Spatial Inference of Nitrate Concentrations in Groundwater

DAWN B. WOODARD, ROBERT L. WOLPERT, AND MICHAEL
A. O'CONNELL

We develop a method for multi-scale estimation of pollutant concentrations, based on a nonparametric spatial statistical model. We apply this method to estimate nitrate concentrations in groundwater over the mid-Atlantic states, using measurements gathered during a period of ten years. A map of the fine-scale estimated nitrate concentration is obtained, as well as maps of the estimated county-level average nitrate concentration and similar maps at the level of watersheds and other geographic regions. The fine-scale and coarse-scale estimates arise naturally from a single model, without refitting or ad-hoc aggregation. As a result, the uncertainty associated with each estimate is available, without approximations relying on high spatial density of measurements or parametric distributional assumptions.

Several risk measures are also obtained, including the probability of the pollutant concentration exceeding a particular threshold. These risk measures can be obtained at the fine scale, or at the level of counties or other regions.

The nonparametric Bayesian statistical model allows for this flexibility in estimation while avoiding strong assumptions. This method can be applied directly to estimate ozone concentrations in air, pesticide concentrations in groundwater, or any other quantity that varies over a geographic region, based on approximate measurements at some locations and perhaps of associated covariates.

Key words: Bayesian, geostatistics, kriging, Lévy processes, nonparametrics, response surface, spatial moving average.

Dawn B. Woodard is Assistant Professor of Operations Research and Information Engineering at Cornell University, Ithaca, NY (email woodard@orie.cornell.edu). Robert L. Wolpert is Professor of Statistical Science and Professor of the Environment at Duke University, Durham, NC (email wolpert@stat.duke.edu). Michael A. O'Connell is President, Waratah Corporation, Durham, NC (email michaelo@waratah.com).

1 INTRODUCTION

Nitrate is the most common contaminant of groundwater, and is of concern due to its environmental impact and its health effects when consumed in high levels in drinking water. Nitrate occurs naturally in groundwater, but elevated levels can be caused by contamination by agricultural fertilizer, animal manure, or septic systems, as well as deposition from fossil fuel combustion (Ator and Ferrari 1997). Addressing this concern, measurements of nitrate concentrations in groundwater have been compiled over large geographic regions by the U.S. Geological Survey's National Water Quality Assessment (NAWQA) program (U.S. Department of the Interior, U.S. Geological Survey 2003).

Depending on the particular regulatory or scientific purpose, estimation of nitrate concentrations can be desired at a fine scale, or at the level of counties, watersheds, or other geographic regions. Frequently other measures of risk are also of interest, for instance the probability that a site or region has nitrate concentration exceeding a particular threshold. We propose a method for simultaneous inference of pollutant concentrations and risk measures at multiple scales, based on approximate measurements at a set of locations. We apply this method to obtain inferences for nitrates in groundwater over the mid-Atlantic states.

More precisely, the inferences that we obtain include the following, along with the associated uncertainties:

- The concentration at a fine scale;

- The average concentration over specified regions, *e.g.*, counties, aquifers or watershed regions (indicated by USGS-assigned Hydrological Unit Codes, or *HUC*s), census blocks, states, *etc.*;

- The probability that a particular site has concentration exceeding specified thresholds (*e.g.*, 1, 3, or 10 mg/L), or the average of this probability over a particular region;

- The regions with the highest concentrations.

Since the various regions of interest (counties, HUCs, *etc.*) are not nested (e.g. a county is not necessarily contained within a single HUC), these multiple goals cannot be met by modeling average

concentrations only at a single specified level of spatial aggregation; instead we construct a statistical model for the uncertain nitrate concentration at all locations in the region, from which we can compute average concentrations and other summaries of interest. In our Bayesian formulation, uncertainty about the nitrate concentration and its average over various regions are all random variables, for which we can compute both expected values (best overall estimates) and probabilities of exceeding specified thresholds.

In order to avoid strong assumptions about the distribution of nitrate concentrations, we use a nonparametric model. Our model and estimation methods can be applied directly to estimate the concentrations of other types of pollutants, both in water and in air.

In the Bayesian model, a joint prior distribution for the unobserved spatially-varying concentration is constructed as a moving average of independent-increment random measures (Wolpert, Clyde, and Tu 2006; Clyde, House, and Wolpert 2006). A reversible jump Markov chain Monte Carlo computational approach (Green 1995) is used for approximating the posterior distribution of the concentration at all spatial locations.

We compare our approach to alternative statistical methods for pollutant level estimation. Methods that estimate the pollutant concentration for each county or watershed separately ignore the fact that measurements in neighboring regions are often highly correlated, providing mutually relevant evidence especially in the case of sparse data. Alternative spatial methods include universal kriging and lattice methods (Stein 1999; Cressie and Chan 1989). Our nonparametric Bayesian method makes less restrictive assumptions than these, and allows for inference of multiple risk measures at multiple scales as described above. It is not dependent on the choice of an arbitrary grid size, and is relatively computationally efficient, allowing for the interpretation of large data sets. It also naturally handles co-located and closely-located data without numerical instability.

In Section 2 we describe the nitrates data. Then in Section 3 we give existing methods for spatial estimation of pollutants, and in Section 4 we introduce the moving-average model. The details of implementation are given in Section 5, and in Section 6 we obtain inferences for the nitrates data. In Section 7 we summarize and compare the moving-average model to alternative methods, in the context of pollutant level estimation.

2   NITRATE MEASUREMENTS IN GROUNDWATER

We use nitrate measurements for water samples taken from 929 wells in mid-Atlantic and surrounding states, taken between the years of 1985 and 1996 and compiled from a number of regional studies, as documented in USGS Open File Report 98-158 (Ator 1998). All the data were collected by or in cooperation with the USGS, ensuring a degree of consistency in the measurement methodology. However, due to the multiple studies, the locations of the sampling sites were not chosen using a consistent sampling design. We address the consequences for inference in Section 6.

Frequently during analysis of nitrate concentrations in groundwater multiple samples from the same site, or multiple sites at nearby locations, are removed in order to avoid numerical problems with estimation techniques, or to avoid weighting that site or area too heavily in the statistical analysis. Often the most recent measurement for a particular well, or the shallowest well within a small area, is used in lieu of the full set of measurements (Ator and Ferrari 1997; Nolan, Hitt, and Ruddy 2002). However, multiple measurements at a single site or at nearby sites give us a valuable source of information about the measurement variability and well-to-well variability, so in our analysis we do not remove repeat measurements at a single site or nearby locations.

3   EXISTING SPATIAL METHODS

Lattice methods.   Denote the geographic region by $\mathscr{X}$ and the nitrate concentration at a location $x \in \mathscr{X}$ by $\Lambda(x)$. One common approach begins by dividing the geographic region into a fixed collection of basic subregions $\mathscr{X} = \cup \mathscr{X}_k$, sufficiently fine that variations of the nitrate surface within each $\mathscr{X}_k$ are regarded as unimportant for the desired inferences; denote by Λ_k the average value of $\Lambda(x)$ on $\mathscr{X}_k$. The associated covariates $X_j(x)$ are also taken to be sufficiently slowly-varying that they may be summarized on $\mathscr{X}_k$ by a typical value X_{kj}, and the dependence of Λ_k on the X_{kj}'s can be modeled with linear regression either directly, $\Lambda_k = \sum_{j \in J} X_{kj}\beta_j + Z_k$, or more commonly on a logarithmic scale, $\log \Lambda_k = \sum_{j \in J} X_{kj}\beta_j + Z_k$, with correlated mean-zero residuals $\{Z_k\}$ (reflecting spatial variability, measurement error, and modeling error). Typically a nearest-neighbor spatial dependence is assumed for $\{Z_k\}$ in order to ensure sparsity of the precision matrix and to permit a routine implementation of Bayesian inference (Besag, York, and Mollié 1991).

Lattice models have been applied to nitrates in groundwater by Faulkner (2003).

Kriging.   The kriging approach provides smooth surface estimates for point measurement data, e.g. nitrate concentration, over the entire region. It models $\log \Lambda(x) = \sum_{j \in J} X_j(x) \beta_j + Z(x)$ at all locations $x \in \mathscr{X}$, where $Z(x)$ is a mean-zero Gaussian random field. This random field is specified through a covariance function, which has parameters corresponding to scale, range, and possibly shape and/or smoothness.

The regression coefficients $\{\beta_j\}_{j \in J}$ and the parameters of the covariance function are fit via maximum likelihood or other means, and then the (conditional) means and variances of $\Lambda(x_{i'})$ at any sites $\{x_{i'}\}_{i' \in I'}$ of interest, or the average values of $\Lambda(x)$ over any regions of interest, are computed. Standard kriging does not handle co-located data, so repeated measurements at a single site must be removed or averaged. See Chilès and Delfiner (1999, §2.5) or Cressie (1993, §2.3) for details. Kriging has been applied to groundwater nitrates by LaMotte and Greene (2007).

4   MOVING-AVERAGE BAYESIAN MODELS

Ickstadt and Wolpert (1997) and Wolpert and Ickstadt (1998a) introduced methods for interpolating unobserved intensities of spatial point patterns, based on modeling the intensity $\Lambda(x)$ continuously in space as a moving average of an underlying unobserved independent-increment stochastic process. The idea was later extended to a spatial regression tool (Ickstadt and Wolpert 1999; Best, Ickstadt, and Wolpert 2000). The underlying mathematical and statistical methods have been used in one-dimensional non-point-process applications including identifying proteins in mass spectroscopy (House, Clyde, and Wolpert 2006), as well as a spatio-temporal (three-dimensional) non-point-process application, namely inferring temporal fluctuations in sulfur dioxide air pollution levels in the mid-Atlantic states (Tu 2006). The Inverse Lévy Measure algorithm that underlies the method is described in Wolpert and Ickstadt (1998b).

The spatial model we use is similar to the spatio-temporal model in Tu (2006); our main contribution is to use this model to perform inference for multiple risk measures, at multiple spatial

scales. In the moving-average approach the concentrations $\Lambda(x)$ are modeled as

$$\Lambda(x) = \sum_{j \in J} X_j(x)\beta_j + \sum_{m \in M} k(x, s_m)\,\gamma_m \tag{4.1}$$

for a specified kernel function $k(x, s)$ on $\mathscr{X} \times \mathscr{S}$ and with uncertain regression coefficients $\{\beta_j\}_{j \in J}$ and a number $|M| \leq \infty$ of locations s_m and magnitudes $\gamma_m > 0$ of mixture components. Here $\mathscr{S}$ can be any space, but for the pollutant application we take it to be equal to $\mathscr{X}$. Covariates can be included multiplicatively in the model as well as additively (Best et al. 2000). To complete the statistical model, a likelihood function based on $\Lambda(x)$ and prior distributions for the uncertain quantities must be specified. Additionally, a computational method must be proposed for evaluating the posterior distribution (Section 5.1).

We must first specify a likelihood function, *i.e.*, the probability density function for the observations $\{Y_i \approx \Lambda(x_i)\}_{i \in I}$. Data analysis suggests that the log discrepancies $\{\log[Y_i/\Lambda(x_i)]\}_{i \in I}$ are homoskedastic (that is, the magnitude of variations does not seem to differ markedly with either location x_i or with the magnitudes of the $\{Y_i\}$ themselves), suggesting a log-normal error model $\log Y_i \sim \mathsf{No}\big(\log \Lambda(x_i), \sigma^2\big)$ for some scale parameter σ and leading to the likelihood function

$$L(\omega) = (2\pi\sigma^2)^{-|I|/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i \in I} \big[\log Y_i - \log \Lambda(x_i)\big]^2\right) \tag{4.2}$$

depending on the uncertain parameter vector ω that includes the measurement-error scale σ, the regression coefficient vector $\vec{\beta}$, and the set $\{(s_m, \gamma_m)\}_{m \in M}$ of locations and magnitudes or, more succinctly, the discrete measure

$$\Gamma(ds) \equiv \sum_{m \in M} \gamma_m \delta_{s_m}(ds)$$

with a point mass of magnitude γ_m at each location s_m (δ denotes the Dirac delta distribution function). We can now rewrite Equation 4.1 in integral form as

$$\Lambda(x) = \sum_{j \in J} X_j(x)\beta_j + \int_{\mathscr{S}} k(x, s)\,\Gamma(ds). \tag{4.3}$$

The moving-average (second) term of this model can be viewed in the context of groundwater pollution as being composed of the sum of an unknown number of point sources with unknown locations and magnitudes, where the pollutant concentration decreases with distance from each

source in a manner consistent with the shape of the kernel k. The components can also be viewed as area sources that are highest at a particular location and decrease away from that location according to the shape of k. For the kernel we choose the function

$$k(x,s) = \exp\left\{-\frac{1}{2d^2}\|x-s\|^2\right\} \tag{4.4}$$

for $d > 0$ a fixed constant and distance measure $\|x-s\|$ taken to be great-circle distance. This kernel is a particularly reasonable choice in the context of water pollution since it decreases smoothly as a function of the distance from the center. However, different choices for $k(\cdot,\cdot)$ could eventually be considered, and one could even put a prior distribution over possible choices of $k(\cdot,\cdot)$, parameterizing the uncertainty about the kernel.

All uncertain features of the model can now be expressed explicitly in terms of the parameter vector $\omega = (\sigma, \vec{\beta}, \Gamma)$. Sometimes we will indicate the ω-dependence of $\Lambda(x)$ explicitly by writing $\Lambda(x, \omega)$.

For a Bayesian model we must also specify a joint prior distribution for ω. We do not include covariates in the current analysis of nitrate concentrations in groundwater, so the term involving β in the above model description drops out and we do not need to specify a prior for β. However, available covariate information could easily be added to the model as described above; see Best et al. (2000) for an example of a moving-average regression model in the point-process context.

The prior distributions of σ and Γ are chosen as follows. We adhere to common practice (Gilks, Richardson, and Spiegelhalter 1996) in choosing the inverse gamma distribution for the measurement-error variance σ^2, *i.e.*, model $\sigma^{-2} \sim \mathsf{Ga}(\alpha_\sigma, \rho_\sigma)$ for constant shape parameter $\alpha_\sigma > 0$ and inverse scale parameter $\rho_\sigma > 0$.

Finally for Γ we take a Lévy distribution $\Gamma \sim \mathsf{Lv}(\nu)$, parameterized by a measure $\nu(d\gamma, ds)$ on $\mathbb{R}_+ \times \mathscr{S}$, under which the number $|M|$ of discrete mass points $(\gamma_m, s_m) \in \mathbb{R}_+ \times \mathscr{S}$ has the Poisson distribution $|M| \sim \mathsf{Po}(\nu_+)$ with expectation $\mathsf{E}[|M|] = \nu_+ \equiv \nu(\mathbb{R}_+ \times \mathscr{S})$ and, conditional on $|M|$, the points $\{(\gamma_m, s_m)\}_{m \in M}$ are drawn independently from the probability distribution $\nu(d\gamma, ds)/\nu_+$. Such Lévy distributions assign (a priori) independent infinitely-divisible random variables $\Gamma(A) \perp\!\!\!\perp \Gamma(B)$ to disjoint sets $A, B \subset \mathscr{S}$, $A \cap B = \emptyset$; every such independent assignment can be written uniquely as the sum of a Lévy-distributed $\Gamma(ds)$ and a Gaussian $W(ds)$ (see, for example, Wolpert

and Ickstadt 1998a,b). The specific Lévy random field we will use is the well-known gamma random field on a bounded set $\mathscr{S} \subset \mathbb{R}^d$, whose Lévy measure has density function

$$\nu(\gamma,s) = \alpha\,\gamma^{-1}e^{-\rho\gamma}, \qquad \gamma > 0, \quad s \in \mathscr{S} \tag{4.5}$$

that assigns independent random variables with gamma distributions $\Gamma(A_i) \sim \mathsf{Ga}\big(\alpha|A_i|,\rho\big)$ to disjoint sets A_i. The quantities $\alpha,\rho > 0$ are taken to be constants.

In practice we must truncate $\nu(\gamma,s)$ by setting it to zero for $\gamma < \varepsilon$ for some small $\varepsilon > 0$, to ensure that $\nu_+ \equiv \nu(\mathbb{R}_+ \times \mathscr{S})$ (and hence the expectation of $|M|$) is finite, but we choose $\varepsilon > 0$ to be so small that the omitted mass is negligible. With this choice the prior expectation of $|M|$ is $\nu_+ = \alpha|\mathscr{S}|E_1(\rho\varepsilon)$, where $E_1(z) \equiv \int_z^\infty e^{-t}t^{-1}\,dt$ denotes the *exponential integral function* (Abramowitz and Stegun 1964, *p.* 228), and the total expected mass lost to truncation is less than $\alpha|\mathscr{S}|\varepsilon$.

With these choices the number $|M|$ of terms included in the latent spatial random field is necessarily finite, so the second term in the definition of $\Lambda(x)$ (given in summation form in Equation 4.1) is as smooth as the kernel k–namely C^∞ for the kernel given in (4.4). The question of smoothness becomes more subtle as one considers the limit as $|M| \to \infty$ (or as the truncation parameter $\varepsilon \to 0$), as explored in §3.1 of Wolpert et al. (2006).

5 IMPLEMENTATION OF THE BAYESIAN MODEL

5.1 COMPUTATIONS

The previous section gives the likelihood function $L(\omega)$ and the joint prior distribution $\pi(d\omega)$ for all the components of ω. We may compute summaries of interest from the joint posterior distribution, which is proportional to the product of prior and likelihood. These include the posterior mean $\mathsf{E}[g(\omega)]$ and distribution $\mathsf{P}[g(\omega) \in A]$ of any quantity of interest $g(\cdot)$, such as the value $g_1(\omega) \equiv \Lambda(x,\omega)$ of the uncertain pollutant concentration at a particular site $x \in \mathscr{X}$ or its average $g_2(\omega) \equiv \int_A \Lambda(x,\omega)\,dx/|A|$ over any region $A \subset \mathscr{X}$. The posterior mean $\mathsf{E}[g(\omega)]$ is equal to the ratio of integrals:

$$\mathsf{E}[g(\omega)] = \frac{\int_\Omega g(\omega)\,L(\omega)\,\pi(d\omega)}{\int_\Omega L(\omega)\,\pi(d\omega)}.$$

Once we succeed in evaluating this ratio of integrals we can generate consistent estimates of a wide variety of quantities, such as

- the average of $\Lambda(x)$ over any specified set A;

- the maximum value of $\Lambda(x)$ within any specified set A;

- the probability that $\Lambda(x)$ exceeds any threshold λ^* at a particular site $x \in \mathscr{X}$.

To compute posterior means and distributions of any quantity $g(\omega)$ we implement the Metropolis-Hastings variation of the Markov chain Monte Carlo (MCMC) simulation-based computational method (Tierney 1994). In this approach we construct a Markov chain $\{\omega^t\}_{t \in \mathbb{N}} = \{(\sigma^t, \Gamma^t)\}_{t \in \mathbb{N}}$ that approaches the posterior distribution $\pi(d\omega \mid \vec{Y})$ as $t \to \infty$; details of the Metropolis-Hastings computation for the moving-average model are given in Appendix A. Then we can evaluate posterior expectations by ergodic averages:

$$\mathsf{E}\big[g(\omega)\big] = \lim_{T \to \infty} \frac{1}{T} \sum_{t \leq T} g(\omega^t). \tag{5.1}$$

For instance, one can estimate the posterior mean of $\Lambda(x, \omega)$ at $x \in \mathscr{X}$ as $\frac{1}{T} \sum_{t \leq T} \Lambda(x, \omega^t)$. Similarly, to obtain a Monte Carlo estimate of the average of $\Lambda(x, \omega)$ over a region $A \subset \mathscr{X}$ one can sample K locations $\{a_i\}_{i \leq K}$ uniformly at random in A, and take $\frac{1}{TK} \sum_{i,t} \Lambda(a_i, \omega^t)$.

Due to computational limitations we approximate the great-circle distance $\|x - s\|$ in Equation 4.4 by taking the easting/northing coordinates of x and s with respect to the centroid of the region $\mathscr{X}$, and calculating the Euclidean distance. For a small geographic area this method yields a very close approximation to the great-circle distance; on the mid-Atlantic nitrate data it is correct to within 20%. This approximation is necessary since we must evaluate the likelihood at each iteration of the Markov chain, and since each likelihood evaluation involves computing the distance between each of the 929 measurement locations and at least several hundred kernel locations, leading to more than 10^{11} distance calculations for a chain with 10^6 iterations. Parallel computation and the use of more efficient Metropolis proposals could be used to mitigate this difficulty (see discussion in Section 7).

5.2 SPECIFICATION OF CONSTANTS

In order to specify the model fully we must choose values for the constants in the prior distributions described in Section 4. There is a prior for the measure $\Gamma(ds)$ and a prior for the outcome variance on the log scale, σ^2. The prior for $\Gamma(ds)$ is determined by the quantities ε, α, and ρ, while the prior for σ^2 is determined by α_σ and ρ_σ. We must also choose the range constant d in Equation 4.4.

The nitrate measurements in the data set show substantial variability from well to well within a short distance; nitrate measurements of wells within a 10 km^2 area commonly vary by a factor of 1.15–2. This range is obtained by dividing the geographic region up into a grid boxes of length 10 kilometers on a side, and for boxes that have more than one measurement, taking the standard deviation of the nitrate measurements within the box on the log scale and exponentiating. This gives a metric of the average variability of local nitrate measurements in terms of multiplicative factors. The factors 1.15 and 2 are the 25th and 75th quantiles of the resulting values over the different grid boxes.

On the log scale the range 1.15-2 corresponds to a standard deviation of 0.14–0.69. Therefore we take the values 0.14 and 0.69 to be the 25th and 75th quantiles, respectively, of the prior distribution of σ. These quantiles for σ imply the prior distribution $\sigma^{-2} \sim \mathsf{Ga}(\alpha_\sigma = 0.39, \rho_\sigma = 0.0098)$ for σ^{-2}.

There is believed to be long-range geographic dependence of nitrate concentrations in groundwater. Nitrate loading is high in some regions but not in others, and geologic factors vary by region and strongly affect nitrate concentrations in groundwater. However, extreme long-distance dependence of nitrate concentrations is not likely since divisions between geologic zones, aquifers, and land-use regions limit this dependence (Ator and Ferrari 1997; Nolan et al. 2002). Therefore we set the kernel radius d to be an intermediate value of 40 km.

We take the prior mean of $\Lambda(x, \omega)$ at all locations $x \in \mathscr{X}$ to be equal to the median of the nitrate measurements (4.4 mg/L). One might consider taking the prior mean to be equal to the mean of the nitrate measurements; however, the mean of the measurements is sensitive to the values of outliers and the median of the data is a slightly lower and more conservative choice that still captures the

centering of the data.

Nitrate concentrations in groundwater are known to be less than 0.4 mg/L in some areas and more than 10 mg/L in other areas (Ator and Ferrari 1997). Therefore we take the prior standard deviation of $\Lambda(x, \omega)$ to be equal to 3, so that the values 0.4 mg/L and 10 mg/L are within two standard deviations of the prior mean.

The prior mean and variance of $\Lambda(x, \omega)$ are approximately $2\pi d^2 \alpha / \rho$ and $\pi d^2 \alpha / \rho^2$, respectively, where the approximation is as $\mathscr{X} \to \mathbb{R}^2$ and $\varepsilon \to 0$ (see Appendix B). In order to have a prior geometric standard deviation of $3/4.4$ for $\Lambda(x, \omega)$, we must have $\alpha = 1.07 \times 10^{-4} \text{ km}^{-2}$ and $\rho = 0.244 \text{ L/mg}$.

The truncation constant $\varepsilon > 0$ should be chosen small enough so that the fraction of mass that is truncated is small, ensuring that the above approximations for the mean and variance of $\Lambda(x)$ are fairly accurate. However, if ε is too small then the prior expectation of $|M|$ is large, increasing the computation time of the MCMC. We choose $\varepsilon = 0.0412$ mg/L, guaranteeing that the fraction of mass that is truncated is less than one percent while yielding a manageable prior expectation for $|M|$ (about 400).

In order to verify that the analytic approximations given above for the mean and variance of $\Lambda(x, \omega)$ are accurate enough for our purposes, we ran the MCMC to sample from the prior, and estimated the prior mean and standard deviation from this sample. We obtained a prior mean of 4.38 mg/L and a prior standard deviation of 3.01 mg/L, which are very close to our desired values. The prior probability that $\Lambda(x, \omega) > 10$ mg/L, as estimated from the prior sample, is approximately 5% at any location x within the study region, which matches our prior belief that 10 mg/L is a high nitrate concentration but does occur in some areas.

The prior choices given here lead to realized prior surfaces $\Lambda(x, \omega)$ such as the one in Figure 1. Visually, there are some areas with high nitrate concentrations and some with quite low nitrate concentrations, and the locations of the high-nitrate regions are unknown a priori. This model, wherein the surface $\Lambda(x, \omega)$ is composed of the sum of an unknown number of point / area sources with unknown centers and magnitudes, is particularly appropriate in the context of estimation of pollutant concentrations.

6 RESULTS OF THE NITRATES ANALYSIS

Nitrates occur naturally in groundwater; the naturally occurring concentration is not well understood and varies from region to region, but has been estimated to be 0.4 mg/L in parts of the Delmarva Peninsula and the Potomac River Basin (Hamilton et al. 1993; Ator and Denis 1997). Most (80%) of the nitrate measurements described in Section 2 are above this naturally occurring concentration.

The data were gathered over a period of 11 years, so a strong trend in nitrate concentrations over that period of time could affect the accuracy of our analysis. However, we do not find a significant trend in the measured nitrate concentrations in the 11 years over which the data were gathered. A Kolmogorov-Smirinov two-sample test does not find a difference between the nitrate distribution in the first half of the data (before July 1, 1992) and that in the second half (after July 1, 1992), at significance level $\alpha = .01$. A linear model fit to the nitrate measurements as a function of the Julian date also finds no significant trend in time.

There is a clear trend in the choice of sampling locations over time. In particular, the second half of the data was gathered from a much larger geographic area than the first half. However, absent a trend in nitrate concentrations or distribution over time this will not cause a bias in our analysis.

As discussed in Section 2, the data have been compiled from a number of regional studies and thus the measurement locations have been chosen using sampling designs that vary by region. Again this will not bias the results of our modeling, so long as the choice of sampling locations is independent of the nitrate concentration. This independence may not hold if there are confounding factors such as land use that affect both the nitrate concentration and the sampling locations. However, if such confounding factors are suspected and data are available for the confounding factors, then one could control for them by including them as covariates in the model, as described in Section 4.

In order to estimate the spatial surface of nitrate concentrations, we first apply kriging. We fit the Gaussian random field model to the log transformation of nitrate concentration in order to make the Gaussian assumption more plausible (fitting the model on the original scale leads to poorer

estimates due to the heavy right-skew of the measurements). Additionally, fitting the Gaussian random field model on the log scale allows for more direct comparison with the moving-average model (for which we have assumed a log-normal likelihood, given in Equation 4.2).

We use a spherical covariance function, and obtain estimates of the parameters of the covariance function via a least squares fit of the theoretical variogram to the empirical variogram. We then use these parameter estimates to obtain a kriged estimate of the transformed nitrate concentration. The resulting estimated nitrate concentration map is shown in Figure 2, along with lower and upper 95% kriged confidence bounds. As is clear from the maps, the confidence limits are so wide in many places as to be practically meaningless, since the lower bound is typically less than the estimated natural nitrate concentration of 0.4 mg/L and the upper bound is above the federal Maximum Contaminant Level of 10 mg/L (U.S. Environmental Protection Agency 1991) in much of the study area. The nitrate measurements are overlaid in Figure 2, showing that the wide confidence intervals are attributable in some areas to lack of data. However, there are also some areas with very wide intervals that have a large number of measurements.

Next we apply the spatial moving-average approach described in Section 4. The Markov chain was run for a burn-in period of 10^5 iterations, followed by a sampling period of 10^6 iterations, saving every hundredth sample. Using this choice of burn-in and sampling length, trace plots show no lack of convergence of the chain although there is substantial autocorrelation of the chain up to lag 50. This autocorrelation does not invalidate inferences obtained from the Markov chain; however, it does increase the standard errors associated with these inferences. One could more formally estimate the standard error of the Monte Carlo approximations and stop the Markov chain when the standard errors fall below a specified value (Jones et al. 2006).

Figure 3 shows a plot of the posterior mean of the nitrate concentration $\Lambda(x)$. Locations with no data nearby are estimated to have mean close to the prior mean (taken to be 4.4 mg/L, the median of the measurements in the data set, as described in Section 5.2). In regions where numerous and consistently low measurements were taken, such as West Virginia and western Pennsylvania, the estimated concentration is very low. In regions where many high measurements were taken, such as southeast Pennsylvania and the Chesapeake region, the estimated concentration is high. The Chesapeake region has the highest estimated concentration, due to numerous high readings

on the Delmarva Peninsula in Virginia and at a particular site on the west coast of the Chesapeake. The highest measurements in the data set (34 and 29 mg/L) were collected on the Delmarva Peninsula, at -76.0 degrees Longitude, 37.2 degrees Latitude. There are 13 other high measurements very close to the same location and thus indistinguishable in the Figure. There are also four high measurements at locations very close to -75.8 degrees Longitude, 37.6 degrees Latitude. In southern Maryland along the western Chesapeake coast is the site mentioned before, where 43 high measurements were taken. The moving-average model interpolates between this site and the high Delmarva measurements, estimating high concentrations in the Chesapeake region of southern Maryland and eastern Virginia. This region is made up of several separate peninsulas whose groundwater is presumably separated by the geographical barrier of the bay, so it may be reasonable in future moving-average analyses to limit the dependence of these peninsulas in the model. This region is not estimated to be a hot spot in the kriged estimates, due to the fact that repeated measurements at the same site are removed for the kriging analysis.

A plot of the posterior standard deviation of the concentration $\Lambda(x)$ is also shown in Figure 3. The standard deviation is high in places where there is large uncertainty about the concentration, namely areas with little or no data, or areas where the measurements have high variability. The area in the Chesapeake region noted above for having high estimated concentrations $\Lambda(x)$ also has high posterior standard deviation of $\Lambda(x)$. The posterior standard deviation in Bayesian modeling plays the same role as standard errors in classical statistical analysis, in the sense that a parameter's 95% posterior interval, for the case where the posterior distribution is Gaussian, is the posterior mean $\pm$ two posterior standard deviations. Most areas in Figure 3 that have numerous measurements have very low standard deviations; this is in contrast to the kriged surface, which has very wide confidence intervals in many of these places, for instance central Maryland.

There is a slight spotty quality to Figure 3 in regions with no data; this is due to noise introduced by the computational method. Running the algorithm for more iterations or improving the efficiency of the software would reduce this spottiness; we leave this to future analyses.

Average nitrate concentrations over regions such as counties can also be obtained. Figure 4 shows such averages at the county level; the counties shown in red are those for which the nitrate concentration is estimated to be the highest. The estimates are only shown for counties within states

for which some data are available (although the estimates can be computed for any other county that is within the modeling region, there is likely little regulatory interest in such estimates).

Figure 5 shows a map of the posterior probability that the nitrate concentration $\Lambda(x)$ exceeds the federal Maximum Contaminant Level of 10 mg/L, averaged by county. In locations with no data nearby, this probability is approximately its prior value of 5% (the counties in green around the edges are due to edge effects of the model). Counties where data were gathered almost all have a very low posterior probability (less than 2%) of exceeding 10 mg/L. The only exceptions are in the Chesapeake coast area that was noted for having high estimated nitrate concentration; this region is predicted to have the highest posterior probability of exceedance, due to the high measurements in the area. This area has 5 counties with average probabilities of exceedance above 8%, namely the Virginia counties Northumberland (73.7%), Lancaster (57.5%), Richmond (35.6%), Middlesex (16.6%), and Westmoreland (16.0%).

The low probabilities of exceedance in the large majority of counties is due to the fact that there are many low nitrate measurements in the data set. Only 15% of the measurements are above 10 mg/L, and fitting the model to the data results in attribution of these high measurements to local well-to-well and sample-to-sample variability, rather than to high values of the nitrate concentration $\Lambda(x)$. Since the measurements in the data set are quite variable, even when restricting to a small area, this well-to-well and sample-to-sample variability, captured by the parameter σ, is estimated to be high. In fact, the (posterior mean) estimate of σ is 1.38, well above its prior 75th quantile of 0.69.

7   DISCUSSION AND CONCLUSIONS

Using the moving-average Bayesian model we have obtained several risk measures, each at multiple scales. For the nitrates data we have estimated the average nitrate concentration both at a fine scale and by county, as well as the probability of exceeding the regulatory threshold, averaged by county.

The fine-scale estimated nitrate concentration map is very reasonable; hot spots are captured effectively, and the posterior intervals are wide in areas with little data and generally narrow in

areas with much data. The nitrates data set exhibits a great deal of variability in measurements taken over even a small area, and the model attributes this variability correctly. For this reason, the high nitrate measurements in the data set are attributed to this measurement variability rather than to high regional nitrate levels, so that few counties have high probability of exceeding the regulatory threshold.

One possible extension of the moving-average model would be the addition of a mean parameter β_0 to capture the background, or natural, nitrate concentration. One could also replace the fixed kernels in the model with kernels specified using a prior on the scale and eccentricity, as in Tu (2006). This would improve the ability of the model to capture, for example, pollutant point sources that spread out more in one direction than another due to wind or water flow patterns. The fixed values of the spatial density parameter α and the kernel height parameter ρ could also be replaced by prior distributions over the same parameters, improving the flexibility of the model. For the nitrates analysis, one could also add covariates such as land use and geologic and climatic factors that are believed to affect the absorption and dispersion of nitrates in groundwater.

The primary advantage of the moving-average Bayesian model is that it addresses a variety of desired inference questions as simple summaries of an estimated posterior distribution (*e.g.*, spatial nitrate concentration distribution) from a single model formulation and without model re-fitting. In particular, we can compute average concentrations over a variety of specified regions as well as probabilities that the concentration exceeds specified thresholds, averaged by region. Since pollutant level estimation is carried out on a variety of scales and regulation scenarios are multi-faceted in nature, the Bayesian modeling framework is particularly useful in this context. By contrast, lattice models require the specification of a single partition (*e.g.*counties) or nested partitions (Nakaya 2000), and offer no way of giving consistent estimates across non-nested partitions, nor of exploring variation within partition elements. Not surprisingly, inferences from lattice models can be sensitive to the choice of lattice size.

The moving-average model is nonparametric, avoiding the Gaussian distributional assumption of kriging. The moving-average model instead assumes that the surface consists of the sum of an unknown number of kernels with unknown locations and heights, a flexible representation and one which is reasonable in the context of pollutant level estimation, since the kernels can be interpreted

as point or area sources.

The moving-average approach also has a computational advantage over the kriging method for large data sets; computing the likelihood for the moving-average model requires $O(|I||M|)$ operations, where $|I|$ is the number of data points and $|M|$ is the number of kernels, compared to $O(|I|^3)$ for kriging. Kriging has even more difficulty computing the conditional mean and variance of $\Lambda(x')$ at the points of interest x', which requires $O(|I \cup I'|^3)$ operations where I' indexes the points of interest (compared to $O(|I'||M|)$ for the moving-average model). This is an obstacle if estimates are required at a large number of points, for instance to permit high-resolution image plots. The efficiency of kriging can be improved by using a short-range covariance function, allowing the use of sparse matrix techniques (Fields Development Team 2004; Lophaven, Nielsen, and Søndergaard 2002), but then long-range dependence in the data cannot be captured. By contrast, the moving-average model can capture long-range dependence for much larger data sets.

Implementation of the reversible jump MCMC method for the moving-average model is no more difficult than implementation of a standard Metropolis-Hastings algorithm (see Appendix A). The convergence and mixing of this Markov chain for the moving-average model can be slow, since kernels are added or deleted one at a time, and the model can include hundreds of kernels for applications like the nitrates example. However, the computation is naturally parallelizable (by simulating multiple chains), and more efficient (non-local) Metropolis moves could be explored.

Due to its modeling flexibility and computational advantages for large data sets, the moving-average Bayesian approach is promising for multi-scale estimation for nitrates and other pollutants, in both air and water.

Acknowledgments

A   DETAILS OF THE COMPUTATIONAL METHOD

Let Ω be the parameter space, so that the elements $\omega \in \Omega$ are the parameter vectors $\omega = (\sigma, \Gamma)$. Here we omit a description of updating for the optional regression coefficients β, but this updating can be performed in a manner analogous to that for the other parameters.

The Metropolis-Hastings (MH) proposal distribution $Q(d\omega^* \mid \omega)$ is specified as follows. Choose probabilities $\{p_\sigma, p_\Gamma\}$ summing to one; with probability p_σ draw σ^2 from its full conditional posterior distribution (inverse gamma), and with probability p_Γ propose a change in Γ as follows.

Proposed moves $\Gamma \to \Gamma^*$ are of three different types: introduce a new mass point (γ_m, s_m), and increment $M^* = M \cup \{m\}$; remove an existing mass point (γ_m, s_m), and decrement $M^* = M \setminus \{m\}$; and move an existing mass point (γ_m, s_m), leaving $M^* = M$ unchanged, with probabilities p_Γ^+, p_Γ^-, and $p_\Gamma^=$, respectively, that sum to p_Γ. Points moved outside $s_m^* \notin \mathscr{S}$ are reflected back into $\mathscr{S}$; masses decreased to values $\gamma_m^* < \varepsilon$ are removed. New points are drawn from a specified "birth" distribution with a density function $b(\gamma, s)$; old points are removed with equal probability; and existing points are moved with normally-distributed random walk steps in $\log \gamma_m$ and s_m, constrained to avoid leaving $([\varepsilon, \infty) \times \mathscr{S})$, and tuned to ensure an acceptance rate of approximately 25–40% (Roberts, Gelman, and Gilks 1997).

The space of possible values for Γ may be described as the disjoint union

$$\mathscr{U} \equiv \bigcup_{m=0}^{\infty} (\mathbb{R}_+ \times \mathscr{S})^m$$

of Cartesian powers of $(\mathbb{R}_+ \times \mathscr{S})$. It is possible to write a density function $\pi_v(\Gamma) = \pi_v(d\Gamma)/d\Gamma$ for the prior distribution $\pi_v(d\Gamma)$ of Γ, with respect to a reference measure $d\Gamma$ given by Poisson measure with rate $e^{-\gamma} d\gamma ds$ (i.e., the sum of Lebesgue measure for s and a standard exponential distribution for γ, scaled by $e^{-|\mathscr{S}|}/m!$ on each component $(\mathbb{R}_+ \times \mathscr{S})^m$ of the union), given there by $\pi_v(\Gamma) = e^{|\mathscr{S}| - v_+} \prod_{m \in M} (\alpha \gamma_m^{-1} e^{(1-\rho)\gamma_m})$ and thus the overall log prior density function for $\omega \equiv (\sigma, \Gamma) \in \Omega \equiv (\mathbb{R}_+ \times \mathscr{U})$ is given by:

$$
\begin{aligned}
\log \pi(\omega) \;=\; & -\log \frac{\Gamma(\alpha_\sigma)}{2} + \alpha_\sigma \log \rho_\sigma - (2\alpha_\sigma + 1)\log \sigma - \rho_\sigma/\sigma^2 \qquad \text{(A.1)}\\
& + |\mathscr{S}| - v_+ + |M| \log \alpha - \sum_{m \in M} \log \gamma_m + (1 - \rho) \sum_{m \in M} \gamma_m
\end{aligned}
$$

with respect to Lebesgue measure on $\mathbb{R}_+$ for σ and the reference measure $d\Gamma$ for $\Gamma \in \mathscr{U}$. Combining the likelihood (Equation 4.2), log prior (Equation A.1), and the above description of the move proposals, we can express the MH acceptance probability as

$$
\frac{\pi(d\omega^*)L(\omega^*)Q(d\omega \mid \omega^*)}{\pi(d\omega)L(\omega)Q(d\omega^* \mid \omega)}
$$

$$
= \frac{L(\omega^*)}{L(\omega)} \times
\begin{cases}
\exp\left[\rho(\gamma_m - \gamma_m^*)\right] & \text{for } \Gamma^= \text{ moves} \\[2ex]
\dfrac{p_\Gamma^- + p_\Gamma^= \Phi\left(\log(\varepsilon/\gamma_m^*)/\delta_\gamma\right)}{p_\Gamma^+ b(\gamma_m^*, s_m^*)M^*/\nu(\gamma_m^*, s_m^*)} & \text{for } \Gamma^+ \text{ additions} \\[2ex]
\dfrac{p_\Gamma^+ b(\gamma_m, s_m)M/\nu(\gamma_m, s_m)}{p_\Gamma^- + p_\Gamma^= \Phi\left(\log(\varepsilon/\gamma_m)/\delta_\gamma\right)} & \text{for } \Gamma^- \text{ deletions}
\end{cases}
$$

where δ_γ is the scale of the normally-distributed proposal for $\log \gamma_m$.

B   RANDOM FIELD MEAN AND VARIANCE

Following Equation 17 of Wolpert et al. (2006), the prior mean of $\Lambda(x)$ is:

$$
\mathrm{E}[\Lambda(x)] = \int_{\mathscr{S}} \int_0^\infty k(x,s)\gamma\nu(d\gamma, ds)
$$

$$
\approx \int_{\mathbb{R}^2} \int_0^\infty \exp\left\{-\frac{1}{2d^2}\|x - s\|^2\right\} \alpha e^{-\rho\gamma} d\gamma\, ds = 2\pi d^2 \alpha \rho^{-1}
$$

and the prior covariance is:

$$
\mathrm{Cov}[\Lambda(x), \Lambda(y)] = \int_{\mathscr{S}} \int_0^\infty k(x,s)k(y,s)\gamma^2\nu(d\gamma, ds)
$$

$$
\approx \int_0^\infty \pi d^2 \exp\left\{-\frac{1}{4d^2}\|x - y\|^2\right\} \alpha\gamma e^{-\rho\gamma} d\gamma
$$

$$
= \alpha\pi d^2 \rho^{-2} \exp\left\{-\frac{1}{4d^2}\|x - y\|^2\right\}.
$$

In particular, $\mathrm{Var}[\Lambda(x)] = \alpha\pi d^2/\rho^2$.

REFERENCES

Abramowitz, M. and Stegun, I. A. (eds.) (1964), *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, vol. 55 of *Applied Mathematics Series*, Washington, DC: National Bureau of Standards.

Ator, S. W. (1998), "Nitrate and pesticide data for waters of the mid-Atlantic region," USGS Open File Report 98-158, Reston, VA: U.S. Geological Survey.

Ator, S. W. and Denis, J. M. (1997), "Relation of nitrogen and phosphorus in ground water to land use in four subunits of the Potomac River Basin," USGS Water-Resources Investigations Report 97-4268, Reston, VA: U.S. Geological Survey.

Ator, S. W. and Ferrari, M. J. (1997), "Nitrate and selected pesticides in ground water of the mid- Atlantic region," USGS Water-Resources Investigations Report 97-4139, Reston, VA: U.S. Geological Survey.

Besag, J., York, J., and Mollié, A. (1991), "Bayesian image restoration, with two applications in spatial statistics (with comments)," *Annals of the Institute of Statistical Mathematics*, 43, 1–59.

Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000), "Spatial Poisson regression for health and exposure data measured at disparate resolutions," *Journal of the American Statistical Association*, 95, 1076–1088.

Chilès, J.-P. and Delfiner, P. (eds.) (1999), *Geostatistics, Modeling Spatial Uncertainty*, New York: Wiley.

Clyde, M. A., House, L. L., and Wolpert, R. L. (2006), "Nonparametric models for proteomic peak identification and quantification," in *Bayesian Inference for Gene Expression and Proteomics*, eds. K. A. Do, P. Muller, and M. Vannucci, Cambridge University Press, pp. 293–308.

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley.

Cressie, N. and Chan, N. H. (1989), "Spatial modeling of regional variables," *Journal of the American Statistical Association*, 84, 393–401.

Dey, D., Müller, P., and Sinha, D. (eds.) (1999), *Practical Nonparametric and Semiparametric Bayesian Statistics*, New York: Springer-Verlag.

Faulkner, B. R. (2003), "Confronting the modifiable areal unit problem for inference on nitrate in regional shallow ground water," in *Groundwater Quality Modeling and Management Under Uncertainty*, ed. S. Mishra, Reston, VA: American Society of Civil Engineers, pp. 248–259.

Fields Development Team (2004), *Fields: Tools for Spatial Data*, National Center for Atmospheric Research, Boulder, CO. Available at `http://www.cgd.ucar.edu/stats/Software/Fields/`.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1996), *Markov Chain Monte Carlo in Practice*, New York: Chapman and Hall.

Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Hamilton, P. A., Denver, J. M., Phillips, P. J., and Shedlock, R. J. (1993), "Water-quality assessment of the Delmarva Peninsula, Delaware, Maryland, and Virginia – Effects of agricultural activities on, and distribution of, nitrate and other inorganic constituents in the surficial aquifer." USGS Open File Report 93-40, Reston, VA: U.S. Geological Survey.

House, L. L., Clyde, M. A., and Wolpert, R. L. (2006), "Nonparametric models for peak identification and quantification in mass spectroscopy, with application to MALDI-TOF," Discussion Paper 2006-24, Duke University, Dept. of Statistical Science, URL `ftp://ftp.isds.duke.edu/pub/WorkingPapers/06-24.html`.

Ickstadt, K. and Wolpert, R. L. (1997), "Multiresolution assessment of forest inhomogeneity," in *Case Studies in Bayesian Statistics, Volume III*, eds. C. Gatsonis, J. S. Hodges, R. E. Kass, R. E. McCulloch, P. Rossi, and N. D. Singpurwalla, New York: Springer-Verlag, pp. 371–386.

— (1999), "Spatial regression for marked point processes (with comments)," in *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Oxford University Press, pp. 323–341.

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006), "Fixed-width output analysis for Markov chain Monte Carlo," *Journal of the American Statistical Association*, 101, 1537–1547.

LaMotte, A. E. and Greene, E. A. (2007), "Spatial analysis of land use and shallow groundwater vulnerability in the watershed adjacent to Assateague Island National Seashore, Maryland and

Virginia, USA," *Environmental Geology*, 52, 1413–1421.

Lophaven, S. N., Nielsen, H. B., and Søndergaard, J. (2002), "DACE: A Matlab Kriging toolbox, Version 2.0," Technical Report IMM-TR-2002-12, Technical University of Denmark. Available at `http://www.imm.dtu.dk/∼hbn/dace/dace.pdf`.

Nakaya, T. (2000), "An information statistical approach to the modifiable areal unit problem in incidence rate maps," *Environment and Planning A*, 32, 91–109.

Nolan, B. T., Hitt, K. J., and Ruddy, B. C. (2002), "Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States," *Environmental Science and Technology*, 36, 2138–2145.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997), "Weak convergence and optimal scaling of random walk Metropolis algorithms," *Annals of Applied Probability*, 7, 110–120.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer-Verlag.

Tierney, L. (1994), "Markov chains for exploring posterior distributions (with discussion)," *Annals of Statistics*, 22, 1701–1762.

Tu, C. (2006), "Bayesian nonparametric modeling using Lévy process priors with applications for function estimation, time series modeling, and spatio-temporal modeling," PhD thesis, Duke University, Dept. of Statistical Science.

U.S. Department of the Interior, U.S. Geological Survey (2003), "USGS national water-quality assessment program (NAWQA)," Home page: `http://water.usgs.gov/nawqa/`.

U.S. Environmental Protection Agency (1991), "Fact sheet: National primary drinking water standards," Washington, D.C.: U.S. Government Printing Office.

Wolpert, R. L., Clyde, M. A., and Tu, C. (2006), "Lévy adaptive regression kernels," Discussion Paper 2006-08, Duke University, Dept. of Statistical Science, URL `http://ftp.stat.duke.edu/WorkingPapers/06-08.html`. Revised, March 2009.

Wolpert, R. L. and Ickstadt, K. (1998a), "Poisson/gamma random field models for spatial statistics," *Biometrika*, 85, 251–267.

— (1998b), "Simulation of Lévy random fields," in Dey, Müller, and Sinha (1999). pp. 227-242.

List of Figures

Figure 1. A concentration surface sampled from the prior distribution for the nitrates example.

Figure 2. The kriged estimate (top left), and kriged lower (top right) and upper (bottom left) 95% confidence bounds for the nitrate concentration. The well locations are shown in each plot, with symbols indicating the magnitude of the measurement(s) at that well.

Figure 3. Posterior expected nitrate concentration (left) and posterior standard deviation of the nitrate concentration (right), in mg/L. The well locations are also shown, with symbols indicating the magnitude of the measurement(s) at that well.

Figure 4. Expected nitrate concentration, averaged over each county.

Figure 5. Posterior probability of the nitrate concentration $\Lambda(x)$ exceeding 10 mg/L, averaged over each county.
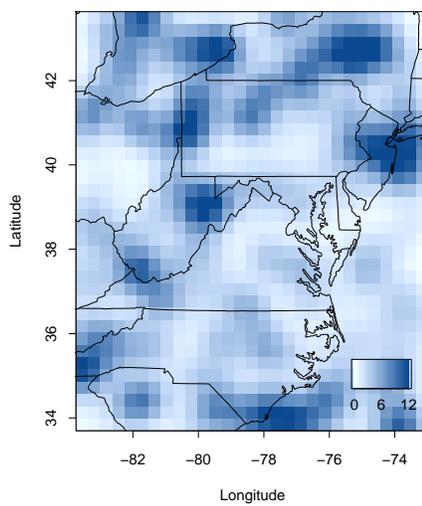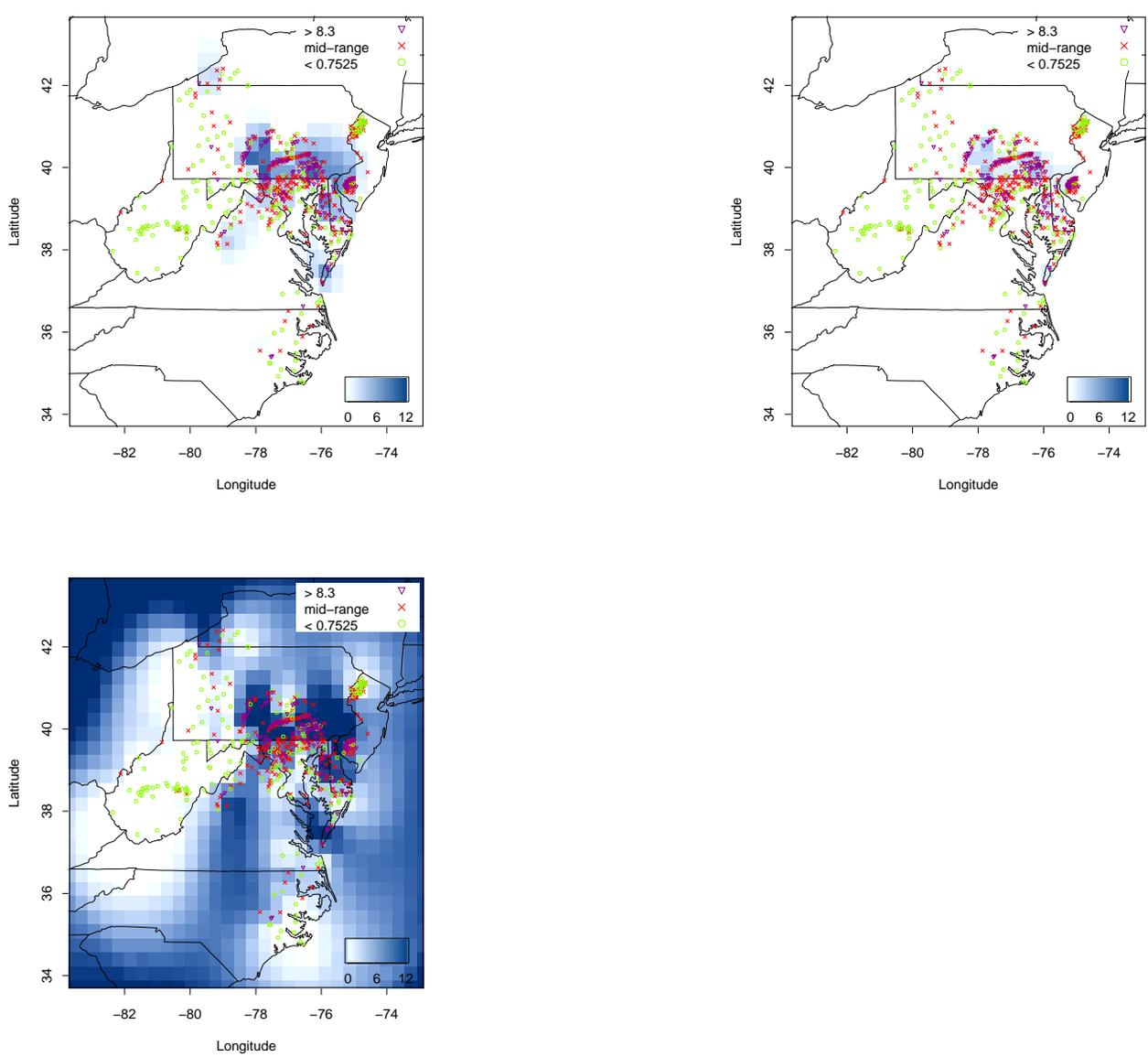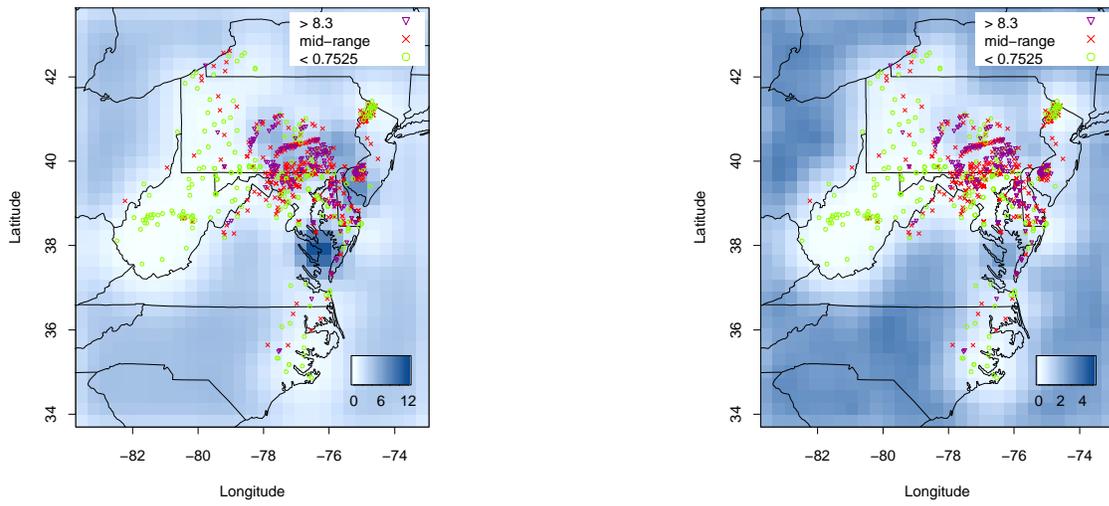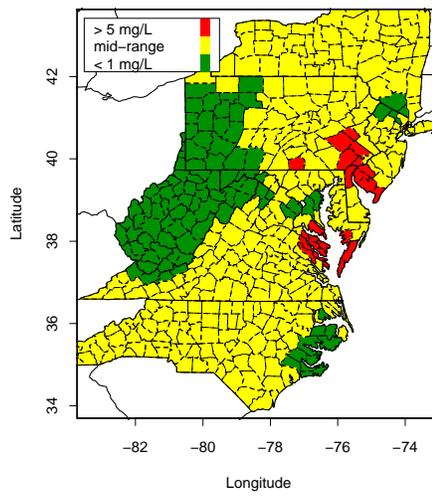
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5