Environmental Pollution 219 (2016) 1148–1155

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Bayesian Maximum Entropy space/time estimation of surface water chloride in Maryland using river distances☆

Prahlad Jat, Marc L. Serre*

Department of Environmental Science and Engineering, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599, United States

A R T I C L E I N F O

Article history:
Received 16 March 2016
Received in revised form
10 August 2016
Accepted 6 September 2016
Available online 9 September 2016

Keywords:
Chloride
Monitoring
River distance
Geostatistics

A B S T R A C T

Widespread contamination of surface water chloride is an emerging environmental concern. Consequently accurate and cost-effective methods are needed to estimate chloride along all river miles of potentially contaminated watersheds. Here we introduce a Bayesian Maximum Entropy (BME) space/time geostatistical estimation framework that uses river distances, and we compare it with Euclidean BME to estimate surface water chloride from 2005 to 2014 in the Gunpowder-Patapsco, Severn, and Patuxent subbasins in Maryland. River BME improves the cross-validation R^2 by 23.67% over Euclidean BME, and river BME maps are significantly different than Euclidean BME maps, indicating that it is important to use river BME maps to assess water quality impairment. The river BME maps of chloride concentration show wide contamination throughout Baltimore and Columbia-Ellicott cities, the disappearance of a clean buffer separating these two large urban areas, and the emergence of multiple localized pockets of contamination in surrounding areas. The number of impaired river miles increased by 0.55% per year in 2005–2009 and by 1.23% per year in 2011–2014, corresponding to a marked acceleration of the rate of impairment. Our results support the need for control measures and increased monitoring of unassessed river miles.

1. Introduction

Chloride contamination of rivers and its effect on the ecosystem health is a great environmental concern. During the winter snow, roads and sidewalks are treated with deicing salts. As the snow melts, more than 50 percent of the chloride in the deicing salt is transported to surface waters, leading to widespread effects on water chemistry. Road salt application practices and a variety of other processes lead to complex spatial and temporal patterns in chloride concentrations (Corsi et al., 2015).

Geostatistical methods provide potential for water quality assessment. Several studies have characterized surface water quality using spatial linear kriging methods (Peterson and Urquhart, 2006; Money et al., 2010). However, spatial kriging studies do not account for space/time autocorrelation and non-Gaussian 'soft' data (interval and censored data etc.). To address this issue, the Bayesian Maximum Entropy (BME) (Christakos, 1990; Christakos and Li, 1998) method is used here to estimate chloride concentration across space/time along a river network in Maryland. BME is a nonlinear estimation method that rigorously accounts for space/time variability and non-Gaussian soft data, and leads to kriging as its linear limiting case (Christakos, 1990; Christakos and Li, 1998; Christakos and Serre, 2000).

Peterson and Urquhart (2006) found that in Maryland the spatial autocorrelation of dissolved organic carbon (DOC) is better described using a covariance based on Euclidean distances rather than using a Weighted Asymmetric Hydrologic Distance (WAHD) covariance model, which is calculated based on the river distance (distance measured along the river network) and the proportion of flow shared between points (Peterson and Urquhart (2006), Money et al., 2009). Therefore, when considering other water quality parameters in Maryland, we expect that the Euclidean distance will better describe the spatial autocorrelation. However, their work did not report results for an autocorrelation using covariance based only on river distances (and not proportion of flow shared between points), unlike several other studies which successfully used river

☆ This paper has been recommended for acceptance by Charles Wong.

* Corresponding author. Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, 1303 Michael Hooker Research Center, Chapel Hill, NC 27599, United States.

E-mail address: marc_serre@unc.edu (M.L. Serre).

http://dx.doi.org/10.1016/j.envpol.2016.09.020
0269-7491/© 2016 Published by Elsevier Ltd.

distances in other river networks (Gardner et al., 2003; Ganio et al., 2009; Yang and Jin, 2010; Money et al., 2011; Chen et al., 2012). Hence, an important remaining question is whether the river distance works better than the Euclidean for the geostatistical estimation of chloride along rivers in Maryland.

The objectives of this study are therefore to introduce a framework for the BME space/time estimation of surface water chloride using river distances in three subbasins located in Maryland, to compare this method with alternate methods using Euclidean distances, to do a sensitivity analysis of methods used to deal with censored data, and to perform a space/time statistical estimation of chloride concentration along all river miles in our study area using the BME method based on river distances.

2. Materials and methods

2.1. Chloride and hydrography data

A total of 390 space/time chloride concentration values were obtained from the Maryland Biological Stream Survey (MBSS) dataset from 2005 to 2014 in stream waters located in the Gunpowder-Patapsco, Severn, and Patuxent subbasins (Fig. 1). The concentration values ranged from 1.5 mg/l to 3251.2 mg/l, with mean 93.69 mg/l and standard deviation 230.44 mg/l. Details on field sampling design, sampling methodology, and lab analysis procedures can be found elsewhere (Taylor-rogers, 1997).

The river network in our study area is described based on flow lines (Fig. 1) obtained from the USGS National Hydrography Data (USGS Hydrography data, 2015). The impervious surfaces are described based on the National Land Cover Database published by the Multi-Resolution Land Characteristics Consortium for the conterminous United States. Details about the NHD flowlines and impervious surface data are provided in the Supplementary Information (SI).

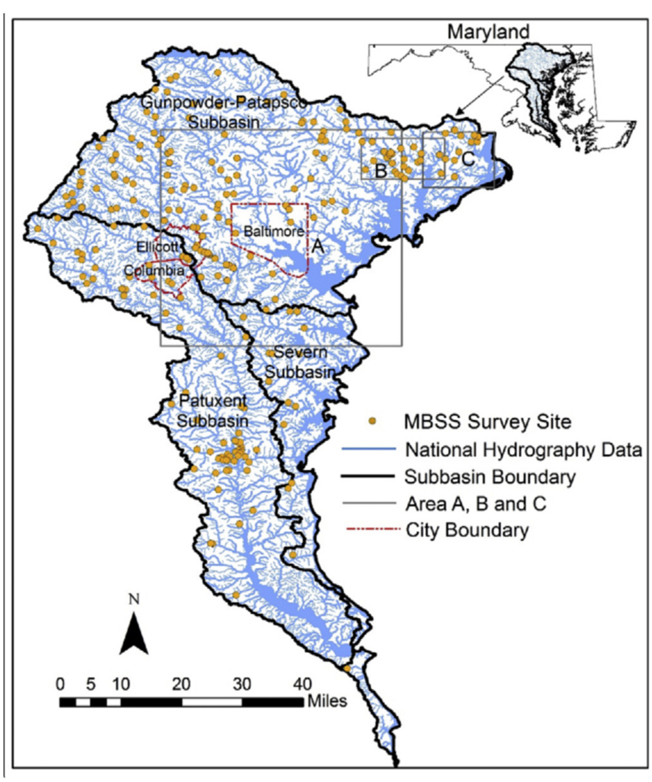

Fig. 1. The Maryland Biological Stream Survey (MBSS) sites in the Gunpowder-Patapsco, Patuxent, and Severn subbasins in Maryland. Baltimore, Ellicott, and Columbia are tree major cities in these subbasins.

2.2. Left-censored data

Left-censored chloride data correspond to data for which the true log-concentration is known only to be below a censoring limit (CL) of interest. Censoring data is a common practice when measured values are below the detection limit (DL) of an instrument. The BME approach has recently been shown to rigorously process left-censored data (Messier et al., 2012). Briefly, the maximum-likelihood estimation (MLE) method is used to estimate the mean (μ) and standard deviation (σ) of stream chloride concentrations by finding the μ and σ values that maximizes the MLE likelihood function (Helsel, 2005; Messier et al., 2012)

$$\mathscr{L}(z|\mu,\sigma) = \left\{ \prod_{z_i|z_i \geq CL_i} f_{\mu,\sigma}(z_i) \right\} * \left\{ \prod_{z_i|z_i \leq CL_i} F_{\mu,\sigma}(CL_i) \right\} \tag{1}$$

where $f_{\mu,\sigma}(z_i)$ denotes the normal probability distribution function (PDF) of observed chloride log-concentrations, z_i, with population mean (μ) and standard deviation (σ), and $F_{\mu,\sigma}(CL_i)$ denotes the CDF of the distribution taken at the log of the censoring limit (CL_i). The uncertainty associated with a left-censored data with CL_i is then fully characterized by the Truncated Gaussian PDF (TGPDF) obtained by truncating a Gaussian PDF above CL_i. The TGPDF(μ,σ,CL_i) has a mean$<\mu$ because of the truncation.

2.3. Space/time BME geostatistical framework for mapping analysis

BME, a space/time geostatistical estimation framework grounded in epistemic principles, reduces to the kriging methods as its linear limiting case. BME theory and its numerical implementation details are given elsewhere (Christakos, 1990; Christakos and Serre, 2000). Details about the application of BME to river networks are given elsewhere (Money et al., 2009).

Our notation to describe a space/time random field (S/TRF) will consist of denoting a single random variable Z in capital letters, its realization, z, in lower case; and vectors in bold faces (e.g., $\boldsymbol{z} = [z_1,\ldots,z_n]^T$). Let $\boldsymbol{z}_d$ be the vector of log-concentrations observed at locations $\boldsymbol{p}_d$, let $o_z(\boldsymbol{p})$ be an known offset function (Messier et al., 2015), where $\boldsymbol{p} = (\boldsymbol{s},t)$, $\boldsymbol{s}$ is the space coordinate and t is time, and let $\boldsymbol{x}_d = \boldsymbol{z}_d - o_z(\boldsymbol{p}_d)$ be the vector of offset removed log-concentrations. The suffix d in $\boldsymbol{p}_d$ is used to specify a location where data is available (i.e. a data point), whereas $\boldsymbol{p}$ without suffix d specify any location in the study domain. We define $X(\boldsymbol{p})$ as a homogenous/stationary S/TRF with realization $\boldsymbol{x}_d$, and we let

$$Z(\boldsymbol{p}) = X(\boldsymbol{p}) + o_z(\boldsymbol{p}). \tag{2}$$

be the S/TRF representing the distribution of stream chloride log-concentrations.

The total knowledge base K characterizing the S/TRF $X(\boldsymbol{p})$ can be divided in the general knowledge base (G-KB) and the site-specific knowledge base (S-KB). The G-KB describes general characteristics of the S/TRF including its mean $m_x(\boldsymbol{p}) = E[X(\boldsymbol{p})]$ and covariance function

$$c_X\left(\boldsymbol{p},\boldsymbol{p}'\right) = E\left[\left(X(\boldsymbol{p}) - m_x(\boldsymbol{p})\right)\left(X\left(\boldsymbol{p}'\right) - m_x\left(\boldsymbol{p}'\right)\right)\right], \tag{3}$$

where E[.] is the stochastic expectation operator. The S-KB refers to the sampling data $\boldsymbol{x}_d$, including both the hard (above detect) data $\boldsymbol{x}_h$ collected at $\boldsymbol{p}_h$, and the soft (left-censored) data $\boldsymbol{x}_s$ collected at $\boldsymbol{p}_s$ with an uncertainty expressed in terms of the PDF $f_s(\boldsymbol{x_s})$

(e.g. TGPDF(μ,σ,CL_i)).

We briefly describe here the main stages of the BME analysis used to estimate chloride log-concentration at unsampled locations $\boldsymbol{p}_k$ along the river network. At the prior stage, the $G-KB=\{E[X(\boldsymbol{p})], C_X(\boldsymbol{p},\boldsymbol{p}')\}$ is examined to obtain the prior PDF $f_G(.)$ describing the S/TRF $X(\boldsymbol{p})$ at mapping points of interest. At the integration stage, the prior PDF is updated using Bayesian epistemic conditionalization on $S-KB=\{\boldsymbol{x}_h, f_s(\boldsymbol{x}_s),\}$, leading to the BME posterior PDF

$$f_K(x_k)=A^{-1}\int d\boldsymbol{x}_s f_G\left(\boldsymbol{x}_h,\boldsymbol{x}_s,x_k\right)f_S\left(\boldsymbol{x}_s\right) \quad (4)$$

where x_k is a value of $X_k=X(\boldsymbol{p}_k)$, $f_G(\boldsymbol{x}_h,\boldsymbol{x}_s,x_k)$ is the multivariate Gaussian PDF for $(\boldsymbol{x}_h,\boldsymbol{x}_s,x_k)$ with mean and variance-covariance given by the G-KB, and $A=\int dx_k \int d\boldsymbol{x}_s f_G\ (\boldsymbol{x}_h,\boldsymbol{x}_s,x_k) f_S(\boldsymbol{x}_s)$ is a normalization coefficient. At the interpretive stage, the relation $Z_k=X_k+o_Z(\boldsymbol{p}_k)$ is used together with $f_K\ (x_k)$ to obtain the BME mean and variance log-concentration at the estimation points, which are then used to produce maps describing the estimated chloride log-concentration and associated estimation uncertainty at space/time locations of interest.

Several approaches exist to calculate an offset function $o_Z(\boldsymbol{p})$. In this work we use the approach described in Akita et al. (2007) and Money et al. (2009), where $o_Z(\boldsymbol{p})=o_Z(\boldsymbol{s},t)$ is the sum of a spatial component $o_{Z,s}(\boldsymbol{s})$ and a temporal component $o_{Z,t}(t)$ that are calculated using an exponential kernel smoothing of the time-averaged and spatially averaged data, respectively. Specifically, the spatial component at a given location $\boldsymbol{s}$ is given by

$$o_{Z,s}\ (\boldsymbol{s})=\sum_i w\ (\boldsymbol{s},\boldsymbol{s}_i)\ \overline{z(\boldsymbol{s}_i)} \quad (5)$$

where $\overline{z(\boldsymbol{s}_i)}$ is the time-averaged log-concentration at location $\boldsymbol{s}_i$, $w(\boldsymbol{s},\boldsymbol{s}_i)$ is an exponential kernel weight given by

$$w\ (\boldsymbol{s},\boldsymbol{s}_i)=B^{-1}\exp\left(\frac{-3d(\boldsymbol{s},\boldsymbol{s}_i)}{k_r}\right), \quad (6)$$

$d(\boldsymbol{s},\boldsymbol{s}_i)$ is the distance between $\boldsymbol{s}$ and $\boldsymbol{s}_i$, k_r is the spatial exponential smoothing range, and $B=\sum_i \exp\left(\frac{-3d(\boldsymbol{s},\boldsymbol{s}_i)}{k_r}\right)$ is a normalization coefficient calculated so that the sum of weights equals 1. In the previous water quality studies (Akita et al., 2007; Money et al., 2009) the distance $d(\boldsymbol{s},\boldsymbol{s}_i)$ in Eq. (6) is based on an Euclidean metric. In this work we extend past works by calculating that distance based on either an Euclidean or a river distance metric, i.e.

$$d\left(\boldsymbol{s},\boldsymbol{s}'\right)=\begin{pmatrix} d_E\left(\boldsymbol{s},\boldsymbol{s}'\right) & \text{Euclidean distance} \\ d_R\left(\boldsymbol{s},\boldsymbol{s}'\right) & \text{River distance}\end{pmatrix} \quad (7)$$

To the best of our knowledge, this is the first study implementing an offset calculated using a kernel smoothing based on a river metric, hence the river offset presented here is novel. Note that the calculation of the temporal component $o_{Z,t}(t)$ is done as described in Akita et al. (2007), i.e. by replacing the spatial distance in Eq (6) with the corresponding time difference. As shown in the SI, the offset function described here captures well the broad spatial and temporal trends in chloride log-concentrations, indicating that this offset function is suitable in this study area.

Alternatively, the offset function can be calculated using a Land Use Regression (LUR) as described in Messier et al. (2012), and Reyes and Serre (2014), where the LUR uses land imperviousness as a predictor, since it has been found to be a predictor of stream water quality degradation (Brabec et al., 2002; King et al., 2005).

The $c_x(\boldsymbol{p},\boldsymbol{p}')$ function describing the covariance of the homogeneous/stationary S/TRF $X(\boldsymbol{p})$ can be expressed as an exponential function of the spatial distance and time difference between space/time points $\boldsymbol{p}=(\boldsymbol{s},t)$ and $\boldsymbol{p}'=(\boldsymbol{s}',t')$, i.e.

$$c_x\left((\boldsymbol{s},t),\left(\boldsymbol{s}',t'\right)\right)=c_0\exp\left(-3\frac{d(\boldsymbol{s},\boldsymbol{s}')}{a_r}\right)\exp\left(-3\frac{|t-t'|}{a_t}\right) \quad (8)$$

where c_0, a_r and a_t are the variance, spatial covariance range, and temporal covariance range, respectively, of the S/TRF $X(\boldsymbol{p})$, and $d(\boldsymbol{s},\boldsymbol{s}')$ can again be either the Euclidean or river distance (equation (7)). In this work we choose an exponential covariance model because it has been shown to be permissible for any river networks (Ver Hoef et al., 2006; Peterson and Urquhart, 2006; Money et al., 2009) and to our knowledge no other covariance model has been shown to fulfill that same property.

To quantify the impact of using either the Euclidean or river distance (eq. (7)) in the offset (eq. (6)) and covariance (eq. (8)), we implement all combinations of offset and covariance models (i.e. Euclidean offset/Euclidean covariance, Euclidean offset/River covariance, River offset/Euclidean covariance, and River offset/River covariance models) and we compare their mapping accuracy.

Another alternative for the covariance model is using a WAHD covariance model (Peterson and Urquhart, 2006; Money et al., 2009), however, we excluded it from detailed analysis because we found it has a lower mapping accuracy than the Euclidean covariance model, which is consistent with what Peterson and Urquhart (2006) found for DOC using the MBSS data.

2.4. Comparison of BME using river versus Euclidean distance

The DL for our MBSS chloride data is very low (0.01 mg/l), and all 390 measured values are above DL. In that case the BME method treats all the data as hard, and no soft data are used. In this baseline case the effect of using a river versus Euclidean distance in the BME estimation method was assessed by performing a leave-one-out cross-validation (LOOCV) whereby each chloride log-concentration value z_j was removed one at a time, and re-estimated using only the remaining data. For a given estimation method (m) that uses either the river or Euclidean distance, the overall estimation error was quantified using the Mean Squared Error, $MSE^{(m)}=\frac{1}{n}\sum_{j=1}^{n}(z_j^{*(m)}-z_j)^2$, the consistent estimation error (i.e. the bias) was quantified using the Mean Error $ME^{(m)}=\frac{1}{n}\sum_{j=1}^{n}(z_j^{*(m)}-z_j)$, and the random error (i.e. lack of precision) was quantified using the squared Pearson coefficient, $R^2=1-\sum_{j=1}^{n}(z_j^{*(m)}-z_j)^2/\sum_{j=1}^{n}(z_j^{*(m)})^2$, where $z_j^{*(m)}$ is the re-estimation of z_j. This cross validation analysis was used to quantify the gain in mapping accuracy when the Euclidean distance is replaced with the river distance in the covariance model, and then in the offset model. This results in four baseline approaches (Euclidean offset/Euclidean covariance, Euclidean offset/river covariance, river offset/Euclidean covariance, and river offset/river covariance) which are all mathematically permissible regardless of their physical meaningfulness.

2.5. Sensitivity analysis with respect to the proportion of left censored data

Methods are needed to deal with situations where there is a large proportion of left censored data. This can happen for cost effectiveness purposes when low-cost data is used (LoBuglio et al., 2007), or when measuring toxic compounds that are difficult to detect.

The usual approaches used to deal with left censored data have been to delete them, or to fabricate numbers for them (equal to half of the CL, or equal to the CL), which are flawed approaches that can introduce a strong bias in mean and standard deviation (Singh and Nocerino, 2002).

On the other hand, the BME approach has recently been shown to rigorously process left-censored data (Messier et al., 2012). However few studies have investigated the loss of accuracy associated with left-censored data (Helsel, 2005; Messier et al., 2012), and this study provides a unique opportunity to do that. As stated earlier, all 390 measured values are above the DL, which provided us an opportunity to investigate the sensitivity of the loss in mapping accuracy with respect to the proportion of censored data. This sensitivity analysis consisted in left censoring a proportion of the data, and comparing the cross validation statistics of the following three methods: (a) BME rigorously modeling the censored data using the TGPDF, (b) kriging replacing the censored data with half the CL, and (c) kriging replacing the censored data with the CL. Comparison of the loss in the mapping accuracy of these three methods revealed whether BME (methods a) better handles left-censored data than its kriging limiting cases (methods b and c).

2.6. Assessment of impaired river miles

The space/time distribution of chloride is governed by complex natural and physical processes. Imperfect knowledge about these complex processes result in a significant uncertainty in chloride estimation. Not accounting for estimation uncertainty in impairment assessment may lead to a wrong conclusion and hence accounting for uncertainty is considered to be an essential aspect of any decision making framework. Our river BME method is a geostatistical approach and as such its advantage is that it provides not only concentration estimates but also the probability that chloride exceeds a specific regulatory level. Using river BME, we calculated the probability that chloride exceeds the EPA guideline level of 230 mg/l along each of the 6018 river miles in the study area from 2005 to 2014, and we classified a given river reach as impaired if the average probability of exceedance of the EPA guideline level along that river reach is greater than 90%, as non-assessed if that probability is between 10% and 90%, and clean if that probability is less than 10%. The average probability of exceedance along a river reach is calculated as the arithmetic average of the probability of exceedance calculated at equidistant points along that river reach.

3. Results and discussion

3.1. Covariance models of offset-removed chloride log-concentrations

Details about LUR analysis ($R = 0.6$), the three offset models (Euclidean, river and LUR), and the weighted least square covariance fitting procedure used to obtain the covariance parameters for each offset model are available in the SI. The sill (i.e. variance) and the spatial covariance range for the Euclidean offset removed chloride log-concentrations are $c_o = 0.41$ (log-mg/l)^2 and $a_r = 19$ km (across land) for the Euclidean covariance model, and $c_o = 0.41$ (log-mg/l)^2 and $a_r = 28$ km (along rivers) for the river covariance model. For the river offset removed chloride log-concentrations, $c_o = 0.25$ (log-mg/l)^2 and $a_r = 28$ km (across land) for the Euclidean covariance model, and $c_o = 0.25$ (log-mg/l)^2 and $a_r = 36$ km (along rivers) for the river covariance model. For the LUR offset removed chloride log-concentrations, $c_o = 0.61$ (log-mg/l)^2 and $a_r = 58$ km (across land) for the Euclidean covariance model, and $c_o = 0.61$ (log-mg/l)^2 and $a_r = 96$ km (along rivers) for the river covariance model. The temporal range is $a_t = 12$ years for all covariance models.

3.2. Cross-validation results contrasting the Euclidean versus river covariance models

The cross validation results (Table 1) obtained in the baseline case (where none of the 390 values are censored) show that using an Euclidean offset (first row of Table 1), space/time BME using a river covariance better predicts chloride ($R^2 = 0.711$) than when using an Euclidean covariance ($R^2 = 0.638$), corresponding to an 11.44% percent change (PC) in R^2. This work is the first to demonstrate that the river covariance model is better than the Euclidean covariance model for chloride estimation in these subbasins. This means that the autocorrelation of chloride is best described using distances measured along the river network, which indicates that processes that are distributed along river networks (e.g. highways -a known source of chloride, vegetation buffers -a known attenuation process, etc.), are important drivers of the distribution of chloride along rivers.

3.3. Cross-validation results contrasting Euclidean versus river offsets

Since we conclude in the baseline case that the covariance should be based on the river distance rather than the Euclidean distance, then the next question is whether the offset should also be calculated based on the river distance rather than the Euclidean distance. To answer that question we implemented space/time BME using our novel river offset (second row of Table 1). The only difference between the first and second row of Table 1 is the introduction of the river offset, and by comparing these two rows we find that the river offset consistently outperforms the Euclidean offset. For example when using a river covariance (second column of Table 1), space/time BME using the river offset better predicts chloride ($R^2 = 0.789$) than when using the Euclidean offset ($R^2 = 0.711$), corresponding to a 10.97% PC in R^2. Our work is the first to introduce the river offset and to demonstrate that it leads to an appreciable improvement over the Euclidean offset used in previous works(Akita et al., 2007; Money et al., 2011). The implication of this finding is that the river network topology should be taken into account for both the offset and covariance models. Doing so results in an overall PC in R^2 of 23.67%, which considerably

Table 1
Leave-one-out cross-validation statistics obtained using the BME method with different offset and covariance models for the estimation of chloride log-concentration.[a]

	Euclidean covariance			River covariance		
	MSE (log-mg/l)^2	ME (log-mg/l)	R^2 (unitless)	MSE (log-mg/l)^2	ME (log-mg/l)	R^2 (unitless)
Euclidean offset	0.343	0.002	0.638	0.264	0.002	0.711
River offset	0.224	0.003	0.760	0.194	0.018	0.789

[a] The Euclidean covariance and river covariance models use the Euclidean and river distance metrics, respectively. The Euclidean offset and the river offset use the Euclidean and river distance metrics, respectively; MSE is the mean squared error; ME is the mean error; R^2 is the squared coefficient of determination between observed and estimated values.

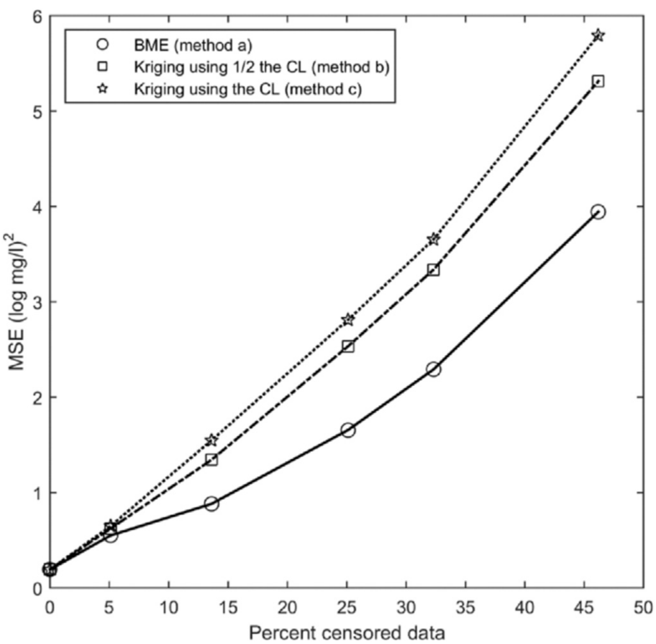

Fig. 2. Cross validation MSE for river BME and its kriging linear limiting cases shown with respect to the proportion of censored data. BME (method a) rigorously models the uncertainty in the censored data using the TGPDF, while kriging treats them as data with no uncertainty by simply replacing them with half of the CL (method b) or by the CL (method c).

improves our ability to accurately predict chloride across space and time.

3.4. Sensitivity analysis results with respect to censoring limit

To assess sensitivity analysis of the estimation accuracy of the river BME and kriging methods with respect the proportion of censored data, we performed a cross validation analysis for 6 different proportions of censored data ranging from 0% (baseline case) to 46.2% of the overall data (Fig. 2). Each censored dataset was generated by selecting a CL, censoring all values below the CL and only providing the CL value. River BME rigorously models the uncertainty contained in censored data using the TGPDF, while the kriging methods simply treat them as data without any uncertainty since these data are replaced with half the CL, or with the CL. As expected, the estimation accuracy degrades with increasing proportion of censored data. However Fig. 2 clearly demonstrates that the rate of deterioration in estimation accuracy is lower for river BME (method a) than for its kriging linear limiting cases (method b and c). This trend can also be seen from the cross validation R^2 which indicates that BME improves the R^2 by a factor of about 2–7.5 over kriging (with censored data replaced by half the CL) when the proportion of censored data ranges from 13.6% to 46.2% (see SI for more details). Overall these results indicate that when a dataset includes censored data, then the BME method used in this work is consistently more efficient than the kriging method at extracting the information contained in these censored data.

3.5. Cross validation results contrasting the river and LUR offsets

The LUR offset is obtained based on the average imperviousness in HUC12 subwatersheds, which is a weak predictor of chloride in our study area (R = 0.6, see SI for more details). LUR is an integral part of many water quality models and is an attractive method because it takes advantage of seemingly free data (e.g. imperviousness calculated for other purposes), but in practice its implementation require dedicated modelers to preprocess these data, which can be time consuming for local regulatory agencies. The cross-validation statistics MSE increases from 0.194 (log-mg/l)^2 for the river offset BME method to 0.313 (log-mg/l)^2 for the LUR offset BME method, and the corresponding R^2 drops from 0.789 to 0.660. These cross-validation statistics indicates that using a LUR offset fails to produce better results than using the river offset presented in this work, however LUR models with river buffers and temporally varying imperviousness maps may improve the LUR based approach.

3.6. Difference in the maps produced using Euclidean versus river BME

To the best of our knowledge, previous studies have not compared, and quantified, the difference in estimated levels obtained using an Euclidean versus river BME methods in that situation. To address this question, we provide here a comparison of the Euclidean versus river BME maps in area B and area C (Fig. 1 depicts where areas B and C are located). The purpose of this comparison is purely to emphasize the difference in chloride estimates using Euclidean versus river BME along unsampled river reaches. These maps are not meant to compare the estimation accuracy of the Euclidean and river BME methods at unsampled locations.

The Euclidean BME and river BME maps for area B are shown in Fig. 3(a) and (b), respectively. In that area we are interested in the assessment of Bynum Run, which lacks monitoring data, and runs parallel to Winters Run where monitoring data are available. Fig. 3(a) and (b) show that in this area major highways (a known source of chloride) are aligned along the river network. The river distance between the monitoring stations on Winter Run and estimation points on Bynum Run are long, resulting in a low autocorrelation in chloride measurements. The situation for the Euclidean BME model is the converse, the estimated values along Bynum Run are strongly affected by what's measured in Winters Run. Fig. 3(a) and (b) show this difference in estimated chloride, and reveal that the chloride levels along Bynum Run are substantially higher in the Euclidean BME map (Fig. 3(a)) than in the river BME map (Fig. 3(b)). To quantify this difference, we calculate the number of river miles with estimates exceeding two thresholds of interest: 230 mg/l (an ambient water quality criteria for chloride defined by the U.S. EPA (U.S. Environmental Protection Agency. Ambient water quality criteria for chloride, 1988)), and 145 mg/l (a concentration level at which declines in survival of salamanders have been documented (Stranko et al., 2013)). We find that according to Euclidean BME, 14% of Bynum Run river miles North of US 40 exceed 230 mg/l, and 62% of these river miles exceed 145 mg/l, whereas none of these river miles exceed either threshold limits according to river BME.

Similarly, the river BME estimates along the Grays and Cranberry Runs (Fig. 3(d)) are low as opposed to the high chloride estimates obtained with Euclidean BME (Fig. 3(c)). According to Euclidean BME, 9% of river miles along the Grays and Cranberry Runs exceed 230 mg/l, and 52% of these river miles exceed 145 mg/l, while none of these river miles exceed either threshold limits according to river BME.

These results demonstrate that there can be big differences in the estimated chloride concentration using Euclidean BME and river BME, which may lead to substantial differences in the assessment of whether a river reach is impaired. For example using the Euclidean approach one might conclude that Bynum Run and the Grays and Cranberry Runs are in need of remedial action, while using the river approach one might conclude that remedial action is less needed and added monitoring is desired. The implication of

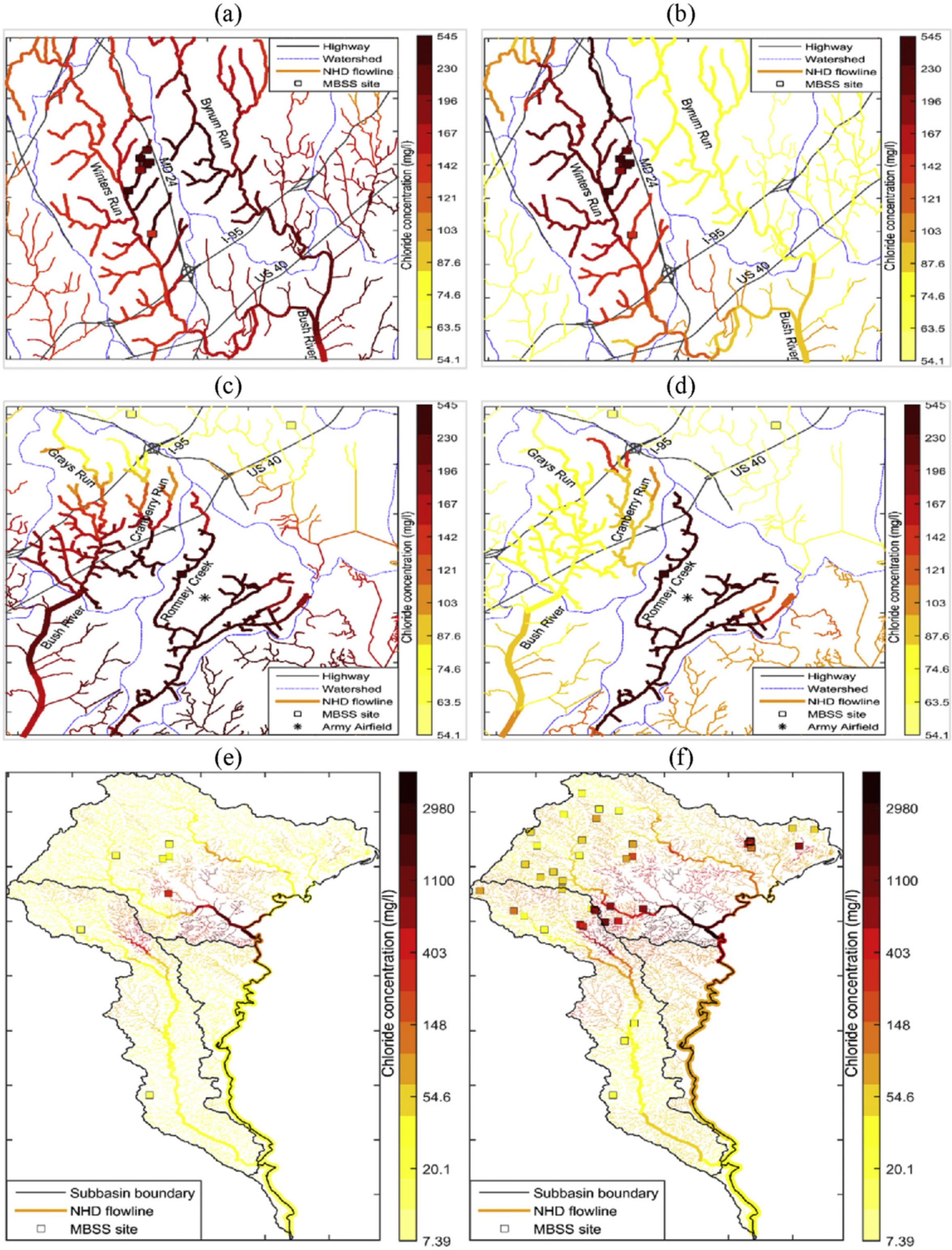

Fig. 3. Maps of the BME mean estimate of chloride concentrations in 2014. The maps on the left panels are estimated using Euclidean BME, the maps on the right are estimated using river BME. Panel (a), (c) and (e) show the Euclidean BME estimate of chloride in area B, area C, and the study domain, respectively. The corresponding river BME maps are in the Panels (b), (d) and (f), respectively. The flow lines in panels (a), (b), (c), and (d) are highlighted (increased width) for better visual appearances of segments compared for estimation accuracy. The width of the flow lines in panels (e) and (f) correspond to their cumulative river miles.

this finding is that using the proper approach does matter, and therefore one should use the river BME approach introduced in this work rather than the classical Euclidean approach when estimating chloride along unmonitored river miles. Another implication of this finding is that using river BME, one will delineate impaired areas that are confined along river reaches, as opposed to spread isotropically across land, which may be easier to remediate because resources will be targeted to a specific subwatershed, rather than spread across multiple subwatersheds.

3.7. Space/time patterns in chloride contamination

The rate of urban development, changes in road salt application practices, and changing climate conditions may drive a variety of spatial and temporal patterns in chloride concentrations (Corsi et al., 2015). Accurate estimation of chloride is crucial to understand these patterns, to improve our understanding of the extent and nature of chloride contamination, and to design effective measures to control the chloride pollution. A series of chloride concentration maps from 2005 to 2014 are constructed using the space/time river BME method introduced in this study. The maps obtained for 2014 are shown in Fig. 3, while maps for other years are in SI. These maps provide the first representation of chloride distribution that fully integrates information about space/time variability and river network topology.

In the study area, the high population density area is made up of Baltimore and Columbia-Ellicott cities, which have a high concentration of impervious surfaces and are separated by a narrow green buffer along the Patapsco River. Conversely the surrounding area is generally green with localized concentrations of impervious surfaces where small towns are located.

Our river BME maps of chloride concentrations reveal that there are two distinct cores of chloride contamination corresponding to Baltimore and Columbia-Ellicott cities, which are persistently contaminated from 2005 to 2014. This indicates that once an area is contaminated it remains contaminated for a long time, which is consistent with what has been reported in previous studies (Harte and Trowbridge, 2010; Perera et al., 2013) These two core areas are initially separated by a clean buffer along the Patapsco River. This buffer is revealed by the river BME estimation method as it accounts for river network topology. These two core areas are expanding outwards at a low rate during 2005–2009, resulting in a narrowing and eventual loss of the green buffer separating Baltimore and Columbia-Ellicott cities. There is a stagnation in 2010 and 2011, followed by an accelerated rate of outward expansion of the two core areas during 2012–2014 up until they coalesce in 2014, resulting in significant contamination over the whole Baltimore-Columbia-Ellicott urban area. Major factors for this significant urban-wide contamination may include increased rate of salt application, as well as the loss of green buffer separating Baltimore and Columbia-Ellicott cities.

Our river BME maps further reveal that at the beginning of the study period (2005) the concentration of chloride is low or inexistent in the streams located outside of the Baltimore-Columbia-Ellicott urban area. However in that area several pockets of high chloride concentration emerge in 2005–2009 and remain contaminated till the end of the study period (2014). Each of these pockets can be visually detected using river BME because they are confined along distinct river branches, whereas it is more difficult to see them when using an Euclidean approach that averages out concentration across river branches. These pockets of contamination illustrate the usefulness of river BME to identify such areas so that they can be targeted for monitoring.

3.8. Probabilistic assessment of impaired river miles

The probabilistic assessment of impaired river miles indicates that there are two distinguishable time periods (2005–2009 and 2011–2014) during which the fraction of unassessed and impaired river miles increased (Fig. 4). In the first time period the impaired river miles increased from 1.3% in 2005 to 3.5% in 2009, corresponding to a 0.55% rate of increase in impaired river miles per year. In second time period, the impaired river miles increased from 2.3% to 6%, corresponding to a 1.23% rate of increase in impaired river miles per year. These results demonstrate that there is a marked acceleration of the impairment of the study area, with a greater than two fold increase in the rate at which river miles become impaired. As stated earlier mechanisms causing this acceleration of impairment include the loss of buffer along the Patapsco River, the coalescence of core impaired areas, and the increased rate of chloride application. The implication of this finding is that there is sufficient evidence of increased impairment to justify taking strong measures to control chloride applications in these watersheds.

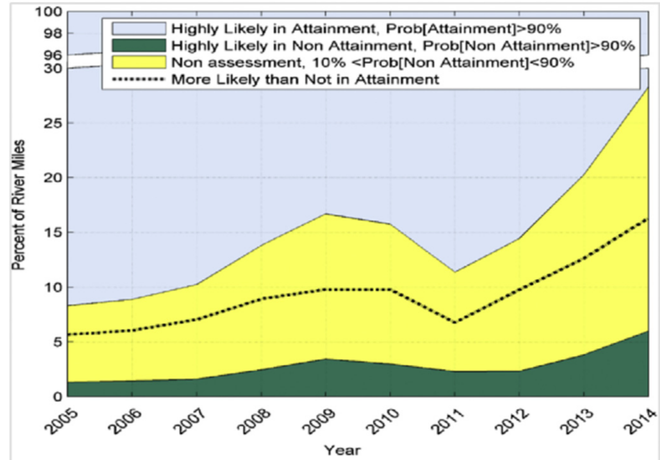

Fig. 4. Time series of average fraction of river miles in Gunpowder-Patapsco, Patuxent, and Severn subbasins in Maryland that are highly likely in non-attainment (the probability of exceedance of the EPA guideline (230 mg/l) is greater than 90%), non-assessed (probability between 10% and 90%), and highly likely in attainment (probability less than 10%) from 2005 to 2014. See Supplementary Information for maps showing for each year from 2005 to 2014 the spatial distribution of the probability that chloride exceeds 230 (mg/l).

Interestingly, there is an even stronger acceleration in the unassessed river miles. There is a 1.05% and 3.17% rate of increase in unassessed river miles per year during the 2005–2009 and 2011–2014 periods, respectively. This dramatic acceleration of the rate of increase of unassessed river miles indicates that the monitoring effort, which in 2005 was sufficient to differentiate between clean and impaired river miles, is becoming insufficient to fulfill its task, and increased monitoring is needed while chloride levels are rising. Hence the overall finding of our work is that there is an urgent need for increased monitoring in areas where chloride is unassessed, and these unassessed areas can efficiently be identified using the river BME approach.

4. Conclusions

This work is making an important methodological contribution for the assessment of water quality along rivers. It consists in the introduction of a river kernel smoothing function used to capture large distance scale variability in water quality. We find that when combined with geostatistical estimation of offset-removed concentrations, the river kernel smoothing is more accurate than earlier approaches that used Euclidean kernel smoothing.

This is because river kernel smoothing better captures river topology than Euclidean kernel smoothing. To our knowledge, this work is the first to perform a mapping analysis using the river kernel smoothing described here in a river geostatistical framework, and to demonstrate that it substantially improves mapping accuracy over an Euclidean approach. This approach is a contribution to the field of river geostatistics, and will be applicable to the estimation a wide range of river water quality parameters.

Another important contribution is our analysis of the mapping efficiency of the BME method of modern geostatistics when dealing with dataset with left censored data, as is the case when measurements are below the DL. We demonstrate that when a proportion of data is left censored, then BME always outperforms its kriging linear limiting case. This is a widely applicable finding of our work because there are many instances where environmental agencies have to measure trace level toxic constituents that have concentrations less than the DL of the measuring instruments. In such cases we recommend that these agencies use the full non-linear and non-Gaussian BME approach rather than arbitrarily setting the left censored data to half the CL or to the CL value.

Turning to the analysis of river chloride in Maryland, we find that there are big differences in the estimated chloride concentration using Euclidean BME versus river BME, particularly along unmonitored river reaches that run parallel to a river reach with monitoring data. We demonstrate that the differences in estimated chloride concentrations lead to substantial differences in the assessment of whether a river reach is impaired. Hence, an appropriate estimation method is important as estimates change the outcome of regulatory or policy decisions and the remediation strategy selected.

Using the river BME approach we find that chloride contamination in Maryland is characterized by wide contamination throughout Baltimore and Columbia-Ellicott cities, the disappearance of a clean buffer separating these two large urban areas, and the emergence of multiple localized pockets of contamination in surrounding areas. The number of impaired river miles increased by 0.55% per year in 2005–2009 and by 1.23% per year in 2011–2014, corresponding to a marked acceleration of the rate of impairment that justify taking strong measures to control chloride applications in these watersheds. We also find that the number of unassessed river miles has increased even more drastically over these periods, indicating the need of increased monitoring required as large clean areas become fragmented with pockets with persistently high chloride concentration. These unassessed pockets areas can efficiently be identified using the river BME approach for optimal sampling design for targeted monitoring. Since the river BME approach accounts for river network topology, the areas identified as unassessed are confined along specific river reaches, which will make regulatory effort more targeted and efficient.

Acknowledgements

This work was supported in part by grant number P42ES005948 of the National Institute of Environmental Health Sciences, and by NSF grant 1316318 as part of the joint NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envpol.2016.09.020.

References

USGS Hydrography data [WWW Document], 2015. http://nhd.usgs.gov/data.html (accessed 10.6.15.).

Akita, Y., Carter, G., Serre, M.L., 2007. Spatiotemporal nonattainment assessment of surface water tetrachloroethylene in New Jersey. J. Environ. Qual. 36, 508–520. http://dx.doi.org/10.2134/jeq2005.0426.

Brabec, E., Schulte, S., Richards, P.L., 2002. Impervious surfaces and water quality: a review of current literature and its implications for watershed planning. J. Plan. Lit. 16, 499–514. http://dx.doi.org/10.1177/088541202400903563.

Chen, Y.C., Yeh, H.C., Wei, C., 2012. Estimation of river pollution index in a tidal stream using kriging analysis. Int. J. Environ. Res. Public Health 9, 3085–3100. http://dx.doi.org/10.3390/ijerph9093085.

Christakos, G., 1990. A Bayesian/maximum-entropy view to the spatial estimation problem. Math. Geol. 22, 763–777. http://dx.doi.org/10.1007/BF00890661.

Christakos, G., Li, X., 1998. Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators? Math. Geol. 30, 435–462. http://dx.doi.org/10.1023/A:1021748324917.

Christakos, G., Serre, M.L., 2000. BME analysis of spatiotemporal particulate matter distributions in North Carolina. Atmos. Environ. 34, 3393–3406. http://dx.doi.org/10.1016/S1352-2310(00)00080-7.

Corsi, S.R., De Cicco, L.A., Lutz, M.A., Hirsch, R.M., 2015. River chloride trends in snow-affected urban watersheds: increasing concentrations outpace urban growth rate and are common among all seasons. Sci. Total Environ. 508, 488–497. http://dx.doi.org/10.1016/j.scitotenv.2014.12.012.

Ganio, L.M., Torgersen, C.E., Gresswell, R.E., 2009. Geostatistical Approach in Stream Pattern for Describing Networks 3, pp. 138–144. http://dx.doi.org/10.1890/1540-9295(2005)003[0138:AGAFDS]2.0.CO;2.

Gardner, B., Sullivan, P.J., Lembo Jr., A.J., 2003. Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. Can. J. Fish. Aquat. Sci. 60, 344–351. http://dx.doi.org/10.1139/f03-025.

Harte, P.T., Trowbridge, P.R., 2010. Mapping of road-salt-contaminated groundwater discharge and estimation of chloride load to a small stream in southern New Hampshire, USA. Hydrol. Process 24, 2349–2368. http://dx.doi.org/10.1002/hyp.7645.

Helsel, D.R., 2005. Nondetects and data analysis. Statistics for censored environmental data.

King, R.S., Baker, M.E., Whigham, D.F., Weller, D.E., Jordan, T.E., Kazyak, P.F., Hurd, M.K., 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. Ecol. Appl. 15, 137–153.

LoBuglio, J.N., Characklis, G.W., Serre, M.L., 2007. Cost-effective water quality assessment through the integration of monitoring data and modeling results. Water Resour. Res. 43, 1–16. http://dx.doi.org/10.1029/2006WR005020.

Messier, K.P., Akita, Y., Serre, M.L., 2012. Integrating address geocoding, land use regression, and spatiotemporal geostatistical estimation for groundwater tetrachloroethylene. Environ. Sci. Technol. 46, 2772–2780. http://dx.doi.org/10.1021/es203152a.

Messier, K.P., Campbell, T., Bradley, P.J., Serre, M.L., 2015. Estimation of groundwater radon in North Carolina using land use regression and Bayesian maximum entropy. Environ. Sci. Technol. 49, 9817–9825. http://dx.doi.org/10.1021/acs.est.5b01503.

Money, E.S., Carter, G.P., Serre, M.L., 2009. Modern space/time geostatistics using river distances: data integration of turbidity and E. coli measurements to assess fecal contamination along the Raritan River in New Jersey. Environ. Sci. Technol. 43, 3736–3742. http://dx.doi.org/10.1021/es803236j.

Money, E., Carter, G.P., Serre, M.L., 2010. NIH public access. Water 43, 1948–1958. http://dx.doi.org/10.1016/j.watres.2009.01.034.Using.

Money, E.S., Sackett, D.K., Aday, D.D., Serre, M.L., 2011. Using river distance and existing hydrography data can improve the geostatistical estimation of fish tissue mercury at unsampled locations. Environ. Sci. Technol. 45, 7746–7753. http://dx.doi.org/10.1021/es2003827.

Perera, N., Gharabaghi, B., Howard, K., 2013. Groundwater chloride response in the highland creek watershed due to road salt application: a re-assessment after 20 years. J. Hydrol. 479, 159–168. http://dx.doi.org/10.1016/j.jhydrol.2012.11.057.

Peterson, E.E., Urquhart, S.N., 2006. Predicting water quality impaired stream segments using landscape-scale data and a regional geostatistical model: a case study in Maryland. Environ. Monit. Assess. 121, 613–636. http://dx.doi.org/10.1007/s10661-005-9163-8.

Reyes, J.M., Serre, M.L., 2014. An LUR/BME framework to estimate PM2.5 explained by on road mobile and stationary sources. Environ. Sci. Technol. 48, 1736–1744. http://dx.doi.org/10.1021/es4040528.

Singh, A., Nocerino, J., 2002. Robust estimation of mean and variance using environmental data sets with below detection limit observations. Chemom. Intell. Lab. Syst. 60, 69–86. http://dx.doi.org/10.1016/S0169-7439(01)00186-1.

Stranko, S., Bourquin, R., Zimmerman, J., Kashiwagi, M., McGinty, M., Kauda, R., 2013. Do Road Salts Cause Environmental Impacts? Maryl. DNR.

Taylor-rogers, S., 1997. State of the Streams.

U.S. Environmental Protection Agency, 1988. Ambient Water Quality Criteria for Chloride. Washington D.C. EPA 440/5-88-001.

Ver Hoef, J.M., Peterson, E., Theobald, D., 2006. Spatial statistical models that use flow and stream distance. Environ. Ecol. Stat. 13, 449–464. http://dx.doi.org/10.1007/s10651-006-0022-8.

Yang, X., Jin, W., 2010. GIS-based spatial regression and prediction of water quality in river networks: a case study in Iowa. J. Environ. Manag. 91, 1943–1951. http://dx.doi.org/10.1016/j.jenvman.2010.04.011.