



# Suburban watershed nitrogen retention: Estimating the effectiveness of stormwater management structures

Benjamin J. Koch<sup>1,2,3\*</sup> • Catherine M. Febria<sup>1,2,4</sup> • Roger M. Cooke<sup>5,6,7</sup> • Jacob D. Hosen<sup>1,2</sup> •  
Matthew E. Baker<sup>8</sup> • Abigail R. Colson<sup>6,9</sup> • Solange Filoso<sup>1</sup> • Katharine Hayhoe<sup>10</sup> • J.V. Loperfido<sup>11,12</sup> •  
Anne M.K. Stoner<sup>10</sup> • Margaret A. Palmer<sup>1,2,13</sup>

<sup>1</sup>Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomons, Maryland, United States

<sup>2</sup>Department of Entomology, University of Maryland, College Park, Maryland, United States

<sup>3</sup>Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona, United States

<sup>4</sup>School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

<sup>5</sup>Resources for the Future, Washington, District of Columbia, United States

<sup>6</sup>Department of Management Science, University of Strathclyde, Glasgow, Scotland

<sup>7</sup>Department of Mathematics, Delft University of Technology, Delft, Netherlands

<sup>8</sup>Department of Geography and Environmental Systems, University of Maryland, Baltimore County, Baltimore, Maryland, United States

<sup>9</sup>Center for Disease Dynamics, Economics & Policy, Washington, District of Columbia, United States

<sup>10</sup>Climate Science Center, Texas Tech University, Lubbock, Texas, United States

<sup>11</sup>U.S. Geological Survey, Eastern Geographic Science Center, Reston, Virginia, United States

<sup>12</sup>McAdams Company, Durham, North Carolina, United States

<sup>13</sup>National Socio-Environmental Synthesis Center, University of Maryland, Annapolis, Maryland, United States

\*ben.koch@nau.edu

## Abstract

Excess nitrogen (N) is a primary driver of freshwater and coastal eutrophication globally, and urban stormwater is a rapidly growing source of N pollution. Stormwater best management practices (BMPs) are used widely to remove excess N from runoff in urban and suburban areas, and are expected to perform under a wide variety of environmental conditions. Yet the capacity of BMPs to retain excess N varies; and both the variation and the drivers thereof are largely unknown, hindering the ability of water resource managers to meet water quality targets in a cost-effective way. Here, we use structured expert judgment (SEJ), a performance-weighted method of expert elicitation, to quantify the uncertainty in BMP performance under a range of site-specific environmental conditions and to estimate the extent to which key environmental factors influence variation in BMP performance. We hypothesized that rain event frequency and magnitude, BMP type and size, and physiographic province would significantly influence the experts' estimates of N retention by BMPs common to suburban Piedmont and Coastal Plain watersheds of the Chesapeake Bay region.

Expert knowledge indicated wide uncertainty in BMP performance, with N removal efficiencies ranging from <0% (BMP acting as a source of N during a rain event) to >40%. Experts believed that the amount of rain was the primary identifiable source of variability in BMP efficiency, which is relevant given climate projections of more frequent heavy rain events in the mid-Atlantic. To assess the extent to which those projected changes might alter N export from suburban BMPs and watersheds, we combined downscaled estimates of rainfall with distributions of N loads for different-sized rain events derived from our elicitation. The model predicted higher and more variable N loads under a projected future climate regime, suggesting that current BMP regulations for reducing nutrients may be inadequate in the future.

## Domain Editor-in-Chief

Donald R. Zak, University of Michigan

## Associate Editor

Julian D. Olden, University of Washington

## Knowledge Domains

Earth & Environmental Science  
Ecology

## Article Type

Research Article

## Part of an *Elementa* Special Feature

Urban Aquatic Ecosystems

Received: February 26, 2015

Accepted: July 6, 2015

Published: July 30, 2015

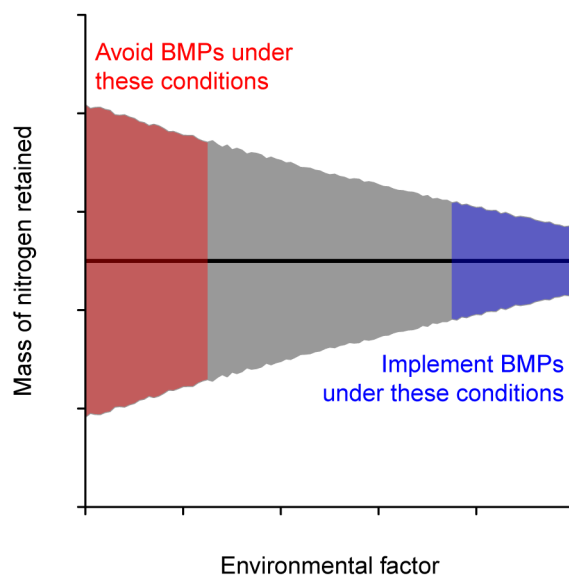
## Introduction

Excess nitrogen from human activities is one of the leading contributors to freshwater and coastal eutrophication worldwide (Conley et al., 2009). Managing nitrogen-rich runoff from the landscape is a top priority for water quality regulators in many regions, although the costs can be substantial especially under projected future climate change (Rabotyagov et al., 2014). Despite this prioritization, the science underlying where and why some nitrogen retention practices are maximally effective is still an active area of research with many knowledge gaps (Giri et al., 2014; Koch et al., 2014). For example, best management practices (BMPs) such as regenerative stormwater conveyance structures and detention ponds intended to control nutrient runoff in urban areas are widely implemented, despite a lack of quantitative data demonstrating they can remove a significant amount of nitrogen (Filoso and Palmer, 2011; Palmer et al., 2014).

BMPs are constructed in heterogeneous landscapes defined by regionally and temporally varying hydrologies, climates, and geomorphologies. While environmental context is thought to strongly influence the performance of stormwater best management practices (SW BMPs; Hunt et al., 2012; Loperfido et al., 2014), there is insufficient data to test that hypothesis, much less make management decisions (Koch et al., 2014). As a result, BMPs are often assumed to function with the same efficiency in all settings (CBP, 2013) and this assumption is built into water quality models that establish the basis for setting nutrient reduction targets and water pollution regulations (e.g., PLRM Development Team, 2009; USEPA, 2010a). Lacking data to validate this assumption, there is the potential for substantial mismatch between how much nitrogen regulators believe is being removed and the amount actually retained by a BMP. Likewise, determining the optimal biophysical environment and placement for individual BMP designs can improve nutrient mitigation practices without added costs (Chiang et al., 2014) by enabling water quality managers to implement BMPs where they are most likely to achieve the highest performance.

Clearly, long-term empirical studies of BMP performance are needed to better define the relationships between nutrient retention and site-specific environmental variables for different types of BMPs. However, until those data are available, managers must continue making decisions about which types, how many, and where to build SW BMPs that will achieve mandated nutrient reduction goals. One potential solution to this dilemma is to use a risk-management approach to quantify the degree of uncertainty in the performance of different BMPs. Defining the bounds of uncertainty in BMP performance would give managers a clear picture of the variability in nutrient retention, thereby enabling them to make more informed decisions that minimize the risk of not meeting a particular water quality target or over-investing in a particular BMP (Walker and Selman, 2014).

While better defining the bounds of scientific uncertainty in performance would not provide an unambiguous recipe for SW BMP implementation, estimating the expected variability of BMP performance under different environmental conditions would provide two distinct avenues for facilitating wiser management of limited pollution-reduction funds. First, if it is possible to identify site-specific factors that result in high uncertainty in BMP performance, then managers may be able to avoid implementing BMPs when those conditions exist (Figure 1). Similarly, if it is possible to identify factors or situations that yield relatively low uncertainty in BMP performance, then managers may choose to take advantage of those factors when implementing BMPs. Second, if it is possible to estimate the uncertainty in BMP performance, then researchers



**Figure 1**

Conceptual figure illustrating how knowledge of uncertainty (gray) around BMP performance (black line) due to different environmental factors can improve management decisions.

Water resource managers can capitalize on environmental conditions that confer a narrow range in expected N retention (blue) by implementing BMPs when those conditions exist. Conversely, managers may choose to avoid implementing BMPs when environmental factors confer a high degree of uncertainty in their performance (red).

doi: 10.12952/journal.elementa.000063.f001

can account for that uncertainty by building it into model-based estimates of nutrient loading. Doing so would enable managers to ask: “Given the estimated level of variability in performance, what degree of BMP implementation would it take to achieve a particular water quality target with a specified level of certainty?” Managers can then evaluate whether such actions are feasible or cost-prohibitive, or whether alternative nutrient reduction strategies may be more cost-effective. For example, it is possible that high uncertainty in the real-world performance of SW BMPs may suggest that managers invest more heavily in other, proven practices for achieving water quality goals like agricultural BMPs, disconnecting impervious surfaces from stream channels, or reducing atmospheric inputs of N.

Here, we employ the technique of structured expert judgment (SEJ) to robustly estimate the expected variability of BMP performance under a range of site-specific environmental conditions. SEJ has been used widely in the field of decision science to inform management decisions where empirical data are lacking or are impossible to collect (e.g., Bamber and Aspinall, 2013; Cooke and Goossens, 2004; Rothlisberger et al., 2012; Wittmann et al., 2014, 2015).

In earlier work, we quantified the variability in nitrogen removal efficiency by different types of BMPs using data from empirical studies and found we could only explain a small fraction of that variability because few studies provided information on the environmental context in which their measurements were made (Koch et al., 2014). For this reason, in this study, we turned to performance-weighted SEJ to determine how experts who study, implement, or prioritize BMP placement believe environmental factors drive performance variability. Our elicitation tool was based on the hypotheses that rain event frequency and size, BMP type and size (relative to watershed drainage area), and physiographic province would significantly influence the experts’ judgments on the N-retention efficiencies of BMPs common to the Chesapeake Bay watershed. Although additional environmental factors such as soil permeability and vegetation type may also drive variability in BMP performance, these were not explicitly examined in the elicitation. Specifically we focused on dry ponds, wet ponds, wetlands, bioretention cells, and in-channel regenerative stormwater conveyance structures in suburban headwater catchments. Our findings provide relevant information for local water resource managers seeking to understand the uncertainties in meeting nutrient reduction targets for impaired waters, including the Chesapeake Bay. Furthermore, our analysis reveals the extent to which nutrient retention by stormwater BMPs is likely to be altered under projected climate change scenarios.

## Methods

### *The structured expert judgment approach*

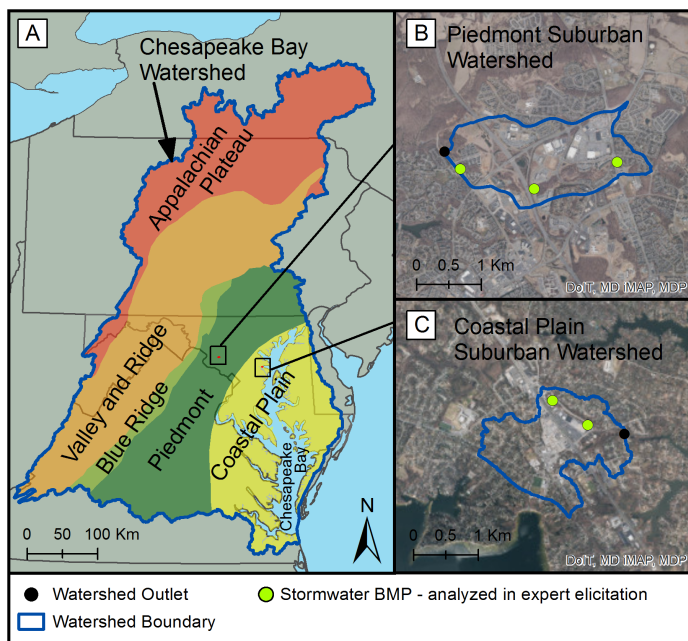
Structured expert judgment (SEJ) is a method for quantifying uncertainty. Here we use SEJ to estimate the uncertainty in stormwater BMP performance under a range of environmental conditions. To date, SEJ has not been applied widely in ecological studies (but see Bamber and Aspinall, 2013; Rothlisberger et al., 2010, 2012; Teck et al., 2010; Wittmann et al., 2014, 2015). Nonetheless, over the last 20 years, the approach has been used to successfully estimate uncertainty in many disciplines including nuclear safety, volcanic forecasting, and human health (Cooke and Goossens, 2008); and it is well established within the discipline of risk assessment. While SEJ cannot supplant the critical need for more empirical research on, e.g., BMP performance, it can be a very helpful tool for quantifying uncertainty in the absence of empirical data. A comprehensive justification of SEJ theory and methodology is beyond the scope of this study; however, we give a concise overview of the approach below. We provide a more thorough treatment of the technique’s theoretical basis, methods, and common misconceptions in Appendix S1. Comprehensive explanations of the method can also be found in Cooke (1991), Aspinall (2010), and Aspinall and Cooke (2013).

### *Elicitation process*

We conducted a SEJ elicitation using the ‘Classical Model’ (Cooke, 1991). We interviewed 10 experts in stormwater management (Appendix S2) and elicited quantitative estimates of uncertainty for nitrogen loads under various watershed conditions. This type of SEJ aggregates the estimates for all experts according to their performance on a subset of calibration variables, for which true nitrogen loads are known to the elicitors, but unknown to the experts. For each question posed in the interview (Appendix S3), the resulting performance-weighted uncertainty estimate reflects the optimal balance of accuracy and information for the entire group of experts (Cooke, 1991; Wittmann et al., 2015).

We sought to represent the widest possible skill set among the experts selected to participate in the elicitation. We therefore compiled a list of candidates with stormwater BMP-related experience and assigned to each a primary area of expertise: practitioner, modeler, engineer, or research scientist. We maintained a roughly even distribution among those categories in confirming the final group of experts for the SEJ.

Several weeks prior to the interviews, we provided experts with a written protocol document (Appendix S3) containing 60 questions on total nitrogen (TN) loading rates within two small (1.5–3.5 km<sup>2</sup>) suburban watersheds in the Chesapeake Bay region (Figure 2, Appendix S3). Experts were asked to provide estimates



**Figure 2**

Map of the two suburban watersheds that were the focus of the elicitation.

The study areas were located in the Piedmont and Coastal Plain physiographic provinces within the Chesapeake Bay watershed (A). The stormwater BMPs analyzed in the expert elicitation are shown for the Piedmont watershed (B) and the Coastal Plain watershed (C).

doi: 10.12952/journal.elementa.000063.f002

of the 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles of the probability distribution envisioned to best describe the TN load for each question. In addition to the questions, the protocol contained detailed background data and information about the watersheds and the BMPs within them. We also provided additional documents with relevant data on nitrogen fluxes and stormwater control in the Chesapeake Bay as well as information on the theory and methodologies of SEJ (all materials available upon request). Experts were encouraged to use any additional resources they judged to be helpful in arriving at their best estimates for each question.

We designed the protocol document to minimize the cognitive burden required of the experts to easily distinguish the 60 questions. Accordingly, the protocol was divided into two sections, each corresponding to one of the physiographic provinces. Questions were further grouped by precipitation events (Appendix S3). Such organizational choices have the potential to influence the variables deemed important by the experts; however, the wide range of rationales provided by the experts indicated that such potential biases were minimal.

Eleven of the 60 questions in the protocol document were calibration variables, where the true TN load was known to the elicitors, but unknown to the experts at the time of the interview. In all cases those values came from unpublished data shared with us by researchers working in the focal watersheds. Experts were not aware of which questions in the protocol document represented calibration variables.

Elicitors (2-4) conducted an in-person interview individually with each expert to record the rationale, assumptions, and data sources used in deriving each answer within the protocol document. All experts completed the elicitation prior to the in-person interview; we used the 1-3 hour interview to review with each expert his reasoning, calculations, spreadsheets, and/or computer code.

### *Expert calibration and information scoring*

To calculate a performance-based combination of expert estimates, each expert received a *calibration* score and an *information* score, which were used to assign weights to each expert's assessment. *Calibration* measures the statistical likelihood that the set of known values for the calibration variables correspond with an expert's assessment of those variables. Thus, the calibration score is calculated as the probability (p-value) of falsely rejecting the hypothesis that the expert's uncertainty estimates were accurate (Cooke, 1991; Wittmann et al., 2015). Calibration scores take values between 0 and 1 and signify the degree to which an expert's uncertainty estimates (5<sup>th</sup>, 50<sup>th</sup>, 95<sup>th</sup> percentiles) are accurate relative to the calibration variables. A low calibration score, near 0, indicates that the expert's estimates are unlikely to be correct. A high score, near 1, indicates good support for the expert's estimates. Further details of the properties of calibration scores and how they are calculated are provided in Appendix S1.

*Information* scores are analogous to measures of precision for each expert's estimates. Experts provided estimates of uncertainty (e.g., 5<sup>th</sup>, 50<sup>th</sup>, 95<sup>th</sup> percentiles), and the distributions defined by those percentiles may be very narrow (signifying a high degree of certainty) or very wide (signifying a low degree of certainty). Information scores provide a way of assessing the width of those distributions by comparing them to a common reference range (Cooke, 1991; Wittmann et al., 2015). In the present study, the reference range was set separately for each question to be the maximum 90% interval (i.e., 5<sup>th</sup> to 95<sup>th</sup> percentile) given by any expert, plus an additional 10% of that maximum interval. Unlike calibration scores, which are absolute, information

scores depend on the reference range used, and are therefore not comparable among different studies. Higher information scores indicate a narrower interval between the 5<sup>th</sup> and 95<sup>th</sup> percentiles (i.e., greater certainty), but they say nothing about the accuracy of that interval (i.e., how likely it is to encompass the ‘true’ value). Further details of how information scores are calculated are provided in Appendix S1. For each expert, we calculated information scores for each question. Information scores were then averaged across all questions for each expert.

For each expert, the product of the calibration and information scores represented that expert’s overall score. The overall scores of the experts were then combined to produce a performance-based combination according to the weighting scheme described in Appendix S1. Briefly, the weights in the performance-based combination are proportional to the experts’ overall scores, and non-zero weights were only assigned to those experts with calibration scores greater than a specified threshold (e.g.,  $\alpha = 0.05$ ). The actual value of  $\alpha$  was set by optimization; we found the value of  $\alpha$  that maximized the overall score of the resulting combination. See Appendix S1 for additional details. In addition to the performance-based combination, we also derived an equal-weights combination of expert estimates, in which all expert assessments were weighted equally (Wittmann et al., 2015). Elicitation results were analyzed using the freely available Excalibur software (v1.5.7; <http://www.lighttwist.net/wp/excalibur>).

To assess the robustness of the performance-weighted combination relative to the equal-weights combination of expert estimates, we conducted an out-of-sample cross validation of the calibration variables. For all possible subsets of calibration variables, we calculated the performance-based combination and the equal-weights combination as described above and compared how these two decision-making methods scored (in units of calibration and in units of information) on the excluded set of calibration variables (Appendix S1).

Among SEJ studies, it is rather common that only a few experts are able to give informative and statistically accurate uncertainty assessments. In roughly 30% of cases, the performance-weighted combination assigns a single expert a weight of 1 (and all others a weight of 0; Cooke and Goossens, 2008). Those unfamiliar with SEJ may find such ‘lack of consensus’ troubling; however, the primary objective of SEJ is not consensus. Rather, it is to achieve the most informative and statistically accurate estimates of uncertainty that are possible given the set of experts participating in the elicitation.

To assess the sensitivity of the performance-based combination to the exclusion of certain experts and calibration variables, we conducted a robustness analysis. We removed one expert at a time, and recalculated both the performance-weighted and equal-weights combinations for the remaining experts (Appendix S4). We performed the same calculations without the top two performing experts and also for the cases where one calibration variable was removed at a time (Appendix S4).

### *Environmental factors and BMP efficiency*

The design of the SEJ protocol included scenarios and questions to elicit information on variability in the influent and effluent N loads for: different types of BMPs in the Piedmont and Coastal Plain physiographic provinces, rain events that varied in intensity, duration and timing, and rain events that had been preceded by prior rain (Appendix S3). We included variation in precipitation amounts because that is the primary factor within current stormwater regulations that determine BMP designs (MDE, 2009). We calculated removal efficiencies using the performance-based combination of influent and effluent TN loads for each point in the watershed:

$$\text{removal efficiency} = \frac{(\text{influent N} - \text{effluent N})}{\text{influent N}} \times 100 \quad (1)$$

Because the drivers of influent and effluent TN loads are likely to be strongly correlated, we assumed comonotonicity in combining 5<sup>th</sup>, 50<sup>th</sup>, 95<sup>th</sup> percentile estimates to calculate removal efficiencies (Dhaene et al., 2002). Thus, the three elements of each vector of quantiles were treated as completely dependent and were combined with corresponding vectors in an additive manner. For example, the 5<sup>th</sup> percentile for removal efficiency was calculated from only the 5<sup>th</sup> percentiles of influent and effluent TN loads using equation 1. We then compared the elicited removal efficiencies among the scenarios.

### *Climate change and BMP efficiency*

Any robust assessment of the ability of stormwater BMPs to remove N must account for hydrological conditions anticipated with climate change (Liu et al., 2014). To illustrate how managers could use expert-derived estimates of uncertainty in N removal by stormwater BMPs, we evaluated the relationship between storm size and uncertainty in N retention in the context of climate change. Most climate projections for the mid-Atlantic region encompassing the two watersheds in the elicitation predict increases in the amount and intensity of heavy rain events (Min et al., 2011; Moglen and Rios Vidal, 2014; Najjar et al., 2010; Walsh et al., 2014). To evaluate the effect of those predictions on BMP performance, we used a Monte Carlo approach to combine downscaled estimates of precipitation for the two watersheds with distributions of TN loads for

different-sized rain events from the SEJ results. These calculations estimated contemporary (1990–2019) and future (2050–2079) average annual TN loads imported to and exported from each BMP and watershed in the elicitation.

For each of the 60 questions on TN influent and effluent loads in the SEJ protocol, we defined the full distribution of predicted TN loads by fitting a lognormal distribution to the performance-based estimates of the 5%, 50%, and 95% percentiles. This process yielded best-fit parameters (mean and standard deviation) describing the expected distribution of influent and effluent TN loads for each unique BMP ( $n=5$ ) and watershed ( $n=2$ ) under specific levels of precipitation. We categorized the set of distributions for each site into three levels of precipitation: 0–19 mm (0–0.75 in.), 19–102 mm (0.75–4 in.), and 102+ mm (4+ in.). All BMPs and watersheds had at least one distribution of TN influent and effluent load associated with each precipitation level, and several had TN load distributions from multiple storms within a precipitation level. For sites with >1 distribution associated with a precipitation level, we ran Monte Carlo simulations (described below) for each distribution separately, and then calculated the mean.

We defined the magnitude of precipitation events according to the total rainfall depth on each day. Consecutive days with non-zero precipitation were combined into a single event by summing the rainfall depths for each day. The minimum rainfall depth needed to cause runoff was estimated at 1mm for the region encompassing both watersheds (Loperfido et al., 2014). Thus, we assumed the TN load was zero for any rain events  $\leq 1$ mm.

We obtained downscaled daily cumulative precipitation estimates for individual long-term weather stations for the years 1950–2100 from 9 Global Climate Models (CMIP5: CCSM4, CNRM-CM5, CSIRO-Mk3-6-0, HadGEM2-CC, INMCM4, IPSL-CM5A-LR, MIROC5, MPI-ESM-LR, MRI-CGCM3) for the higher and lower (RCP 8.5 and 4.5) future representative concentration pathways as used in the National Climate Assessment (based on the downscaling methodology described by Stoner et al., 2013). Downscaled data from NOAA station USC00181032 was applied for sites in the Piedmont watershed; data from NOAA station USW00093721 was applied for sites in the Coastal Plain watershed.

For each site, we randomly drew an annual set of precipitation estimates from the 30-year rainfall distribution for the corresponding watershed. Each rain event in the randomly drawn set was then matched with the SEJ-predicted distributions of TN influent and effluent loads for that site at the appropriate level of precipitation (0–0.75 in., 0.75–4 in., 4+ in.). We then randomly drew predicted influent and effluent TN loads from the influent and effluent distributions for that precipitation level. Lastly, we used the exact magnitude of the rain event associated with the SEJ-predicted distributions to linearly scale the predicted TN loads to the depth of the randomly sampled rain event in order to account for differences in water, and therefore nitrogen, delivery to watersheds and BMPs from different sized storms.

To generate an annual estimate of TN influent and effluent load for a site, we summed the resulting TN loads for all rain events in the randomly selected year. We repeated the entire process 1,000 times to generate distributions of annual TN loads expected for each site, climate model, and representative concentration pathway for the contemporary (1990–2019) and the future (2050–2079) time periods.

## Results

### *Expert rationales*

Experts used a variety of methods and reasoning to generate their estimates of N loads (Appendix S5). These included using statistical and process models to calculate mass balances. Each expert identified factors that they believed influenced N loads and each took various approaches to quantify the influence of those factors. The most common approach was to estimate event mean concentrations (EMCs) for each individual rain event using a set of assumptions along with a combination of data provided in the protocol document (Appendix S3), additional published data, and unpublished datasets from their own research (8 experts). Those experts then used a variety of techniques to calculate 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles of expected TN loads from the estimated EMCs. Two experts parameterized detailed models of hydrology and N fluxes for each watershed, using information from the protocol document along with assumptions from existing urban stormwater models to derive statistical distributions of TN loads at various points in the watersheds. Notably, the top two performing experts relied on substantially different approaches to arrive at their uncertainty estimates. One constructed a detailed hydrologic model for each suburban watershed. The other relied almost exclusively on experience and intuition to adjust his basic load calculations by applying approximate correction factors for the effects of different environmental variables.

### *Expert elicitation*

The experts' calibration scores ranged from  $3 \times 10^{-14}$  to 0.706. Only two of ten experts scored higher than 0.05 (Table 1, Appendix S4). Higher numbers indicate more probabilistically accurate information was provided as measured against the 11 calibration variables. The performance-weighted combination of expert estimates

Table 1. Calibration and information scores for expert assessments, excluding expert #10<sup>a</sup>

Expert ID	Calibration score	Information score		Normalized weight
		All variables	Calibration variables	
1	0.492	1.61	1.58	0
2	7.54×10 <sup>-12</sup>	4.62	3.81	0
3	2.50×10 <sup>-6</sup>	1.47	2.18	0
4	7.99×10 <sup>-4</sup>	0.900	0.574	0
5	3.21×10 <sup>-14</sup>	4.13	3.45	0
6	2.79×10 <sup>-5</sup>	2.75	2.74	0
7	7.99×10 <sup>-4</sup>	2.59	2.40	0
8	0.706	1.90	1.99	1
9	0.001	0.318	0.461	0
EW	0.068	0.750	0.804	
PW	0.706	1.90	1.99	

<sup>a</sup>EW represents the combined estimate of all expert assessments weighting each equally. PW indicates the combined estimate of experts' assessments using the performance-weighted criterion. Larger calibration scores indicate a greater probability that an expert's estimates encompassed the 'true' values of the calibration variables. Larger information scores indicate a narrower interval between the expert's 5<sup>th</sup> and 95<sup>th</sup> percentiles (i.e., greater certainty), regardless of whether that interval actually encompassed the 'true' values of the calibration variables. See the text and Appendix S1 for further explanation of how the weights, calibration scores, and information scores were calculated.

doi: 10.12952/journal.elementa.000063.t001

included only one expert and had a higher calibration score (0.706) than the equal-weights combination (0.068). The scores of both schemes were sufficiently high that we did not reject the hypothesis that their probability assessments were accurate (Table 1). Performance-based weighting was more than twice as informative as the equal-weights combination, and since informativeness grows slowly with the spread of the 5–95% confidence bands, this difference in informativeness was substantial. The substantially higher calibration and information scores of the performance-weighted combination led us to focus our interpretation and subsequent analyses on the performance-weighted combination of expert estimates.

One of the experts (#10) provided estimates that consistently extended several orders of magnitude above the estimates of all other experts, across all variables (Figure 3). The extremely large ranges associated with this expert's assessment defined the reference range used in computing information scores. Including this expert did not affect the resulting combination, but it led to very large ranges outside the 90% confidence bands. These were judged unrealistic by the study team (and indeed by the other experts – as compared to their estimates). The assessment of tail probabilities outside the 90% central confidence region is not part of the expert elicitation but reflects a modeling decision by our study team. We therefore decided to simply exclude this expert's influence on the reference range. Exclusion of this expert did not alter the normalized weightings for the performance-weighted combination of expert estimates, and the performance-based weighting scheme outperformed the equal-weights combination regardless of whether this expert was included. We also report the calibration and information scores with this expert included (Appendix S4); however, all analyses are based on the performance-weighted combination with expert #10 excluded (Table 1).

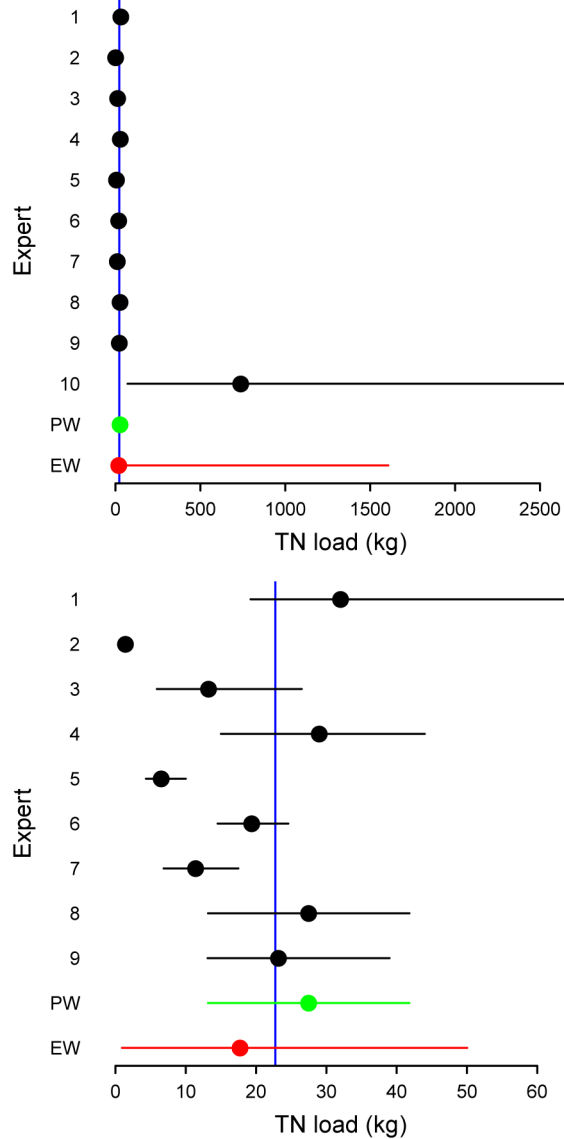
Results of the out-of-sample cross validation analysis revealed that the performance-weighted combination strongly outperformed the equal-weights combination in predicting values of calibration variables which were not used in computing the performance weights (Appendix S1).

The robustness analysis showed that even when the top-performing expert was removed, the resulting performance-weighted combination still outperformed the equal-weights combination (Appendix S4). This result also held when the top two performing experts were removed from the analysis (Appendix S4).

### *Expert assessments of BMP efficiency*

The performance-weighted expert estimates suggest that, even when detailed information on BMP design, sizing, placement on the landscape, and precipitation conditions is available, there is substantial uncertainty in BMP performance (Appendix S6). Furthermore, although median estimates of N removal varied by BMP type, the experts believed the magnitude of variation would not change substantially with environmental factors such as geographic province, season, or location (Figure 4). For several BMP types, the performance-based estimates of TN removal efficiencies extended into negative values indicating the possibility of those structures serving as net sources of N during a given rain event.

For all BMP types, experts agreed that the magnitude of the rain event (e.g., total event precipitation) was the most important factor driving TN removal and variability in TN removal (Figure 4, Appendix S5, Appendix S7). Expert responses indicated that the relationship between watershed-level TN export and



**Figure 3**

Raw results from the SEJ for question #4 (a calibration variable) in the protocol document.

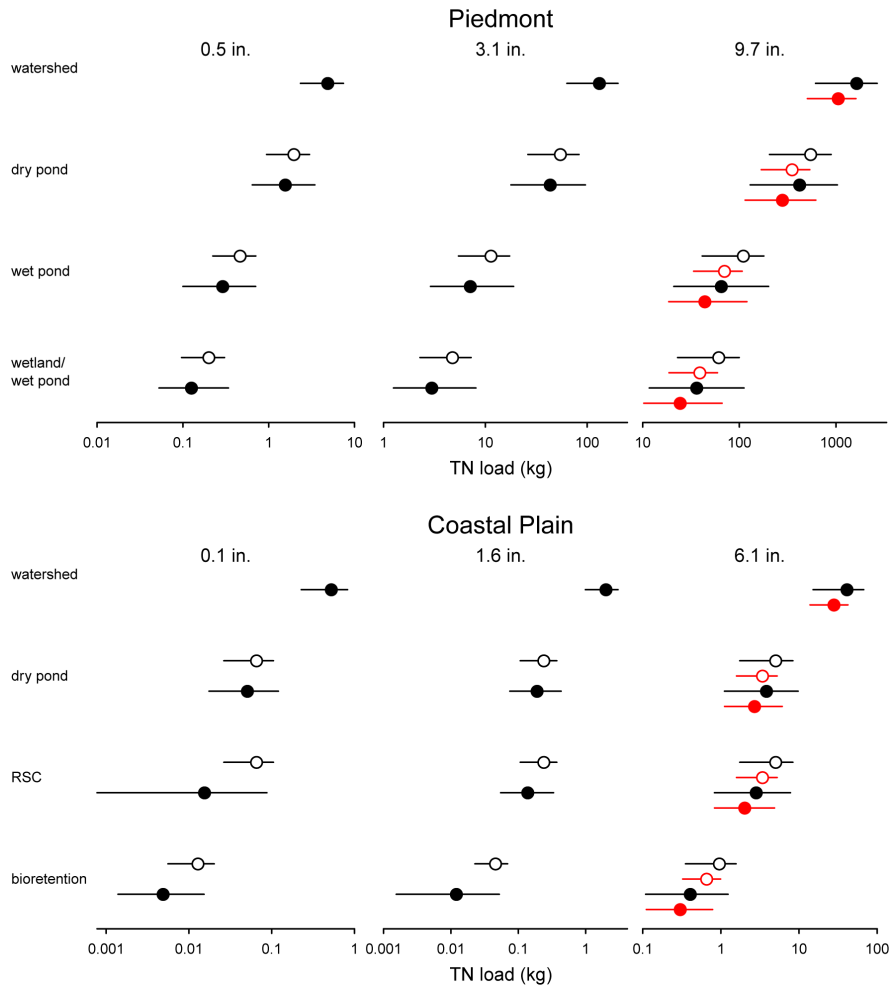
The question asked experts to estimate the 50<sup>th</sup> (black dot), 5<sup>th</sup>, and 95<sup>th</sup> (black line) percentiles of the expected outgoing total nitrogen load (kg TN) from the Piedmont watershed over the entire duration of a 1.1-in. rain event. All 60 questions were of a similar format (Appendix S3). **Top:** Estimates given by expert 10 were consistently inaccurate and uninformative relative to the calibration variables. **Bottom:** Removing expert 10 from the analysis reduced the range against which information scores were calculated and improved the equal-weights decision maker (EW; red dot and line); however, the performance-weighted decision maker (PW; green dot and line) was unchanged. The performance-weighted decision maker (PW) represents the combined estimate of experts' assessments using the performance-weighted criterion. The equal-weights decision maker (EW) represents the combined estimate of experts' assessments weighting each expert equally.

doi: 10.12952/journal.elementa.000063.f003

precipitation varied with geographic province (Figure 5), with smaller estimated TN loads for the Coastal Plain watershed. Assuming all other factors are equal (e.g., infiltration), the volume of water flowing through a BMP is determined by two component variables: (1) total precipitation in the rain event, and (2) drainage area to the BMP. The elicitation showed that experts believed BMPs with smaller drainage areas tend to have greater TN removal (Figure 4) and BMPs tend to function more efficiently for small rain events (Figure 5). Despite this, estimates of the variability of retention rates were relatively invariant across precipitation amounts and drainage areas for dry ponds, wet ponds, and wet ponds/wetlands (Figure 6). In contrast, variability declined with storm size for bioretention cells and regenerative stormwater conveyance structures (Figure 6); experts were more certain of the reduced performance of those BMP types during high flow events resulting from large storms.

Expert estimates suggested that antecedent conditions had a modest effect on TN removal efficiencies. Estimated BMP removal efficiencies were slightly lower for very large rain events (>6 in. total precipitation), when preceded by a similarly large rain event (Figure 4, Figure 6, Figure 7). According to the combined performance-weighted SEJ estimates, large storms flushed N from the watershed, and those effects persisted for at least a month. Thus, expert knowledge indicated that similar storm events during that time exported less N than would otherwise be expected because less N was available in the watershed. Although this phenomenon is apparent from the combined performance-weighted data and was articulated in the rationales provided by some experts (Appendix S5), it is worth noting that the experts disagreed on how long the flushing effects of large storms would persist. Some experts contended that N flushing by large storms would have no effect on N loads in subsequent storms, regardless of the time interval (Appendix S5).





**Figure 4**  
Performance-based expert estimates of TN loads flowing into and out of BMPs for different storm sizes in the two watersheds.

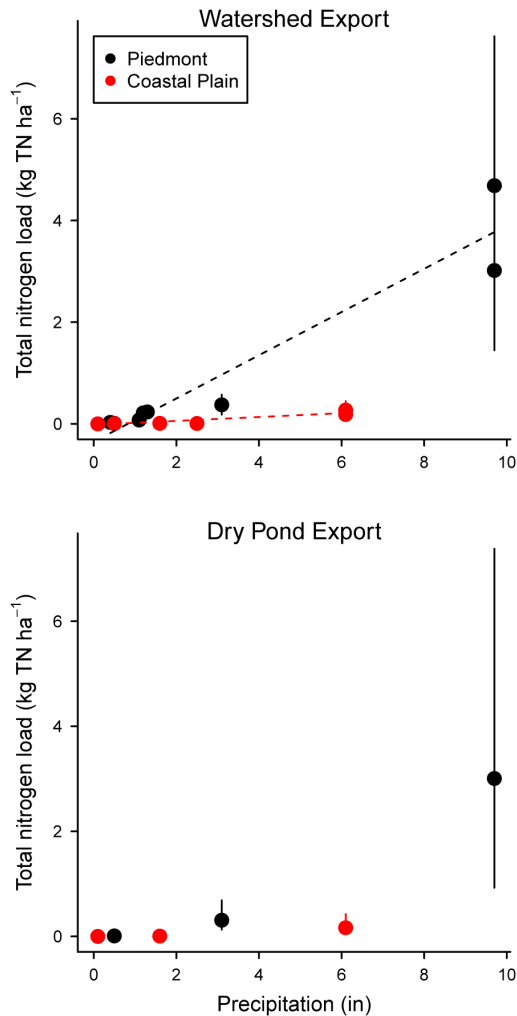
**Top:** Piedmont watershed; **Bottom:** Coastal Plain watershed. Total precipitation for each storm is given above each plot. Open symbols represent influent TN loads; closed symbols represent effluent TN loads. Red symbols are the estimates for a rain event preceded by a storm of identical magnitude one month earlier. RSC: regenerative stormwater conveyance structure.  
doi: 110.12952/journal.elementa.000063.f004

Estimates of TN removal by dry ponds (the only BMP type common in both the Piedmont and Coastal Plain watersheds) showed little effect of geographic province on TN removal efficiency (Figure 5). When scaled by contributing watershed area, performance-weighted expert estimates of TN loads for the Coastal Plain were low, compared to those for the Piedmont watershed, for similarly sized precipitation events (Figure 5, Appendix S7).

Performance-weighted expert estimates indicated that seasonal effects were not strong enough to substantially influence estimated TN loads and removal efficiencies for any of the BMP types. However, many experts also highlighted the lack of field data on seasonal performance of BMPs (Appendix S5). For the Coastal Plain watershed, we asked experts to assess the effect of retrofitting existing BMPs to more recent designs such as bioretention cells and regenerative stormwater conveyance structures. Across all storm sizes, experts estimated that retrofits reduced TN export by 37% (Figure 7).

### *Estimated BMP efficiency with climate change*

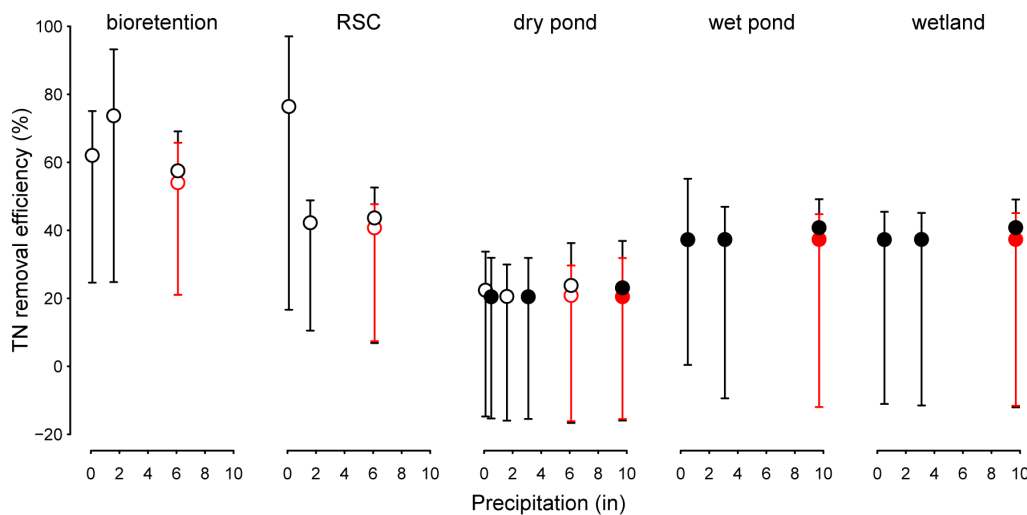
Both historical trends as well as high-resolution future projections for individual weather stations for daily rainfall for the Piedmont and Coastal Plain watersheds indicated a shift toward more heavy precipitation events during 2050–2079 compared to 1990–2019, with greater changes under the higher (RCP 8.5) as compared to the lower (RCP 4.5) scenario (Appendix S8). Under the high scenario, median rain event size averaged across all nine climate models for each 30-year period increased by 0.8mm for the Coastal Plain watershed and by 1.3mm for the Piedmont watershed (Figure 8). Expert estimates showed a clear pattern of TN removal declining with increasing storm size for nearly all BMP types (Appendix S7). Using a Monte Carlo resampling procedure in combination with the performance-based distributions of TN export for different-sized storms, we estimated expected annual TN removal for the BMPs and the entire watersheds under both contemporary and future scenarios of daily rainfall estimates. The resulting estimates of annual TN loads show a pattern of increasing TN export under the projected future precipitation regime, with median estimates of projected TN loads rising by 15–31% across all BMPs and both watersheds for the high scenario



**Figure 5**  
Performance-based expert estimates of watershed- and BMP-level TN export varied with storm size.

**Top:** Expert estimates of drainage area-weighted TN loads exported from the Piedmont and Coastal Plain watersheds increased with precipitation; however, TN export from the Coastal Plain was lower than Piedmont TN export for a given storm size (ANCOVA:  $p < 0.001$ ). **Bottom:** Expert estimates of drainage area-weighted TN loads exported from dry ponds increased with precipitation similarly in the Coastal Plain and Piedmont watersheds (ANCOVA:  $p > 0.05$ ).  
doi: 10.12952/journal.elementa.000063.f005

(Figure 8). For the Piedmont watershed (3.5 km<sup>2</sup>), this projected climate-driven increase in N transport translated to an additional 580 kg of N exported each year (Figure 8). Projected N export increased by 15 kg N for the Coastal Plain watershed (1.5 km<sup>2</sup>; Figure 8). Furthermore, variation in predicted TN loads for the 2050–2079 period was as large or larger than that for 1990–2019, reinforcing the need for better understanding of the controls on BMP performance under climate change.



**Figure 6**  
TN removal efficiencies calculated from the performance-based estimate of incoming and outgoing TN loads for each BMP varied widely with precipitation for all BMP types.

Expert estimates of TN removal efficiencies for RSCs and bio-retention cells declined with increasing storm size. For all BMP types, uncertainty in TN removal was high for all levels of precipitation. Points are median estimates and error bars encompass the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the combined performance-based estimate. Open symbols represent BMPs in the Coastal Plain watershed; filled symbols represent BMPs located in the Piedmont watershed. Red symbols indicate TN removal efficiency for rain events preceded by a storm of identical magnitude one month earlier. Positive TN removal efficiencies represent net retention of TN by the BMP during the rain event, and negative TN removal efficiencies indicate that the BMP acted as a net source of TN during the rain event.  
doi: 10.12952/journal.elementa.000063.f006

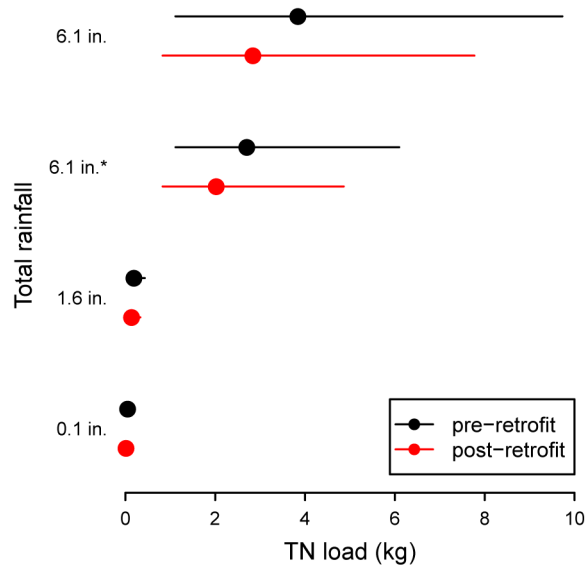


Figure 7

Expert-estimated effluent TN loads and variability were marginally higher for a dry pond before it was retrofitted to a regenerative stormwater conveyance structure (RSC) in the Coastal Plain watershed.

This effect was greatest for large storms. Asterisk (\*) denotes a rain event preceded by a storm of identical magnitude (6.1 in.) one month earlier.

doi: 10.12952/journal.elementa.000063.f007

## Discussion

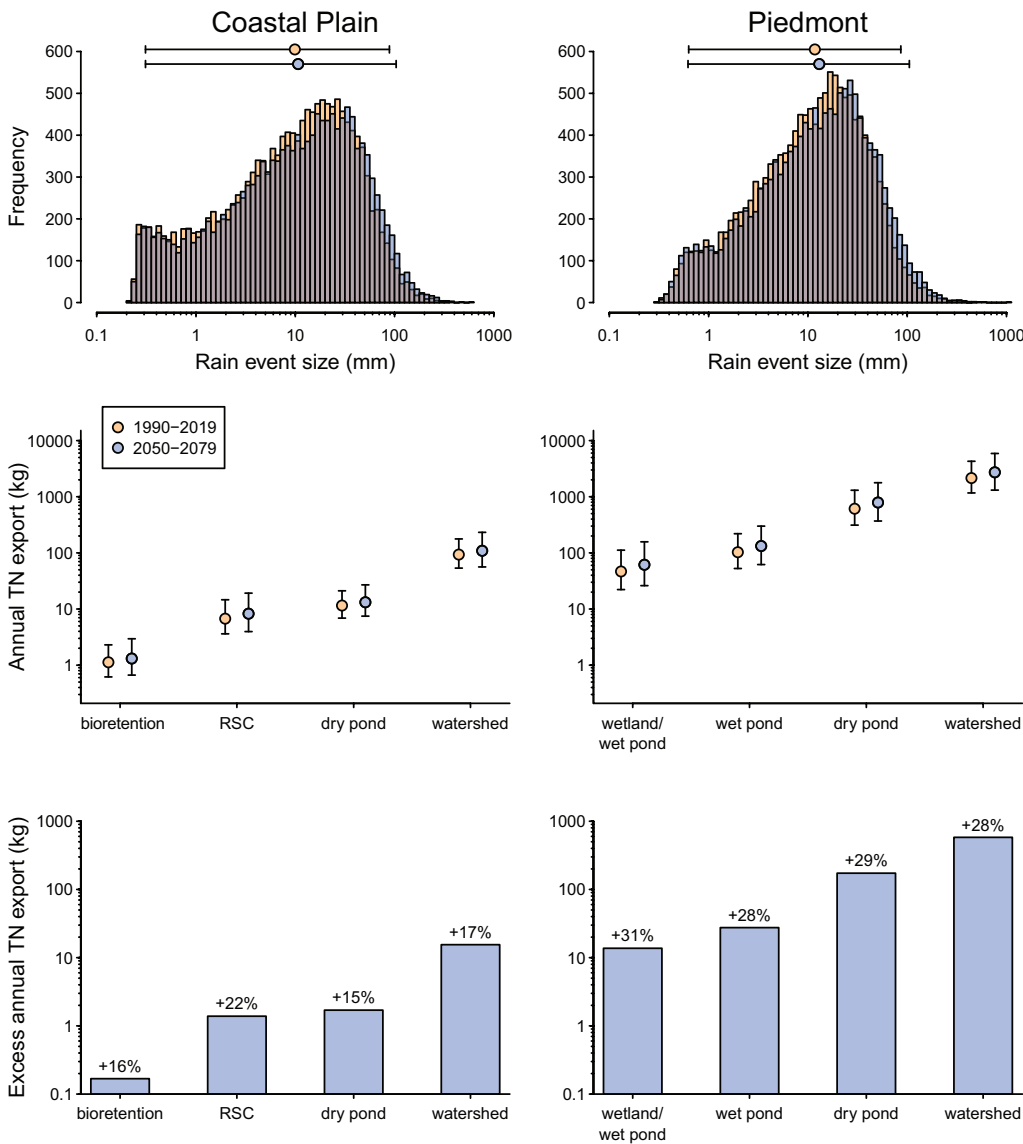
Stormwater BMPs are expected to perform under a wide variety of environmental conditions, yet the extent to which BMPs vary in their capacity to retain excess N and the drivers of that variability are largely unknown. Here, we have quantified the uncertainty in BMP performance under a range of site-specific environmental conditions and we have identified environmental factors that influence variation in BMP performance. Contrary to our hypotheses, expert knowledge suggests that physiographic province, BMP type, seasonality, and antecedent precipitation had little influence on overall retention. Experts believed that the amount of rain was the primary identifiable source of variability in BMP efficiency.

### *Environmental drivers of BMP performance*

There are good reasons to expect that rainfall amount, and hence the volume of water moving through BMPs will be an important driver of N retention across sites (Collins et al., 2010; Newcomer Johnson et al., 2014). Most stormwater detention structures are designed to provide water quality treatment for smaller storm events (e.g., 0.9–1.0 in. in Maryland; MDE, 2009). For rain events that exceed design specifications, water and nutrient fluxes should increase rapidly, though not necessarily at the same rate. Many of the N removal mechanisms for stormwater BMPs (e.g., biotic uptake, denitrification, absorption) depend on water residence time, which controls the duration of contact with transformation or absorption loci (Bettez and Groffman, 2012; Collins et al., 2010; Passeport et al., 2013). As the flux of water and nutrients through a BMP increases, residence time declines, leading to a precipitous drop in nutrient retention efficiency (Carleton et al., 2001). The SEJ results were in line with these findings; the water holding capacity of a BMP was judged by the experts to be an important factor determining retention efficiency at higher flows. Nonetheless, the expert knowledge elicited in the present study indicated that even after precipitation exceeded a 2-in. storm, BMPs still retained a median estimate upwards of 40% TN, and continued to do so with triple the rainfall. Although elevated rainfall intensity may lower removal efficiency estimates, expert knowledge suggested that it may not do so consistently.

The experts judged the BMPs to perform similarly in the Coastal Plain and Piedmont despite differences in stratigraphic properties and the potential for distinct mechanisms of runoff between the provinces. Watershed budgets indicate little difference in precipitation or runoff among watersheds in the Piedmont or Coastal Plain (Correll et al., 1999; Dougherty et al., 2007; Jordan et al., 2003). However, geologic and stratigraphic differences suggest greater runoff capacity from infiltration excess overland flow in the Piedmont (Markewich et al., 1990; Swain et al., 2004) and greater capacity for saturation excess overland flow in much of the Coastal Plain (Ator et al., 2005b; Markewich et al., 1990). Very high runoff yield can occur when saturation is achieved. Utz et al. (2011) found that Coastal Plain streams with low impervious cover (i.e., up to 20%) were generally more stable (i.e., less frequent high flows, longer high flow duration) than Piedmont streams with a similar degree of urbanization. This pattern was supported by the data provided to experts in this study, which indicated higher water retention capacity in the Coastal Plain watershed compared to the Piedmont watershed.

Nonetheless, the boundaries of physiographic provinces may be too coarse to make generalizations about BMPs for management. Both the Coastal Plain and the Piedmont provinces have multiple sub-regional



**Figure 8**

Projected increases in heavy precipitation events under climate change translate to greater export of TN from individual BMPs and from entire watersheds in the Coastal Plain and Piedmont physiographic provinces.

**Top:** Downscaled daily cumulative precipitation estimates from nine CMIP5 climate models (CCSM4, CNRM-CM5, CSIRO-Mk3-6-0, HadGEM2-CC, INMCM4, IPSL-CM5A-LR, MIROC5, MPI-ESM-LR, MRI-CGCM3) under RCP 8.5 for long-term NOAA weather stations USW00093721 (Coastal Plain) and USC00181032 (Piedmont) for the periods 1990–2019 (orange) and 2050–2079 (blue). Precipitation occurring on consecutive days was combined into a single event. Rain event sizes for all nine models are included in the histograms. **Middle:** Bootstrapped median (and 95% CI) annual TN export across all climate models for the individual BMPs and watersheds in the Coastal Plain and Piedmont provinces under contemporary and future climatic conditions (RCP 8.5). Expert-derived estimates of TN export for a range of storm sizes along with projected rain event sizes (top panel) were used as the basis for Monte Carlo simulations of annual export (see text for modeling details). The Coastal Plain watershed (1.5 km<sup>2</sup>) was smaller than the Piedmont watershed (3.5 km<sup>2</sup>). **Bottom:** Median projected increase in annual TN load from the 1990–2019 period to the 2050–2079 period for each BMP and watershed. Percentages indicate the percent increases in annual TN loads over 1990–2019 values.

doi: 10.12952/journal.elementa.000063.f008

hydrogeomorphic properties that affect the impacts of urbanization on N loads. For example, the Coastal Plain province varies geographically in the degree of subsurface hydrologic storage (Ator et al., 2005a), which may lead to localized differences in the capacity to retain N. Similar hydrogeomorphic variation occurs within the Piedmont province, where base flow indices can vary substantially by watershed (Schwartz and Smith, 2014). Such differences may attenuate or enhance the performance of certain BMP designs. For example, regenerative stormwater conveyance structures would not be expected to perform well in low-lying regions with highly saturated soils but may be more effective when located in upland areas with greater hydrologic storage. Thus, it may be necessary for water resource managers to first establish the particular hydrogeomorphic properties of a specific site to enable maximizing the effectiveness of BMPs implemented there.

At the watershed scale, the experts judged the Coastal Plain watershed to have lower N exports than the Piedmont watershed. This perception is consistent with observations of lower Coastal Plain N yields when compared to Piedmont watersheds in rural contexts (Jordan et al., 1997; Liu et al., 2000). The pattern of smaller estimated TN loads for the Coastal Plain watershed may be driven by a greater capacity for subsurface flow in some areas of this province as well as the prevalence of streamside riparian forests (Weller et al., 2011; Weller and Baker, 2014). Indeed, Utz et al. (2011) found that TN yield from mid-Atlantic Coastal Plain watersheds was significantly lower than Piedmont watersheds across the same region. The extent to which the N contained in subsurface flows is retained before reaching coastal waters is variable (Harden and Spruill, 2008), and subsurface N export from some Coastal Plain sites may be substantial (Hubbard and Sheridan, 1983). Thus, while the experts' estimates suggest that the Coastal Plain watershed in this study

may be more retentive of N, unmeasured subsurface flow paths may mean that actual TN export is as great as or even greater than that of the Piedmont watershed.

Indeed, subsurface flow paths may be such that N-rich groundwater can bypass biogeochemically active zones in some Coastal Plain areas (Ator et al., 2005a). In addition, the low relief of the Coastal Plain may also mean that discharge points can occur in neighboring basins, so TN export is realized in a different stream and is spatially disconnected from the source area. In those cases, apparent water quality benefits may be realized if within-basin groundwater flows are attenuated to such a degree that they never contribute a sizeable fraction of observed discharge (because other, cleaner water is arriving faster in greater amounts). Such effects would be difficult to capture empirically in the local watershed-monitoring scheme typically used.

One of the most informative findings of our elicitation was that detailed information on BMP design, sizing, placement in the watershed, additional stormwater infrastructure, land use/land cover, precipitation timing and amount was insufficient for experts to provide narrow ranges in expected BMP N removal efficiency or watershed TN export. Detailed site-specific information could not make up for the lack of data and understanding of the degree to which these specific environmental factors influence BMP performance. At the watershed scale, expert-derived estimates of TN export varied by a factor of three or more under all conditions (Figure 4), and BMP removal efficiencies spanned broad ranges and in some cases were negative (Figure 6), indicating that a BMP could act as a net source of N during a particular storm event. Nevertheless, expert estimates of removal efficiencies were all fairly optimistic in that many experts assumed there was no environmental downside to specific BMP structures, despite some empirical evidence to the contrary (e.g., regenerative stormwater conveyance structures, Palmer et al., 2014). Such optimism may have parallels in the history of rural riparian buffer management, where the water quality management expectations of buffers soared (CBP, 2014) even after the context dependence of many studies was emphasized (e.g., Lowrance et al., 1997).

### *The utility of measuring uncertainty in BMP performance*

Wide uncertainty in BMP performance, despite the availability of detailed site-specific environmental data, suggests that improving the mismatch between expected and actual BMP performance will require more than collecting additional site-specific data. Specifically, sound management of water quality requires accounting for the estimated magnitudes of uncertainty when setting the nutrient removal credits for particular BMPs implemented in particular locations. The fallacy perpetuated in many local and regional water quality regulations that a certain type of BMP will remove a fixed proportion of influent N (PLRM Development Team, 2009; USEPA, 2010a) should instead be updated to follow a probabilistic perspective on nutrient loading. Regulations also do not account for site-specific factors, assuming consistent performance of all similarly designed BMPs, despite evidence to the contrary. More empirical research is needed to fully understand the context-dependencies of BMP performance. In the meantime, a quantitative approach, like the one taken here, for estimating the BMP-scale and watershed-scale variability in N loads and N retention has the capability to transform current model-based, single-value predictions of annual N export to coastal waters into probability distributions of N export that capture the uncertainty of watershed- and BMP-level processes.

For example, SEJ-derived estimates of uncertainty would prove useful to regulators in two contexts. First, they would help regulators understand the regional nutrient loading implications of widespread BMP implementation under proposed policy change scenarios (e.g., the Phase 5 Chesapeake Bay model; USEPA, 2010a). Second, in cases where a particular sub-watershed is targeted for reductions in nutrient runoff to meet a specified total maximum daily load (TMDL), uncertainty estimates would help regulators determine if implementing certain BMPs is sufficient to meet those reduction goals. Characterizing the uncertainty in BMP performance with a distribution enables officials to assess the risks associated with “best case” and “worst case” scenarios that are not well represented by single numbers like proportional reductions to event-mean concentrations.

Similarly, equipped with a known level of uncertainty in BMP performance, local officials can make more informed decisions about how to spend limited funds for reducing nutrient exports. For example, officials may choose to invest more heavily in proven practices such as reducing impervious cover (Collins et al., 2010; Craig et al., 2008) or implementing agricultural BMPs (Lowrance et al., 1997; Woltemade, 2000). More complete knowledge of the variation in BMP performance would also enable water resource managers to more accurately gauge the catchment-wide effects of multiple individual BMPs within the watershed.

### *The role of empirical studies of BMP performance*

Empirical data on stormwater BMP performance are only available for a limited number of sites under a limited number of environmental conditions (Barrett, 2008). A critical question for water resource managers and regulators is how well those limited empirical measurements represent the performance of the many thousands of BMPs currently in operation. We recently conducted a meta-analysis of empirical estimates of stormwater BMP N removal efficiencies (Koch et al., 2014) and can compare those results with N removal efficiencies for a subset of BMP types derived from the expert elicitation (Table 2). The results of this

**Table 2. Median TN removal efficiencies for different BMP types from this study and from a recent meta-analysis of BMP performance<sup>a</sup>**

BMP type	This study	Koch et al., 2014
bioretention structure	62%	–
RSC <sup>b</sup>	51%	–
dry pond	22%	27%
wet pond	38%	40%
wetland	38%	61%

<sup>a</sup>All removal efficiencies are medians

<sup>b</sup>RSC: regenerative stormwater conveyance structure

doi:10.12952/journal.elementa.000063.r002

meta-analysis were made available to all experts in the elicitation and many experts used them as a basis for estimating N retention in their calculations (Appendix S5). In all cases the expert-derived estimates of N removal efficiency were lower than those summarized in the meta-analysis of empirical results. Because removal efficiencies in the meta-analysis made up only one of many inputs to the experts' calculations, this pattern suggests that published empirical estimates of BMP performance may collectively overestimate the ability of BMPs to retain N. Similar to the manner in which newly collected data are used to update prior knowledge in a Bayesian analysis (Bolker, 2008), the experts' estimates had the effect of tempering existing field-based knowledge of BMP performance, and indicated that BMPs may not perform as well as published summaries of field measurements indicate (e.g., Barrett, 2008; Simpson and Weammert, 2009; Winer, 2000). One possible explanation for this mismatch is that BMPs that perform well may tend to be studied more often (CBP, 2013) and results of highly retentive BMPs may tend to be published over those from poorly performing BMPs (CSN, 2012; Koch et al., 2014). Likewise, due to rarity and logistical difficulty, data collection is often hampered during extreme precipitation events when the bulk of N export is hypothesized to occur. Therefore, these high-precipitation events may be under-represented in empirical studies.

Although SEJ provides a means to robustly estimate uncertainty of BMP performance in the absence of more comprehensive empirical measurements, it cannot fully replace well-conceived, long-term field studies of BMP performance over a range of environmental conditions. Such studies are instrumental to understanding the substantial variation in N retention by stormwater BMPs (Strecker et al., 2001). All experts in this study expressed the view that there are many unknowns in calculating N loads and that there is a crucial need for more field data. They indicated a number of environmental factors that may influence BMP performance, but which lacked enough empirical data to factor into their calculations. Those environmental factors included: soil type, patterns of fertilizer use, seasonality, antecedent soil conditions, sediment transport, and vegetation effects on N retention. Future studies are needed to assess the influence of these environmental factors on BMP performance. Specifically, decision makers need long-term measurements of influent and effluent nutrient loads for a range of BMP types under a variety of site-specific and climatic conditions. The cost of such monitoring is small compared to the cost of BMP implementation and those data should be collected routinely when BMPs are installed. Such research would enable stormwater regulators to provide more nutrient reduction credits for certain BMPs in certain locations, thereby encouraging smarter and more effective implementation of BMPs.

In the elicitation, the experts identified rainfall amount (i.e., total depth) as the most important factor influencing BMP performance. However, the SEJ was not designed to exhaustively parse the interactive effects of multiple storm characteristics on BMP N retention, and variables such as the duration and intensity of rainfall may also be important in controlling N export. Although stormwater BMPs are designed according to a standard rainfall depth rather than a standard storm intensity (MDE, 2009), rainfall intensity would also be expected to influence N retention. For example, a 2.5-in. rain event occurring evenly over a period of several days would be expected to export less N from a given BMP than a 2.5-in. rain event occurring over a few hours. Empirical studies are needed to quantify the extent to which factors such as rainfall intensity and the distribution of rain over time influence BMP performance, independent of storm size.

More efficient management of water quality has been realized in some rural areas and highlights the importance of pinpointing the factors controlling BMP performance. In agricultural settings, estimates of uncertainty in BMP performance have been combined with costs to optimize the distribution of BMP placement given a particular budget (Diebel et al., 2008; Maxted et al., 2009). However, as water managers come under increasing pressure to meet higher water quality targets, greater certainty in BMP performance is needed. The knowledge conferring greater certainty must be based ultimately on empirical data.

### *BMP performance under a changing climate*

This elicitation, and indeed most empirical studies, yielded estimates of TN loads and associated uncertainty at the scale of individual rain events. However, water quality regulators are ultimately concerned with the

annual variability in TN loads from a specific watershed. For example, the total maximum daily load (TMDL) of N for the Chesapeake Bay is based on annual estimates of N export from the watershed (USEPA, 2010b), and variation in actual yearly loads determine whether this target is met. In addition, annual estimates of variability are essential for designing stormwater infrastructure that will perform well under projected future climate. Our Monte Carlo-based model of TN loads using high-resolution daily estimates of precipitation for both contemporary (1990–2019) and future (2050–2079) climate scenarios provide estimates of TN export and BMP efficiency on an annual time scale.

Although some studies have assessed the impacts of projected precipitation changes on the effectiveness of urban BMPs to mitigate peak storm flows (Forsee and Ahmad, 2011; Rosenberg et al., 2010; Waters et al., 2003), few studies have examined the likely influence of a changing climate on the capacity for stormwater BMPs to maintain water quality. The model presented here used expert-derived uncertainties in TN loads to predict higher and more variable TN loads from the Coastal Plain and Piedmont watersheds under a projected future climate regime. These predictions suggest that current BMP regulations for reducing nutrients will likely be inadequate for limiting the export of excess N in the future. Similar studies have shown that stormwater BMPs may be insufficiently sized to reduce peak discharge from projected future heavy rain events (Forsee and Ahmad, 2011; Waters et al., 2003). It should be noted that our estimates of TN loads under projected future climate change are conservative in that they only account for the effects of increased precipitation. For example, they do not incorporate projected future increases in atmospheric N deposition (Galloway et al., 2004) nor do they include changes in land use/land cover such as fertilizer use (EMRPC, 2011).

Furthermore, the experts in our elicitation identified the volume of water flowing through a BMP as the primary factor driving TN loads. Because stream flows are determined by precipitation, historical trends and climate projections (Walsh et al., 2014) showing increases in heavy precipitation in the mid-Atlantic are essential to BMP planning. Assuming BMPs will continue to perform in the future as they have in the past ignores the influence of a changing climate (Rosenberg et al., 2010) which may, as our results suggest, be changing the very aspect of climate (precipitation) that most affects BMP N retention.

## Summary

Excess nitrogen is a primary cause of freshwater and coastal eutrophication globally (Conley et al., 2009) and urban stormwater is a rapidly growing source of N pollution (NRC, 2009). Stormwater BMPs are widely used throughout the United States to remove excess nitrogen from runoff in urban and suburban areas (NRC, 2009). Yet the gap between how much N is actually removed by stormwater BMPs and the amount that is regulated to be removed by those BMPs is difficult to quantify and hinders the ability of water resource managers to meet water quality targets. Estimating the uncertainty in BMP performance under a variety of site-specific environmental conditions helps address this problem by establishing likelihoods for different levels of N removal. The expert knowledge from this elicitation indicated substantial uncertainty in BMP performance, with N removal efficiencies ranging from <0% to >40%. Experts believed that the amount of rain was the primary driver of variability in BMP efficiency, and that geographic and seasonal effects were not strong enough to substantially influence estimated variability in BMP performance. When combined with projected increases in heavy precipitation, expert-derived estimates of uncertainty in BMP performance suggested a pattern of future increases in N export for suburban watersheds of the Chesapeake Bay region. SEJ provides a robust way to quantify the uncertainty in stormwater BMP performance and can also be useful in predicting N retention under projected future climate change. However, empirical knowledge of BMP performance is limited and additional field studies are needed to further define the extent to which site-specific conditions and environmental factors influence the N retention by stormwater BMPs.

## References

- Aspinall W. 2010. A route to more tractable expert advice. *Nature* 463:294–295. doi:10.1038/463294a.
- Aspinall WP, Cooke RM. 2013. Quantifying scientific uncertainty from expert judgment elicitation, in Rougier J, Sparks S, Hill LJ, eds., *Risk and Uncertainty Assessment for Natural Hazards*. Cambridge, UK: Cambridge University Press. pp. 64–99.
- Ator SW, Denver JM, Brayton MJ. 2005a. Hydrologic and geochemical controls on pesticide and nutrient transport to two streams on the Delmarva Peninsula. Reston, Virginia: U.S. Geological Survey. Scientific Investigations Report 2004-5051. Available at <http://pubs.usgs.gov/sir/2004/5051/>.
- Ator SW, Denver JM, Krantz DE, Newell WL, Martucci SK. 2005b. A surficial hydrogeologic framework for the mid-Atlantic coastal plain. Reston, Virginia: U.S. Geological Survey. Professional Paper 1680. Available at <http://pubs.usgs.gov/pp/2005/pp1680/>.
- Bamber JL, Aspinall WP. 2013. An expert judgment assessment of future sea level rise from the ice sheets. *Nat Clim Change* 3: 424–427. doi:10.1038/nclimate1778.
- Barrett ME. 2008. Comparison of BMP performance using the international BMP database. *J Irrig Drain Eng* 134: 556–561. doi:10.1061/(ASCE)0733-9437(2008)134:5(556).
- Bettez ND, Groffman PM. 2012. Denitrification potential in stormwater control structures and natural riparian zones in an urban landscape. *Environ Sci Technol* 46: 10909–10917. doi:10.1021/es301409z.

- Bolker BM. 2008. *Ecological Models and Data in R*. Princeton, New Jersey: Princeton University Press.
- Carleton JN, Grizzard TJ, Godrej AN, Post HE. 2001. Factors affecting the performance of stormwater treatment wetlands. *Water Res* 35: 1552–1562. doi:S0043-1354(00)00416-4.
- CBP. 2013. Estimates of county-level nitrogen and phosphorus data for use in modeling pollutant reduction: Documentation for scenario builder version 2.4. Chesapeake Bay Program. Available at [http://www.chesapeakebay.net/documents/SB\\_Documentation\\_V24\\_01\\_04\\_2013.pdf](http://www.chesapeakebay.net/documents/SB_Documentation_V24_01_04_2013.pdf).
- CBP. 2014. Chesapeake Bay Watershed Agreement. Chesapeake Bay Program. Available at <http://www.chesapeakebay.net/chesapeakebaywatershedagreement/page>.
- Chiang L-C, Chaubey I, Maringanti C, Huang T. 2014. Comparing the selection and placement of best management practices in improving water quality using a multiobjective optimization and targeting method. *Int J Environ Res Public Health* 11: 2992–3014. doi:10.3390/ijerph110302992.
- Collins KA, Lawrence TJ, Stander EK, Jontos RJ, Kaushal SS, et al. 2010. Opportunities and challenges for managing nitrogen in urban stormwater: A review and synthesis. *Ecol Eng* 36: 1507–1519. doi:10.1016/j.ecoleng.2010.03.015.
- Conley DJ, Paerl HW, Howarth RW, Boesch DF, Seitzinger SP, et al. 2009. Controlling eutrophication: Nitrogen and phosphorus. *Science* 323: 1014–1015. doi:10.1126/science.1167755.
- Cooke RM. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York, New York: Oxford University Press.
- Cooke RM, Goossens LHJ. 2004. Expert judgment elicitation for risk assessments of critical infrastructures. *J Risk Res* 7: 643–656. doi:10.1080/1366987042000192237.
- Cooke RM, Goossens LLHJ. 2008. TU Delft expert judgment data base. *Reliab Eng Syst Safe* 93: 657–674. doi:10.1016/j.res.2007.03.005.
- Correll DL, Jordan TE, Weller DE. 1999. Transport of nitrogen and phosphorus from Rhode River watersheds during storm events. *Water Resour Res* 35: 2513–2521. doi:10.1029/1999WR900058.
- Craig LS, Palmer MA, Richardson DC, Filoso S, Bernhardt ES, et al. 2008. Stream restoration strategies for reducing river nitrogen loads. *Front Ecol Environ* 6: 529–538. doi:10.1890/070080.
- CSN. 2012. Recommendations of the expert panel to define removal rates for new state stormwater performance standards. Chesapeake Stormwater Network.
- Dhaene J, Denuit M, Goovaerts MJ, Kaas R, Vyncke D. 2002. The concept of comonotonicity in actuarial science and finance: Theory. *Insur Math Econ* 31: 3–33. doi:10.1016/S0167-6687(02)00134-8.
- Diebel MW, Maxted JT, Nowak PJ, Vander Zanden MJ. 2008. Landscape planning for agricultural nonpoint source pollution reduction I: A geographical allocation framework. *Environ Manage* 42: 789–802. doi:10.1007/s00267-008-9186-3.
- Dougherty M, Dymond RL, Grizzard TJ, Godrej AN, Zipper CE, et al. 2007. Quantifying long-term hydrologic response in an urbanizing basin. *J Hydrol Eng* 12: 33–41. doi:10.1061/(ASCE)1084-0699(2007)12:1(33).
- EMRPC. 2011. Urban fertilizers & the Chesapeake Bay: An opportunity for major pollution reduction. Environment Maryland Research & Policy Center. Available at <http://www.environmentmaryland.org/reports/mde/urban-fertilizers-chesapeake-bay>.
- Filoso S, Palmer MA. 2011. Assessing stream restoration effectiveness at reducing nitrogen export to downstream waters. *Ecol Appl* 21: 1989–2006. doi:10.1890/10-0854.1.
- Forsee WJ, Ahmad S. 2011. Evaluating urban storm-water infrastructure design in response to projected climate change. *J Hydrol Eng* 16: 865–873. doi:10.1061/(ASCE)HE.1943-5584.0000383.
- Galloway JN, Dentener FJ, Capone DG, Boyer EW, Howarth RW, et al. 2004. Nitrogen cycles: Past, present, and future. *Biogeochemistry* 70: 153–226. doi:10.1007/s10533-004-0370-0.
- Giri S, Nejadhashemi AP, Woznicki S, Zhang Z. 2014. Analysis of best management practice effectiveness and spatiotemporal variability based on different targeting strategies. *Hydrol Process* 28: 431–445. doi:10.1002/hyp.9577.
- Harden SL, Spruill TB. 2008. Factors affecting nitrate delivery to streams from shallow ground water in the North Carolina coastal plain. Reston, Virginia: U.S. Geological Survey. Scientific Investigations Report 2008-5021. Available at <http://pubs.usgs.gov/sir/2008/5021/>.
- Hubbard RK, Sheridan JM. 1983. Water and nitrate-nitrogen losses from a small, upland, coastal plain watershed. *J Environ Qual* 12: 291–295. doi:10.2134/jeq1983.00472425001200020028x.
- Hunt WF, Davis AP, Traver RG. 2012. Meeting hydrologic and water quality goals through targeted bioretention design. *J Environ Eng* 138: 698–707. doi:10.1061/(ASCE)EE.1943-7870.0000504.
- Jordan TE, Correll DL, Weller DE. 1997. Relating nutrient discharges from watersheds to land use and streamflow variability. *Water Resour Res* 33: 2579–2590. doi:10.1029/97WR02005.
- Jordan TE, Weller DE, Correll DL. 2003. Sources of nutrient inputs to the Patuxent River estuary. *Estuaries* 26: 226–243. doi:10.1007/BF02695964.
- Koch BJ, Febria CM, Gevrey M, Wainger LA, Palmer MA. 2014. Nitrogen removal by stormwater management structures: A data synthesis. *J Am Water Resour Assoc* 50: 1594–1607. doi:10.1111/jawr.12223.
- Liu J, Sample DJ, Zhang H. 2014. Frequency analysis for precipitation events and dry durations of Virginia. *Environ Model Assess* 19: 167–178. doi:10.1007/s10666-013-9390-2.
- Liu Z-J, Weller DE, Correll DL, Jordan TE. 2000. Effects of land cover and geology on stream chemistry in watersheds of Chesapeake Bay. *J Am Water Resour Assoc* 36: 1349–1366. doi:10.1111/j.1752-1688.2000.tb05731.x.
- Loperfido JV, Noe GB, Jarnagin ST, Hogan DM. 2014. Effects of distributed and centralized stormwater best management practices and land cover on urban stream hydrology at the catchment scale. *J Hydrol* 519: 2584–2595. doi:10.1016/j.jhydrol.2014.07.007.
- Lowrance R, Altier LS, Newbold JD, Schnabel RR, Groffman PM, et al. 1997. Water quality functions of riparian forest buffers in Chesapeake Bay watersheds. *Environ Manage* 21: 687–712. doi:10.1007/s002679900060.
- Markewich HW, Pavich MJ, Buell GR. 1990. Contrasting soils and landscapes of the Piedmont and Coastal Plain, eastern United States. *Geomorphology* 3: 417–447. doi:10.1016/0169-555X(90)90015-I.



- Maxted JT, Diebel MW, Vander Zanden MJ. 2009. Landscape planning for agricultural non-point source pollution reduction. II. balancing watershed size, number of watersheds, and implementation effort. *Environ Manage* 43:60–68. doi:10.1007/s00267-008-9174-7.
- MDE. 2009. Maryland Stormwater Design Manual, Volumes I and II. Baltimore, Maryland: Maryland Department of Environment. Available at [http://www.mde.state.md.us/programs/Water/StormwaterManagementProgram/Maryland-StormwaterDesignManual/Pages/Programs/WaterPrograms/SedimentandStormwater/stormwater\\_design/index.aspx](http://www.mde.state.md.us/programs/Water/StormwaterManagementProgram/Maryland-StormwaterDesignManual/Pages/Programs/WaterPrograms/SedimentandStormwater/stormwater_design/index.aspx).
- Min S-K, Zhang X, Zwiers FW, Hegerl GC. 2011. Human contribution to more-intense precipitation extremes. *Nature* 470:378–381. doi:10.1038/nature09763.
- Moglen GE, Rios Vidal GE. 2014. Climate change and storm water infrastructure in the mid-Atlantic region: Design mismatch coming? *J Hydrol Eng* 19:04014026. doi:10.1061/(ASCE)HE.1943-5584.0000967.
- Najjar RG, Pyke CR, Adams MB, Breitburg D, Hershner C, et al. 2010. Potential climate-change impacts on the Chesapeake Bay. *Estuar Coast Shelf Sci* 86:1–20. doi:10.1016/j.ecss.2009.09.026.
- Newcomer Johnson TA, Kaushal SS, Mayer PM, Grese MM. 2014. Effects of stormwater management and stream restoration on watershed nitrogen retention. *Biogeochemistry* 121: 81–106. doi:10.1007/s10533-014-9999-5.
- NRC. 2009. *Urban stormwater management in the United States*. Washington, D.C: National Academies Press.
- Palmer MA, Filoso S, Fanelli RM. 2014. From ecosystems to ecosystem services: Stream restoration as ecological engineering. *Ecol Eng* 65:62–70. doi:10.1016/j.ecoleng.2013.07.059.
- Passport E, Vidon P, Forshay KJ, Harris L, Kaushal SS, et al. 2013. Ecological engineering practices for the reduction of excess nitrogen in human-influenced landscapes: A guide for watershed managers. *Environ Manage* 51: 392–413. doi:10.1007/s00267-012-9970-y.
- PLRM Development Team. 2009. Pollutant load reduction model (PLRM): Model development document. Santa Cruz, California: Tahoe Integrated Information Management System.
- Rabotyagov SS, Kling CL, Gassman PW, Rabalais NN, Turner RE. 2014. The economics of dead zones: Causes, impacts, policy challenges, and a model of the Gulf of Mexico hypoxic zone. *Rev Environ Econ Policy* 8: 58–79. doi:10.1093/reep/ret024.
- Rosenberg EA, Keys PW, Booth DB, Hartley D, Burkey J, et al. 2010. Precipitation extremes and the impacts of climate change on stormwater infrastructure in Washington State. *Clim Change* 102:319–349. doi:10.1007/s10584-010-9847-0.
- Rothlisberger JD, Finnoff DC, Cooke RM, Lodge DM. 2012. Ship-borne nonindigenous species diminish Great Lakes ecosystem services. *Ecosystems* 15:1–15. doi:10.1007/s10021-012-9522-6.
- Rothlisberger JD, Lodge DM, Cooke RM, Finnoff DC. 2010. Future declines of The binational Laurentian Great Lakes fisheries: The importance of environmental and cultural change. *Front Ecol Environ* 8:239–244. doi:10.1890/090002.
- Schwartz SS, Smith B. 2014. Slowflow fingerprints of urban hydrology. *J Hydrol* 515: 116–128. doi:10.1016/j.jhydrol.2014.04.019.
- Simpson T, Weammert S. 2009. Developing nitrogen, phosphorus and sediment reduction efficiencies for tributary strategy practices: BMP assessment, Final Report. College Park, Maryland: University of Maryland. Mid-Atlantic Water Program. Available at [http://watershedbmps.com/wp-content/uploads/2012/03/01582\\_to\\_01597\\_BMP\\_ASSESSMENT\\_FINAL\\_REPORT.pdf](http://watershedbmps.com/wp-content/uploads/2012/03/01582_to_01597_BMP_ASSESSMENT_FINAL_REPORT.pdf).
- Stoner AMK, Hayhoe K, Yang X, Wuebbles DJ. 2013. An asynchronous regional regression model for statistical downscaling of daily climate variables. *Int J Climatol* 33:2473–2494. doi:10.1002/joc.3603.
- Strecker EW, Quigley MM, Urbonas BR, Jones JE, Clary JK. 2001. Determining urban storm water BMP effectiveness. *J Water Resour Plan Manag* 127:144–149. doi:10.1061/(ASCE)0733-9496(2001)127:3(144).
- Swain LA, Mesko TO, Hollyday EF. 2004. Summary of the hydrogeology of the Valley and Ridge, Blue Ridge, and Piedmont physiographic provinces in the eastern United States. Reston, Virginia: U.S. Geological Survey. *Professional Paper 1422-A*. Available at <http://pubs.water.usgs.gov/pro1422a>.
- Teck SJ, Halpern BS, Kappel CV, Micheli F, Selkoe KA, et al. 2010. Using expert judgment to estimate marine ecosystem vulnerability in the California Current. *Ecol Appl* 20: 1402–1416. doi:10.1890/09-1173.1.
- USEPA. 2010a. Chesapeake Bay Phase 5.3 Community Watershed Model. Annapolis, Maryland: U.S. Environmental Protection Agency, Chesapeake Bay Program Office. *EPA 903S10002 - CBP/TRS-303-10*. Available at <http://ches.communitymodeling.org/models/CBPhase5/>.
- USEPA. 2010b. Chesapeake Bay total maximum daily load for nitrogen, phosphorus and sediment. Philadelphia, Pennsylvania: U.S. Environmental Protection Agency, Water Protection Division. Available at [http://www.epa.gov/reg3wapd/pdf/pdf\\_chesbay/FinalBayTMDL/CBayFinalTMDLExecSumSection1through3\\_final.pdf](http://www.epa.gov/reg3wapd/pdf/pdf_chesbay/FinalBayTMDL/CBayFinalTMDLExecSumSection1through3_final.pdf).
- Utz RM, Eshleman KN, Hilderbrand RH. 2011. Variation in physicochemical responses to urbanization in streams between two Mid-Atlantic physiographic regions. *Ecol Appl* 21: 402–415. doi:10.1890/09-1786.1.
- Walker S, Selman M. 2014. Addressing risk and uncertainty in water quality trading markets. Washington, D.C.: World Resources Institute. Available at [http://www.wri.org/sites/default/files/WRI\\_IssueBrief\\_Uncertainty\\_3-9\\_Final.pdf](http://www.wri.org/sites/default/files/WRI_IssueBrief_Uncertainty_3-9_Final.pdf).
- Walsh J, Wuebbles D, Hayhoe K, Kossin J, Kunkel K, et al. 2014. Ch 2: Our changing climate, in Melillo JM, Richmond TC, Yohe GW eds., *Climate Change Impacts in the United States: The Third National Climate Assessment*. Washington, D.C.: U.S. Global Change Research Program. pp. 19–67.
- Waters D, Watt WE, Marsalek J, Anderson BC. 2003. Adaptation of a storm drainage system to accommodate increased rainfall resulting from climate change. *J Environ Plan Manag* 46:755–770. doi:10.1080/0964056032000138472.
- Weller DE, Baker ME. 2014. Cropland riparian buffers throughout Chesapeake Bay watershed: Spatial patterns and effects on nitrate loads delivered to streams. *J Am Water Resour Assoc* 50:696–712. doi:10.1111/jawr.12207.
- Weller DE, Baker ME, Jordan TE. 2011. Effects of riparian buffers on nitrate concentrations in watershed discharges: New models and management implications. *Ecol Appl* 21:1679–1695. doi:10.1890/10-0789.1.
- Winer R. 2000. National pollutant removal performance database for stormwater treatment practices, 2<sup>nd</sup> edition. Ellicott City, Maryland: Center for Watershed Protection. Available at <http://www.stormwatercenter.net/Library/STP-Pollutant-Removal-Database.pdf>.

- Wittmann ME, Cooke RM, Rothlisberger JD, Lodge DM. 2014. Using structured expert judgment to assess invasive species prevention: Asian carp and the Mississippi—Great Lakes hydrologic connection. *Environ Sci Technol* **48**: 2150–2156. doi:10.1021/es4043098.
- Wittmann ME, Cooke RM, Rothlisberger JD, Rutherford ES, Zhang H, et al. 2015. Use of structured expert judgment to forecast invasions by bighead and silver carp in Lake Erie: Forecasting Biological Invasion. *Conserv Biol* **29**: 187–197. doi:10.1111/cobi.12369.
- Woltemade CJ. 2000. Ability of restored wetlands to reduce nitrogen and phosphorus concentrations in agricultural drainage water. *J Soil Water Conserv* **55**: 303–309.

#### Contributions

- Contributed to conception and design: BJK, CMF, RMC, SF, JVL, MAP
- Contributed to acquisition of data: BJK, CMF, RMC, JDH, AC, MAP
- Contributed to analysis and interpretation of data: BJK, CMF, RMC, JDH, AC, KH, AMKS, MAP
- Drafted and/or revised the article: BJK, CMF, RMC, JDH, MEB, AC, SF, KH, JVL, AMKS, MAP
- Approved the submitted version for publication: BJK, CMF, RMC, JDH, MEB, AC, SF, KH, JVL, AMKS, MAP

#### Acknowledgments

We are grateful to the experts for generously sharing their knowledge, experience, and time throughout the elicitation process. We thank B. Bierwagen, J. Braithwaite, D. Hogan, S. Julius, L. Wainger, and J. West for their thoughtful insight that greatly improved this research. The views expressed in this article are those of the authors and do not necessarily reflect the policies of the USEPA. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

#### Funding information

Funding for this work was provided by the USEPA Global Change Research Program (GS-10F-0502N to MAP) and the National Socio-Environmental Synthesis Center (SESYNC; NSF DBI-1052875).

#### Competing interests

The authors have no competing interests.

#### Supplementary material

- Appendix S1. SEJ methodology and out-of-sample validation. doi: 10.12952/journal.elementa.000063.s001
- Appendix S2. List of participating experts and affiliations. doi: 10.12952/journal.elementa.000063.s002
- Appendix S3. SEJ protocol document containing all questions posed to the experts. doi: 10.12952/journal.elementa.000063.s003
- Appendix S4. Robustness analysis with calibration and information scores for expert assessments, including all experts. doi: 10.12952/journal.elementa.000063.s004
- Appendix S5. Summary and table of the rationales used by the experts to generate their estimates of TN loads for the questions in the protocol document. doi: 10.12952/journal.elementa.000063.s005
- Appendix S6. Plots of all expert assessments for each variable (60) in protocol. doi: 10.12952/journal.elementa.000063.s006
- Appendix S7. Plots of expert estimates of drainage area-weighted TN loads versus rain event size for the BMPs in the Piedmont and Coastal Plain watersheds. doi: 10.12952/journal.elementa.000063.s007
- Appendix S8. Table summarizing predicted magnitudes of rain events and TN loads for the Coastal Plain and Piedmont watersheds for the nine climate models and two representative concentration pathways (RCP 4.5 and 8.5). doi: 10.12952/journal.elementa.000063.s008

#### Data accessibility statement

Data available from the Dryad Digital Repository: <http://doi.org/10.5061/dryad.3d22k>

#### Copyright

© 2015 Koch et al. This is an open-access article distributed under the terms of The Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.