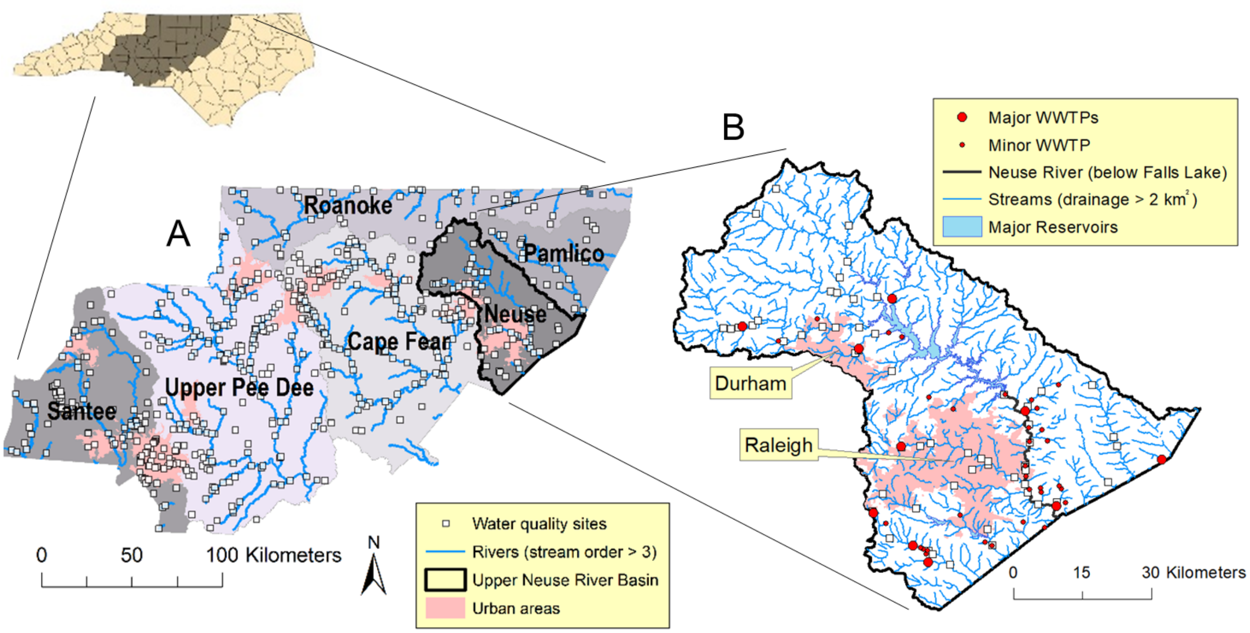


Figure 4.1: (A) Locations of 653 water quality monitoring sites within the six HUC 8 watersheds that intersect the NC Piedmont (shaded gray). (B) The Upper Neuse River Basin stream network where water quality forecasts for potential management scenarios were explored. Also shown are wastewater treatment plants (WWTPs). The main stem of the Neuse River below Falls Lake (black) was considered unwadeable and therefore omitted from the modeling. From Miller (2019).

4.2. Methods

Hierarchical modeling (a.k.a. mixed-effects or multi-level modeling) is an extension of conventional regression modeling that includes both random and fixed effects (Gelman and Hill 2006; Qian et al., 2010). Fixed effects are conventional regression coefficients that broadly characterize relationships between the response variable and various predictor variables (Faraway 2014), while hierarchical random effects account for additional variability in the response across different groups (e.g., watersheds). Initially, baseline models were created for each indicator using impervious cover (fixed effect) and site and basin random effects. Then, predictors from 16 different variable classes representing land covers, natural factors (e.g., soils, hydro-climatology), and point sources were evaluated as potential auxiliary variables using an information criterion approach and an exhaustive search procedure (Faraway 2014). The resulting hierarchical multiple linear regressions (HMLRs; one for each water quality indicator) could then be used to assess the relative importance of the various natural and anthropogenic watershed predictors. Finally, water quality was predicted throughout a subsection of the study area, the Upper Neuse River Basin (UNRB), for current conditions and future watershed management scenarios to evaluate potential improvements in water quality. See Miller (2019) for more details.

4.3. Findings and Results

HMLRs were developed for each of the six water quality indicators. A majority of the variance (i.e., R^2>0.5) in BI and TN is explained by the statistically-selected deterministic (anthropogenic and natural) predictor variables alone (Table 1). Performance for other indictors is lower, but improves if aggregated at the annual level, or if random effects are included (Table 4.1).

In general, impervious cover (IC) in watersheds, canopy loss in stream buffers, point source loadings (i.e., wastewater effluent), and croplands were found to be the most frequently selected anthropogenic predictors of stream water quality and ecological health, while hydro-climatology was the most salient natural predictor (Table 4.1). At the same time, results indicate substantial variation in the natural and anthropogenic factors most relevant to different water quality indicators. Specifically, the BI model shows that the two largest anthropogenic factors affecting stream biodiversity are impervious cover in upstream watersheds and canopy loss in stream buffers. The main driver of variability in both FC and TDU is short-term (i.e., 1-7 day) precipitation. The FC model also indicates pastureland in stream buffers and older IC as important drivers, while the TDU model indicates the impact of recent development. For TN and TP, WWTP discharges are found to be of particular importance. Finally, for SC, both older IC and WWTPs appear to be substantial drivers.

The resulting models can be used to predict water quality across both space and time, which is important because water quality monitoring is often sparse and infrequent. Throughout the UNRB study area, the model was used to estimate water quality under normal and wet hydrologic conditions (Figure 4.2A, 4.2B). Furthermore, the modeling results can be used to assess the benefits of potential watershed management scenarios (Figure 4.2C, 4.2D, 4.2E). In particular, we explored how restoring buffer vegetation, reducing IC impacts, improving WWTPs, and addressing site-specific issues (i.e., represented by site-specific random effects) might improve water quality in our study area. See Miller (2019) for more details.

Table 4.1: Multi-stressor HMLRs for BI, FC, TDU, TN, TP, SC. The variance explained (R^2) by the deterministic portions of the model (predictors, soil and seasonal effects) as well as the full models (with random effects) is provided. Deterministic variance was also calculated averaging samples on a yearly basis at each site (except for BI where sampling did not occur more than once a year). Predictor coefficients (β; unitless) are represented as the standard deviation of their marginal contribution to the predicted response normalized by the standard deviation of the observed response. In this way, coefficients in the table can be compared vertically (within a model) as well as horizontally (across models) to see their relative influence on water quality indicators. The term "basin" represents the % of overall watershed while "buffer" represents the % within a 30-m (per side) stream buffer. Categorical predictors (α) cannot be classified as positive or negative. Of note, the BI model had a quadratic trend (q) for long-term precipitation. Numbers in parentheses for IC (recent) and hydro-climatology (short and long) refer to the number of days, months, or years that define that predictor. For IC (age), the number represents IC constructed before that given date (e.g., 80 represents IC pre-1980). From Miller (2019).

		BI	FC (#/100mL)	TDU (NTU)	TN (ppm)	TP (ppm)	SC (uS/cm)
R^2	Deterministic	0.76	0.25	0.28	0.54	0.32	0.42
	Deterministic (yearly)	-	0.36	0.28	0.68	0.47	0.55
	Deterministic + site + basin	0.92	0.40	0.40	0.72	0.58	0.70
β	Canopy loss, buffer	+0.19	+0.13		+0.16	+0.07	+0.05
	Crop, buffer		-0.07				
	Crop, basin			+0.09	+0.10		-0.18
	Drainage		-0.17				
	Hydro-clim, short (days)		+0.30 (3)	+0.41 (3)	-0.06 (14)	-0.07 (7)	-0.14 (7)
	Hydro-clim, long (months)	-0.05(3; q)	+0.06 (1)	+0.12 (1)	-0.06 (6)	-0.07 (6)	-0.12 (6)
	IC, buffer				-0.10		
	IC, basin	+0.21					
	IC, age (year)	+0.10 (80)	+0.26 (00)				+0.32 (80)
	IC, recent (# of years)			+0.12 (5)		+0.20 (2)	
	Pasture, buffer		+0.19				
	Reservoirs, 5 km	+0.05					
	WWTP loadings		+0.04		+0.66	+0.47	
	# of Major WWTP						+0.36
	WWTP proximity, 1 km	+0.05					
	Wetlands, buffer				-0.06		+0.06
	Year	-0.07	-0.04		+0.04		
α	Sampling method	0.26					
	Season	0.08	0.15	0.16	0.06	0.13	0.12
	Soil type	0.19	0.16	0.14	-	-	0.26
	Site random effects	0.12	0.37	0.19	0.22	0.23	0.29
	Basin random effects	0.16	0.00	0.18	0.25	0.30	0.29
	SD of response variable	1.49	2.76	0.86	0.45	0.09	0.44

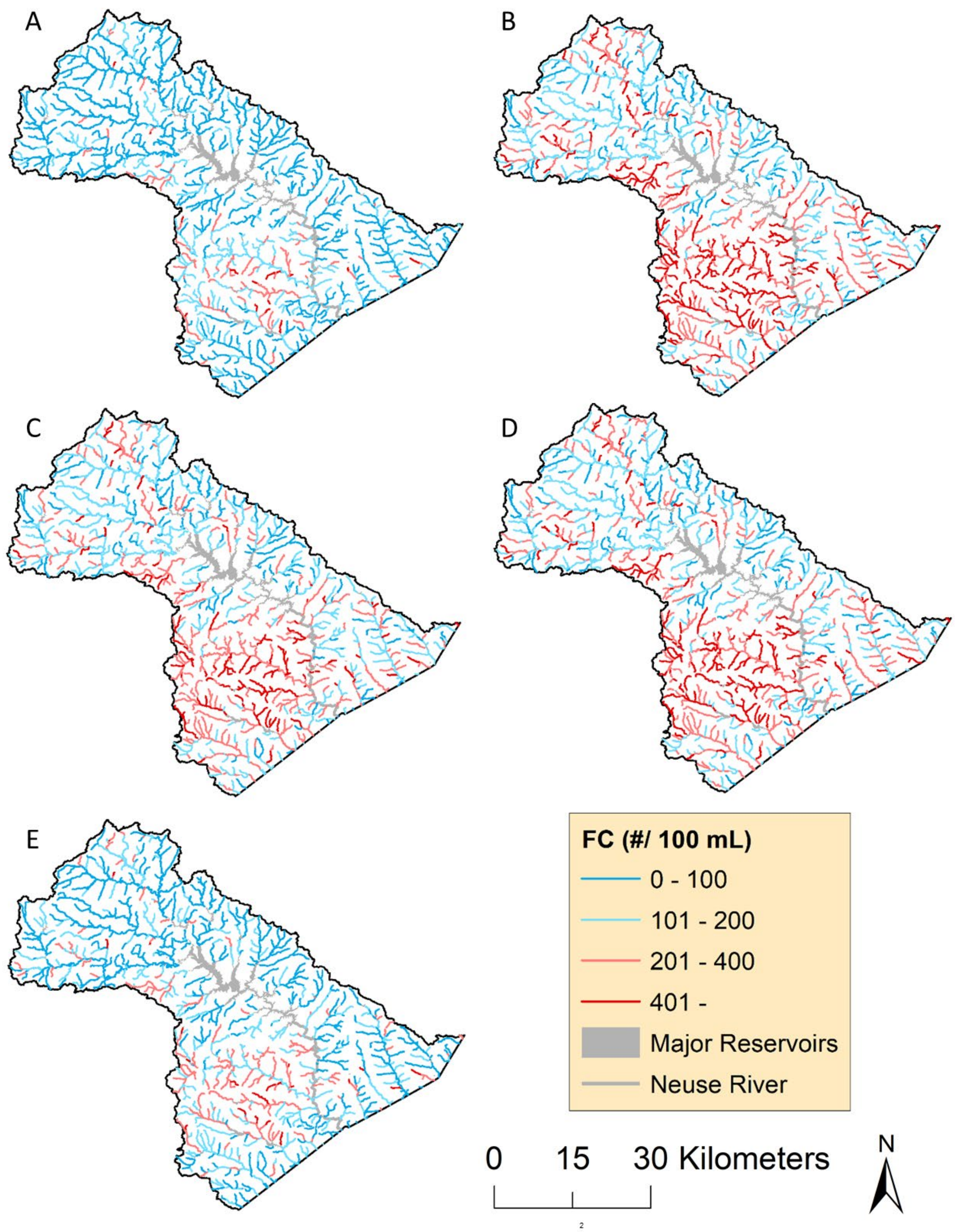


Figure 4.2: Mean FC predictions throughout the Upper Neuse River Basin (UNRB) for current median (A) and wet (B) hydro-climatological conditions. Results are also shown for selected management scenario under wet weather conditions, including (C) 50% reduction in canopy loss, (D) 25% reduction in the effect of IC, and (E) combination of 50% reduction in canopy loss, 25% reduction in IC, and 25% reduction in WWTP. From Miller (2019).

4.4. Summary and Implications

This study demonstrates a data-driven approach for assessing and predicting water quality in wadeable streams. The approach is found to be applicable across a wide range of water quality indicators. It provides an alternative to more mechanistic modeling, which may not be feasible for larger study areas and more complex indicators (e.g., biotic index). The models developed here go beyond relatively simple relationships with impervious cover, to also identify deterministic relationships with stream buffer condition, WWTP discharge, hydro-climatology, age of development, and other watershed attributes. In general, the identified relationships are consistent with a process-based understanding of pollutant sources, fate, and transport. For example, turbidity and phosphorus are uniquely associated with recent development, consistent with construction activities mobilizing sediment and sediment-bound particulate phosphorus.

In addition to the deterministic relationships identified in the HMLRs, the site-specific random effects can also inform watershed management. Water quality in urban areas is degraded by a number of anthropogenic factors, which are not always easy to identify or remedy. The random effects allow for a comparison of expected (i.e., predicted) water quality conditions with actual water quality conditions (Miller et al., 2019). The magnitude of the random effects is also influenced by the number of observations (as more data increases our confidence in potential deviations from expected conditions). Thus, the random effects can help managers identify and prioritize watersheds where additional investigation may be warranted. For example, sites with large positive random effects (indicating worse water quality than expected) can be investigated for localized pollutant sources (e.g., leaking wastewater).

Finally, while the results shown in Figure 4.2 are based on mean predictions, the statistical nature of this modeling approach also provides uncertainty quantification (Miller, 2019; Miller et al., *In prep*.). Results can be presented in terms of the probability of exceeding various water quality criteria (i.e., state or federal standards) or in terms of the probability of observing improvements in water quality under various future scenarios. Thus, watershed improvement scenarios can be evaluated in a risk-based management framework.

Chapter 5. Case Study – Upper Neuse River Basin (UNRB)

In this final chapter we discuss results from the Upper Neuse River Basin (UNRB) case study that we developed to illustrate how our research efforts can be integrated to address policy relevant questions. We consider two scenarios:

> **Scenario #1** - a 25 percent increase in canopy cover combined with a 25 percent decrease in the negative effects of impervious cover.
> **Scenario #2** - a 25 percent reduction in the site and random effects.

The first scenario is consistent with an ambitious streambank restoration policy combined with more aggressive stormwater runoff interventions for roads and the built environment, while the second scenario envisions a policy whereby local managers repair idiosyncratic sources of pollution (e.g., wastewater leaks) identified through extensive monitoring efforts. We apply these scenarios to two sub-watersheds in Wake County – the Crabtree and Walnut Creek sub-watersheds that run through the central and most populous parts of the county and were the focus of the stated preference survey and the Swift and Middle Creek sub-watersheds that run through southern Wake County, is about 22% smaller in terms of stream miles, more suburban, and more rapidly growing.

To begin, we consider baseline water quality conditions in both counties. The first two rows of Table 5.1 reports baseline values under "normal"[2] weather conditions for the six indicators for both sub-watersheds that were generated with the Hierarchical Bayesian water quality model described in Chapter 4. The remaining rows show how these indicators change under the two scenarios.

Table 5.1. Water Quality Indicators: Average Values by Sub watershed and Scenario

Sub-watersheds	Scenario	Biotic Index	Turbidity	Fecal Coliform	Total Nitro.	Total Phos.	Spec. Cond.
Crabtree & Walnut	Baseline	6.78	9.28	217.03	1.32	0.28	132.64
Middle & Swift	Baseline	6.44	9.95	170.07	1.96	0.25	123.87
Crabtree & Walnut	#1	6.35	8.87	161.00	1.24	0.27	126.94
Middle & Swift	#1	6.01	10.64	116.76	2.17	0.26	117.74
Crabtree & Walnut	#2	6.69	8.82	198.46	1.22	0.26	125.74
Middle & Swift	#2	6.27	10.60	136.47	2.14	0.25	115.74

We use the ordered probit results from the expert elicitation reported in Tables 3.1, 3.2, and 3.3 to translate these indicators into the ecosystem condition, human health risk, and murky water day endpoints. The first two rows of Table 5.2 report the percentage of stream miles that fall

[2] "Normal" here is defined in terms of long-term precipitation (mean levels) and short-term precipitation (none).

into each category for each attribute in baseline conditions. In general, the endpoints imply slightly better baseline conditions in Middle and Swift Creeks relative to Crabtree and Walnut Creeks. The remaining rows in Table 5.2 report how the percentages change under both scenarios. The improvements are generally greater for Scenario #1.

Table 5.2. Ecological Endpoints (percentage of watershed stream miles in each category)

Sub-watershed	Scenario	Eco. Cond. Good	Eco Cond. Fair	Eco. Cond. Poor	Health Risk Low	Health Risk Medium	Health Risk Low	Murky Water Days Low	Murky Water Days Medium	Murky Water Days High
Crabtree & Walnut	Baseline	9.83	57.02	33.15	64.89	30.37	4.74	73.39	25.04	1.57
Middle & Swift	Baseline	15.11	60.54	24.35	67.00	28.80	4.20	75.33	23.32	1.35
Crabtree & Walnut	#1	19.18	61.21	19.60	69.92	26.57	3.51	78.44	20.53	1.03
Middle & Swift	#1	25.13	60.33	14.53	71.22	25.55	3.23	79.56	19.50	0.93
Crabtree & Walnut	#2	11.94	58.91	29.16	66.84	28.93	4.24	75.31	23.34	1.35
Middle & Swift	#2	17.96	61.15	20.89	68.90	27.35	3.74	77.24	21.61	1.15

Finally, we report per household and Wake County aggregate willingness to pay measures for the two scenarios in Table 5.3. We use the econometric model that assumes a constant marginal utility of income across the population (i.e., Table 2.1) in this case, although similar results are implied if we used the model that allowed for distance decay (i.e., Table, 2.2). These measures are significantly higher for Scenario #1 relative to Scenario #2 and for Crabtree and Walnut Creeks relative to Swift and Middle Creeks. The latter result is driven in part by the fact that Crabtree and Walnut Creek scenarios impact roughly 100 stream miles whereas the Swift and Middle Creek scenarios impact only 78.5 stream miles.

Table 5.3. Annual Willingness to Pay Measures

	Per Household WTP		Aggregate WTP (millions)	
Sub-watersheds	Scenario #1	Scenario #2	Scenario #1	Scenario #2
Crabtree & Walnut	$110.41	$33.27	$44.2	$13.3
Swift & Middle	$75.44	$26.33	$30.2	$10.8

The point estimates in Table 5.3 do not include standard errors or confidence intervals. These are somewhat challenging to construct given that we have three separate estimations that feed into our welfare estimates and that one of three estimations relies on a Bayesian approach while the other two rely on classical approaches. In ongoing work, we are working to integrate these estimations in a consistent manor.

References

Faraway, J.J. (2014). *Linear models with R*: *Second edition*. Chapman and Hall/CRC.

Gelman, A., and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Keeney, R.L. and Von Winterfeldt, D. (1991). Eliciting probabilities from experts in complex technical problems. *Transactions on Engineering Management*, 38: 191–201.

Kenney, M.A. (2007). Which nutrient criteria should states and tribes choose to determine waterbody impairment? Using science and judgments to inform decision-making, Ph.D., Dissertation, Division of Environmental Sciences and Policy, Duke University.

Merkhofer, M.W. (1987) Social risk management. In: *Decision science and social risk management. Technology, risk, and society (An international series in risk analysis), vol 2*. Springer, Dordrecht.

Meyer, M.A. and Booker, J.M. (1991). *Eliciting and analyzing expert judgment: A practical guide*, London: Academic Press Ltd.

Miller, J.W. (2019). *Assessing natural and anthropogenic drivers of regional water quality using hierarchical modeling*. Ph.D. Dissertation, Department of Civil and Environmental Engineering, North Carolina State University.

Miller, J.W., Paul, M.J., and Obenour, D.R. (2019). Assessing potential anthropogenic drivers of ecological health in Piedmont streams through hierarchical modeling. *Freshwater Science*, *38*(4): 771-789.

Miller, J.W., Paul, M.J., and Obenour, D.R. (*In Prep.*). Forecasting water quality in urban streams under future management scenarios through hierarchical modeling.

Morgan, M.G. and Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*, Cambridge: Cambridge University Press.

Phaneuf, D.J., von Haefen, R.H., Van Houtven, G., Mansfield, C. (2013). "Measuring nutrient reduction benefits for policy analysis using linked non-market and environmental assessment models," Final Report under Grant #X7-83381001-0 from the U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Qian, S.S., Cuffney, T.F., Alameddine, I., McMahon, G., and Reckhow, K H. (2010). On the application of multilevel modeling in environmental and ecological studies. *Ecology*, 91(2):355-361.

Van Houtven, G., Mansfield, C., Phaneuf, D.J., von Haefen, R.H., Milstead, B., Kenney, M.A., and Reckhow, K.H. (2014). Combining expert elicitation and stated preference methods to value ecosystem services from improved lake water quality. *Ecological Economics,* 99:40-52.