Water Resources Research Institute
of The University of North Carolina

Report No. 444

SPACE/TIME GEOSTATISTICAL ESTIMATION OF NITRATE AND RADON GROUNDWATER CONTAMINANTS

By
Mark L. Serre[1], Yasuyuki Akita[1], Kyle Messier[1], Evan O. Kane[2], Rick Bolich[2], and Ted Campbell[2]

[1]Department of Environmental Sciences and Engineering
University of North Carolina at Chapel Hill
Chapel Hill, NC

[2]Division of Water Resources
North Carolina Department of Environment and Natural Resources

UNC-WRRI-444

The research on which this report is based was supported by funds provided by the North Carolina General Assembly and/or the US Geological Survey through the NC Water Resources Research Institute.

This report fulfills the requirements for a project completion report of the Water Resources Research Institute of The University of North Carolina. The authors are solely responsible for the content and completeness of the report.

WRRI Project No. 11-05-W
October 2013

Contents

i. List of Tables

ii. List of Figures

iii. Abstract: Space/time Geostatistical Estimation of Nitrate and Radon Groundwater Contaminants

Objectives: 1) Increase the database for groundwater nitrate and radon. 2) Develop a land use regression model for point level nitrate and radon. 3) Integrate the land use regression into a Kriging model of groundwater nitrate and radon, and demonstrate its improvement over land use regression and Kriging methods alone.

Methods: We first coordinate with our collaborators in the North Carolina Department of Environment and Natural Resources (NC DENR) and obtain the best monitoring data for groundwater nitrate and radon. We then use a multi-stage geocoding process to geocode private well and private well samples for nitrate and radon, respectively.

Explanatory variables are created for nitrate based on point and non-point source data. For point sources including the treated sewage sludge field application sites, we create variables as the sum of exponentially decaying contributions. For non-point source data we create variables as the percent of a land cover or geology feature or age within a circular buffer. Explanatory variables for radon variables consist solely of the percent of geological feature or age within a circular buffer.

We then use a variation on stepwise regression referred to as the Stepwise Hyperperameter Optimization and REgression (SHORE) procedure to select the best land use regression (LUR) model for nitrate and radon. Then the LUR model is integrated into the Kriging method geostatistics as a global offset.

Results: The LUR model for nitrate had an r-squared of 0.18 with 6 explanatory variables. The LUR model for radon had an r-squared of 0.26 with 5 explanatory variables. By integrating the LUR model into the Kriging method for nitrate, we changed the cross-validation mean squared error (MSE) over Kriging alone by -5.2%, resulting in a reduced MSE of 0.267 $(\text{log-mg/L})^2$. After integrating the LUR model into the Kriging method for radon, we changed the MSE over LUR alone (which was better than Kriging alone) by -15%, resulting in a reduced MSE of 1.49$(\text{log-pCi/L})^2$.We created the first point level estimate maps of groundwater nitrate and radon across North Carolina.

Conclusions: The SHORE procedure created a land use regression model for point level groundwater nitrate and radon that when integrated into the Kriging methodology improved our ability to predict point level groundwater nitrate and radon across North Carolina to date.

Recommendations: We recommend that our nitrate map be used to determine areas where private well owners may be at risk of being exposed to elevated groundwater levels, and to determine streams that may be vulnerable to recharge from legacy nutrients contained in the groundwater. Future work should address the need for more information on application rates of wastewater treatment residuals and on well depths before the model can be refined and any policy recommendations are made for Nitrate. For groundwater radon, we recommend that our map be used as basis of comparison for analysis techniques presented in the report and to help better understand existing data, maps, and their limitations.

Acknowledgements

This work was funded by grant 11-05-W of the Water Resources Research Institute of North Carolina. It included data provided by the North Carolina Department of Environmental and Natural Resources.

1. Nitrate Introduction

In North Carolina it is estimated that between one third and one half of citizens rely on untreated groundwater as their main drinking water source (Kenny et al. 2005). In addition, groundwater discharge to streams (baseflow) accounts for roughly two-thirds of annual streamflow in the Coastal Plain of North Carolina (Giese et al. 1993) and may be contributing to excess nutrient loads in streams (Tesoriero et al. 2013). A critical mission of the Division of Water Resources (DWR) of the North Carolina Department of Environment and Natural Resources (NC DENR) is to continually assess the quality of the state's groundwater in order to ensure that the state has a clean water supply. DENR maintains about 300 ambient monitoring wells across the state. However, due to budget limitations, not all have been sampled recently for nitrate. Even so, this network of wells represents a fairly limited network for making detailed projections of nitrate concentrations throughout the state. This work presents modeling methods that utilize all of the best available data to access the State's groundwater for a common groundwater contaminant, nitrate, which is of considerable concern in North Carolina and beyond.

Nitrate is one of the most commonly detected nutrients in groundwater in the United States (B. Nolan & Stoner 2000; Spalding & Exner 1993). Furthermore, many areas of North Carolina are known to have detectable levels of nitrate above the US Environmental Protection Agency (EPA) 10 mg/L drinking water maximum contamination level (MCL) (B. T. Nolan & Hitt 2006). Nitrate sources include point and non-point sources. Of special concern in the last decade is the non-point source application of wastewater treatment residuals (WTR). Nitrate in drinking water is associated with infant methameoglobinemia (blue baby syndrome) and negative reproductive effects (Fan & Steinberg 1996).

Previous studies have related land use characteristics to nitrate contamination with success in surface waters(Cressie & Majure 2012; Howarth et al. 1996; Smith et al. 1997; Qian 2005) and groundwater. However, the studies relating land use to groundwater nitrate are conducted at a large area spatial scale; that is, the nitrate value being evaluated as the dependent variable is averaged over some large area (B. T. Nolan & Hitt 2006; McLay et al. 2001; Gurdak & Qi 2012). This study presents a land use regression model that is consistent to the previous literature on groundwater nitrate when nitrate is spatially averaged so as to describe large area variability of nitrate, but it also presents a model of point level nitrate which emphasizes small area variability of nitrate. To the authors' knowledge this is the first attempt to create a state-wide model for small area variability in groundwater nitrate. The results of this study will provide a

detailed map of groundwater nitrate predictions across the entire state of North Carolina which can be used by policy-makers in determining regulations, in providing recommendations to protect citizens' public health, and in managing nutrient loading of surface waters.

2. Nitrate Materials and Methods

2.1. Nitrate Monitoring Data

A groundwater nitrate monitoring database was created from three different data sources. The largest gain in monitoring information came from the address geocoding of private well nitrate measurements, which resulted in over 22,000 nitrate samples. The second main source of data is from groundwater monitoring data from sites permitted to apply wastewater treatment residuals (WTR). The third source is data from the USGS ambient monitoring wells. We originally included a fourth source, EPA's National Contaminant Occurrence Database (NCOD), which is public water system data. We chose not to include the NCOD data because the measurements are clearly not comparable to the other three sources, representing some possible pre-treatment process before sample measurements. Table 1 summarizes the compiled data from each source. Figure 1 displays the spatial distribution of each source. Figure 1 shows that, at a state-wide scale, we have obtained good data coverage.

Table 1. Summary of Groundwater Nitrate Data Sources

Source	Total # of Space/Time Samples	Percent of Samples Above Detection Limit
USGS Ambient Monitoring Wells	1,985	61.4
NC DHHS Private/Domestic Wells	22,684	30.6
NC DENR Land Application	10,368	79.7

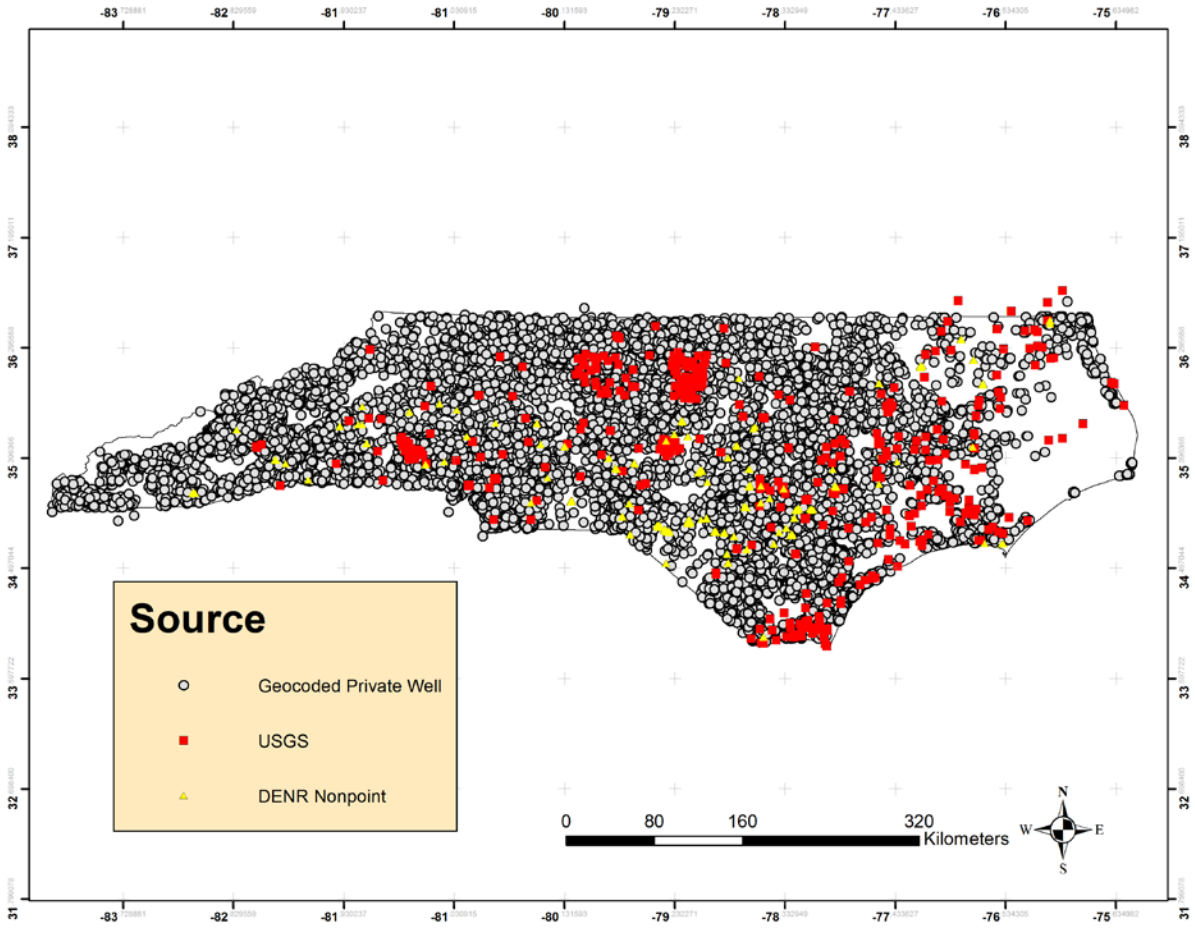


Figure 1 Spatial distribution of compiled groundwater nitrate data in North Carolina. Yellow Triangles are from the DENR/DWR land application database, Grey circles are the geocoded private well nitrate data, and the red squares are USGS monitoring wells.

The groundwater nitrate monitoring data are combined to form a comprehensive database on groundwater nitrate in North Carolina. The distribution of nitrate in groundwater is log-normally distributed; therefore, the data is log-transformed. Furthermore, when data is reported as below detect, we treat the nitrate as ½ the reported detection limit if it is available. If the detection limit is unavailable, or if it is reported as zero, then we treat it as zero. To be able to log transform the data with zeros in the dataset, all of the data has an arbitrary small value of 1mg/L added to it before taking the log, which is subsequently subtracted back out after the results are back transformed for presentation of results in the untransformed space.

2.2 Land Cover Explanatory Variables

We construct explanatory variables based on the National Land Cover Database (NLCD) satellite imagery file that characterizes land cover types at 30 meter resolution. We create variables for every NLCD land cover type and aggregated land cover types such as agriculture and developed. For a NLCD variable (*l*) of interest we calculate

$$LC_i^{(l)}(\lambda_l) = \frac{\sum_{j=1}^{n_i(\lambda_l)} I_j^{(l)}}{n_i(\lambda_l)} \qquad (1)$$

where $LC_i^{(l)}(\lambda_l)$ is the percent of land cover of type (*l*) within a radius λ_l of nitrate point *i*, $I_j^{(l)}$ is an indicator variable equal to 1 if the j^{th} pixel surrounding nitrate point *i* is of type *l*, and zero otherwise, and $n_i(\lambda_l)$ is the number of pixels within the circular buffer of radius λ_l around nitrate point *i*. It is clear from this equation that for a given NLCD variable (*l*) of interest we can construct that variable at many buffer sizes λ_l and eventually pick the construct that leads to the best explanatory power.

2.3 Other Non-Point Source Explanatory Variables

In a similar fashion to the land cover percentage we also calculate the geological formations, soil types and population density within various buffer sizes around each nitrate point. See section 6.3 on geological explanatory variables in the radon part of this report for a more detailed explanation on how geological variables are created.

2.4. Point Source Explanatory Variables

We also calculate the sum of exponentially decaying contribution from various potential nitrate point sources such as permitted treated sewage residuals land application fields, animal operation permit points, concentrated animal feeding operations (CAFO's), swine lagoon locations, waste water treatment facilities, golf courses, and landfills. Equation 2 shows the general form of the point source variables (Messier et al. 2012),

$$PS_i^{(l)}(\lambda_l) = \sum_{j=1}^{n_l} C_{Oj}^{(l)} \exp\left(-3 * \frac{D_{ij}}{\lambda_l}\right), \qquad (2)$$

where $PS_i^{(l)}(\lambda_l)$ is the sum of exponentially decaying contribution from point sources type (*l*) at nitrate point *i*, n_l is the total number of point sources of type (*l*), D_{ij} is the distance between the *j*-th point source of type (*l*) and the nitrate point *i*, C_{0j} is a proxy for the initial nitrate concentration at the point source if available, or equal to 1 otherwise, and λ_l is the exponential

decay range corresponding to the distance it takes for nitrate released by source of type (l) to be reduced by 95%. It is clear in this model that we can calculate the variable for the same point source for varying ranges λ_l since this is a hyperparameter we do not know *a priori.* We determine the best value for λ_l based on the value which maximizes the r-squared regression statistics.

Residual Land Application Permits

North Carolina Division of Water Resources (DWR) maintains data on permits for land application of wastewater treatment residuals. These data, referred to herein as wastewater treatment residuals (WTR), have been studied in previous literature by Keil et al(2011) in which they evaluated its potential use in epidemiologic research. We obtained data that have been updated since the Keil et al. (2011) study; however, the database currently maintained does not contain information on the amount and when WTR is applied. Keil et al. (2011) obtained a small subset of the amount applied for 8 counties over a 3 year period, but this required electronic digitizing of paper reports. Our data contains information on the size of the field in acres, which we use as an estimate for the amount applied and a proxy for the resulting increase in the concentration of nitrate in the underlying groundwater.

WTR is critical information because it has become widely used as a replacement for conventional fertilizers. WTR contains large amounts for nitrogen in the form of nitrate, nitrite, and ammonia. The location and size of the permitted field locations for WTR is shown in figure 2.

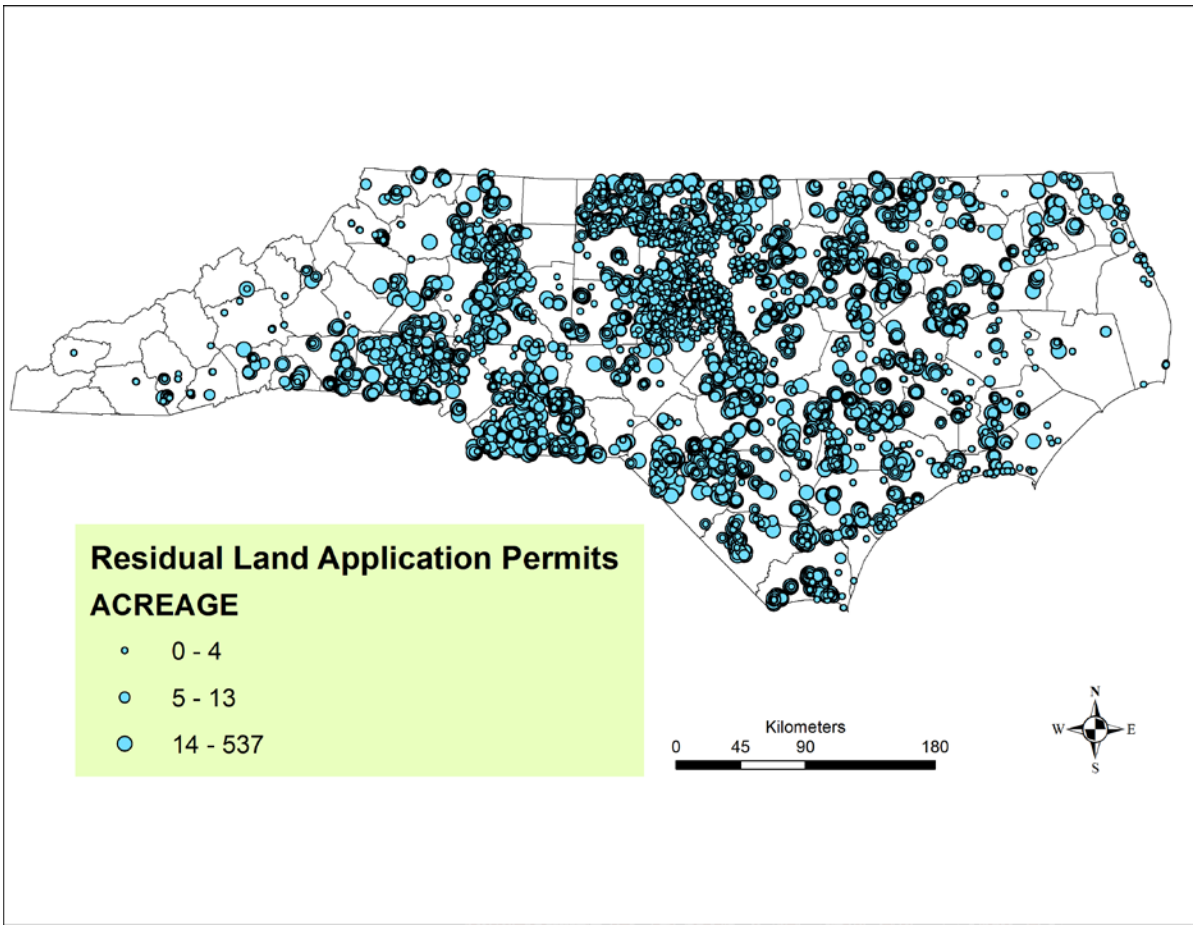


Figure 2. Location and size of residual land application permitted fields. These data represent the most up to date information available to North Carolina Division of Water Resources (DWR)

2.5. Stepwise Hyperparameter Optimization and REgression (SHORE)

In order to develop a land use regression model for nitrate we need a multivariate model selection procedure. As mentioned above, in the case of nitrate we do a first step transformation that consists of adding 1mg/L to the measured nitrate concentration and then taking the log transform, i.e. for any sampling point i of interest, we calculate $Y_i = g_1(Z_i)$, where $g_1(Z_i)=\log(Z_i+1\text{mg/L})$ is the g_1-transform of nitrate concentration Z_i. The spatial regression model for the g_1–transform of nitrate concentration can then be written as follows:

$$Y_i = \beta_0 + \sum_{l=1}^{n} \beta_l X_i^{(l)}(\lambda_l) + \varepsilon_i \qquad (3)$$

where Y_i is the g_1–transform of nitrate concentration Z_i at point i, $X_i^{(l)}(\lambda_l)$ is the l-th predictor variable at point i, β_l is its regression coefficient, λ_l is its hyperparameter value (generally

indicating the length scale at which $X_i^{(l)}(\lambda_l)$ is constructed), and ε_i is an error term. For example $X_i^{(l)}(\lambda_l)$ may be equal to the land cover variable $LC_i^{(l)}(\lambda_l)$ (Eq 1), or the point source variable $PS_i^{(l)}(\lambda_l)$ (Eq 2), or the fraction of area in a circular buffer of radius λ_l around well *i* that is of geological formation type (*l*), etc.

Traditional statistical methods to select predictor variables include forward selection, backwards selection, and stepwise selection. These methods however can lead to erroneous models with high multicollinearity when the potential variables are related, such as is our case with variables that differ only by a hyperparameter. Therefore, we developed and implemented a Stepwise Hyperparameter Optimization REgression (SHORE) procedure that constrains related variables from having more than one selected for the final model, in essence optimizing the distance parameter and preventing multicollinearity. Figure 3 is a flow diagram showing the basic process for creating the final nitrate model starting with all of the potential variables. The procedure starts with a null or intercept only model. Then every candidate variable including variables that differ only by their hyperparameter (i.e. buffer or decay range) is regressed against the response variable. The candidate variable that contributes to the largest increase in r-squared is included in the model. Then each previously included variable is checked for significance and removed if any became non-significant. This component of the procedure represents the stepwise procedure. Additionally, after a candidate variable is included in the model, any remaining candidate variables that differ only by hyperparameter from the newly included variable is removed from the available candidate variables, which prevents multicollinearity and represents a form of optimization. Ties in r-squared (to the thousandths place) are settled by the lowest p-value. The selection criterion used to include a candidate variable is a user-defined input that will affect the results. We start with 0.01, or a 1% increase in r-squared, but it can be varied depending on the stringency desired.

This report represents the first formal presentation of the SHORE procedure. SHORE is similar to the ADDRESS (A Distance Decay Regression Selection Strategy) procedure detailed in Su et al. (Su et al. 2009), which replaces the response variable with prediction residuals at each iteration and mentions nothing in the case of r-squared ties.

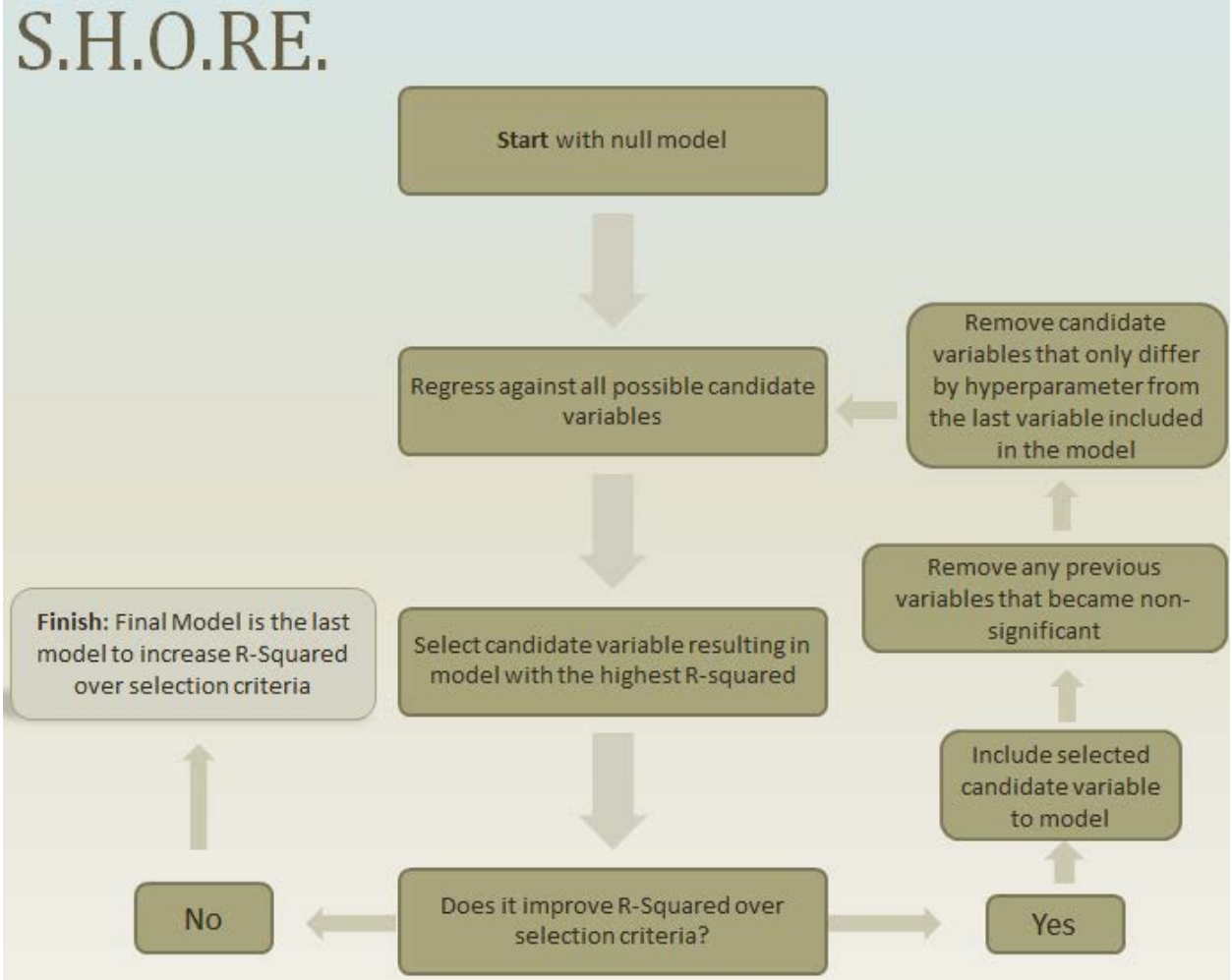


Figure 3. Flow schematic diagram for stepwise hyperparameter optimization and regression (SHORE).

2.6. Spatial/Temporal Mapping

The Kriging method of geostatistics is widely used to estimate exposure in many environmental health studies because it explicitly accounts for spatial dependency and is therefore more accurate than naïve estimations such as nearest neighbor and inverse distance weighted average. However, Kriging has many assumptions including linearity and normality, which limits its effectiveness when dealing with space/time relationships and distributions that are potentially non-linear and non-Gaussian. To address the limitations of Kriging, our group at UNC Chapel Hill has pioneered the Bayesian Maximum Entropy method of spatiotemporal geostatistics(G. Christakos et al. 2002). In this study we use the BME methodology with *BMElib* (M. L. Serre & G. Christakos 1999; George Christakos 1990; de Nazelle et al. 2010), a powerful MATLAB

numerical toolbox implementing the BME theory. In our particular case, we use linear estimators and the data are assumed to be Gaussian; therefore, we are conducting a Kriging analysis, which is a limiting case of the BME methodology.

The theory of space/time random field (S/TRF) is used to model the variability and uncertainty associated with the distribution of groundwater nitrate concentration across space and time. Let $Z(\boldsymbol{p})$ be the concentration of groundwater nitrate at point $\boldsymbol{p}=(\boldsymbol{s},t)$, where $\boldsymbol{s}=(s_1,s_2)$ is the spatial location and t is time. For each point $\boldsymbol{p}_i=(\boldsymbol{s}_i,t_i)$ for which there is an available measured value for Z_i, we defined the transformed data value

$$X_i = g(Z_i) = g_1(Z_i) - o(\boldsymbol{s}_i) \qquad (4)$$

where $g_1(\,)$ is a first-step transformation of Z_i and $o(\,)$ is an global spatial offset that can be set to any arbitrary function of the spatial location $\boldsymbol{s}$. As described above in the case of groundwater nitrate we set $g_1(Z_i)=log(Z_i)+1$mg/L. As for the global spatial offset $o(\boldsymbol{s})$, we first set it equal to zero, i.e. $o(\boldsymbol{s})=0$mg/L, and we then compare that approach to setting it equal to the LUR model (Eq. 3), i.e. $o(\boldsymbol{s})=LUR(\boldsymbol{s})=\beta_0+\sum_{l=1}^{n}\hat{\beta}_l X_{\boldsymbol{s}}^{(l)}(\hat{\lambda}_l)$, where $\hat{\beta}_l$ and $\hat{\lambda}_l$ are the SHORE estimates of the regression coefficients β_l and their corresponding hyperparameter values λ_l. We then model the variability and uncertainty associated with $X(\boldsymbol{p})$ using a homogeneous/stationary space/time random field (S/TRF) for which the set of observed transformed value X_i represents one realization.

The knowledge available about the S/TRF $X(\boldsymbol{p})$ is organized in the general knowledge base (G-KB) describing its space/time variability and the site-specific knowledge base (S-KB) corresponding to the hard and soft data available at sampling wells. The G-KB consists of the mean trend function $m_X(\boldsymbol{p})=\mathrm{E}[X(\boldsymbol{p})]$ and the covariance function $c_X(\boldsymbol{p},\boldsymbol{p}')=\mathrm{E}[(X(\boldsymbol{p})-m_X(\boldsymbol{p}))(X(\boldsymbol{p}')-m_X(\boldsymbol{p}'))]$ of the S/TRF $X(\boldsymbol{p})$, where E[ ] is the stochastic expectation operator. In this work the S/TRF $X(\boldsymbol{p})$ is homogeneous/stationary so that its mean $\mathrm{E}[X(\boldsymbol{p})]=m_X$ is constant, and its covariance function $c_X(\boldsymbol{p},\boldsymbol{p}')=c_X(r=\|\boldsymbol{s}-\boldsymbol{s}'\|,\tau=|t-t'|)$ is only a function of the spatial lag $r=\|\boldsymbol{s}-\boldsymbol{s}'\|$ and temporal lag $\tau=|t-t'|$ between points $\boldsymbol{p}=(\boldsymbol{s},t)$ and $\boldsymbol{p}'=(\boldsymbol{s}',t')$. Furthermore, in this work, the S-KB consists of hard data values only, i.e. values with negligible measurement errors. In this case the BME method reduces to the kriging method of linear geostatistics. Using kriging we obtain the mean and variance of $X(\boldsymbol{p}_k)$ at any unsampled space/time point $\boldsymbol{p}_k$ of interest. Next we use the back transformation $Z(\boldsymbol{p}_k)=g_1^{-1}(X(\boldsymbol{p}_k))+o(\boldsymbol{s}_k)=\exp(X(\boldsymbol{p}_k))-1\text{(mg/L)}+o(\boldsymbol{s}_k)$ relationship to obtain the median of nitrate concentration Z at $\boldsymbol{p}_k$. Finally by locating the estimation points along a regular grid covering NC we are able to calculate and display groundwater nitrate concentration across the State and for any time of interest.

A major assumption in the Kriging methodology is the assumption for the form of the global mean trend. In classical Kriging such as Simple Kriging, Ordinary Kriging, and Universal Kriging, a global mean trend is assumed to be known, constant, or polynomial function of space

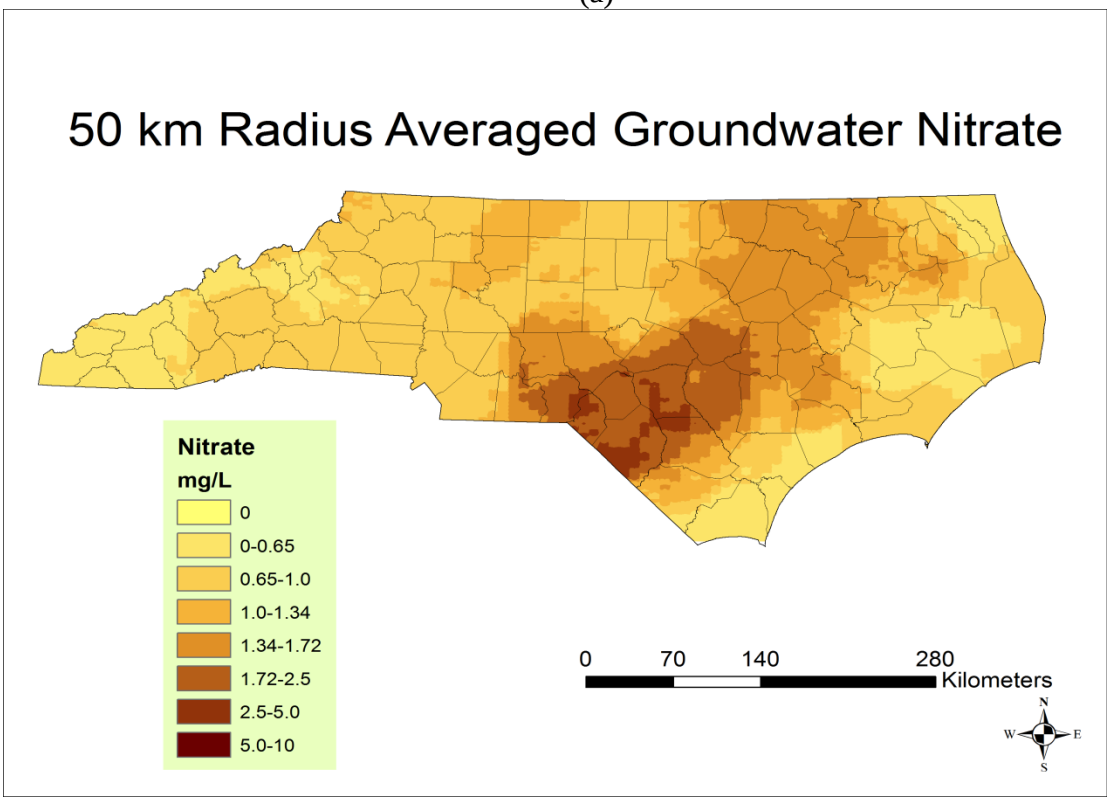
and time. To account for secondary information that can potentially help estimation accuracy, many geostatisticians have implemented a Kriging in which the global mean trend is based on a land use regression (LUR) model of secondary data (De Nazelle et al. 2010; Messier et al. 2012; Hengl 2004). This method may be referred to as Regression-Kriging, or LUR-Kriging. In this work we implement Simple kriging by setting the global spatial offset $o(\boldsymbol{s})$ to a constant value, and we implement LUR-Kriging by setting $o(\boldsymbol{s})$ to the LUR model (Eq. 3).

3. Nitrate Results

3.1. Nitrate Averages at 25 km and 50 km scales

We evaluated the large-area trends in groundwater nitrate by averaging nitrate over 50 km and 25 km radii and over the entire study time period. This resulted in purely spatial values of nitrate averages that are shown in Figures 4a and 4b, respectively. These maps are presented to provide a view of the general spatial trends in groundwater nitrate across North Carolina.

(a)



(b)

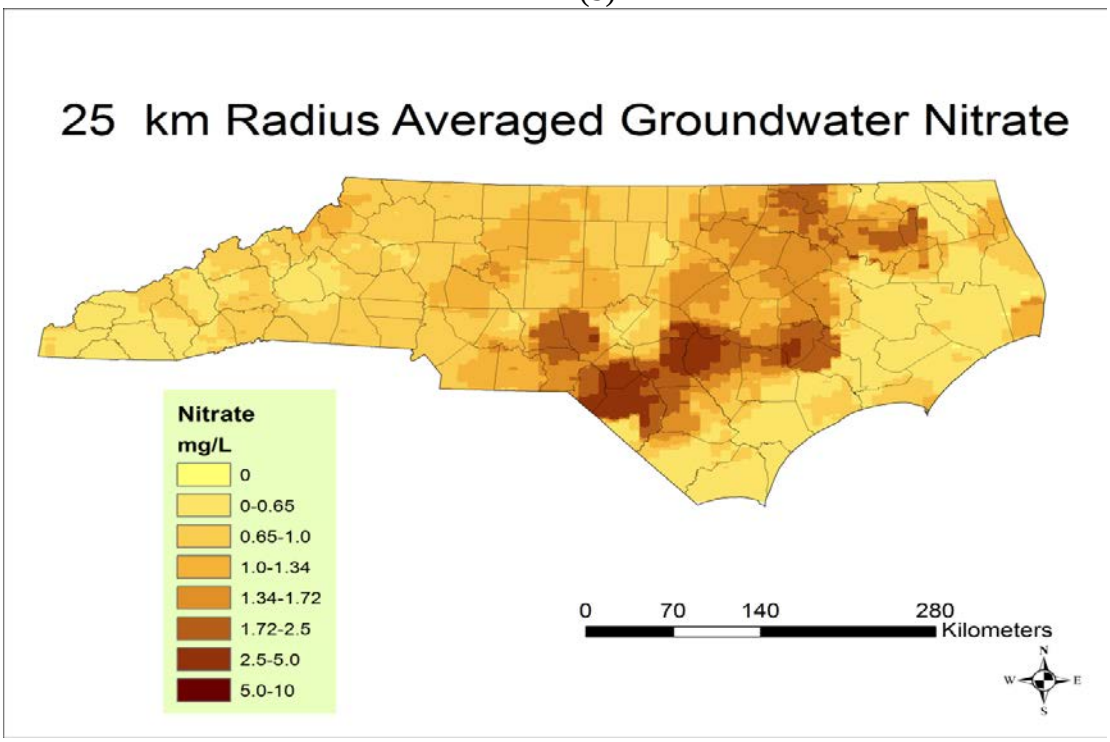


Figure 4. (a) Purely Spatial 50 Kilometer Radial Averaged Groundwater Nitrate, (b) Purely spatial 25 kilometer radial averaged groundwater nitrate

3.2. SHORE Model of the Large Area Variability of Nitrate

The averages of nitrate over 50 km and 25 km radii (Figure 4a and 4b) were used as the dependent variable in the SHORE procedure to evaluate how large-area trends are associated with land use and other explanatory variables.

The application of the SHORE procedure on the 50 km radial averaged nitrate data results in a total r-squared of 0.74, which is comparable to previous literature on spatially averaged nitrate data. The variables resulting from the SHORE procedure are summarized in table 2. Cretaceous Sedimentary a geological feature that is dominant in the southeastern portion of our state. All of the variables represent land cover effects that make physical sense. The beta coefficients are in the direction that one would expect based on previous literature. All of the variable names are based directly on National Land Cover Database classifications (http://www.mrlc.gov/nlcd06_leg.php). Agriculture represents the combination of pasture/hay and crops.

Table 2. 50 Km radial averaged nitrate land use regression results.

		R-Squared = 0.74
Variable Name	**Buffer Size (Km)**	**Beta**
Intercept	**-**	**0.40**
Percent Cretaceous Sedimentary	**50**	**0.24**
Percent Agriculture	**50**	**0.98**
Percent Herbaceous	**50**	**3.44**
Percent Crops	**25**	**0.31**
Percent Herbaceous Wetlands	**50**	**-5.96**
Percent Developed Low	**50**	**-4.28**
Percent Barren	**50**	**12.90**

The results from SHORE on the 25 km radial averaged nitrate data yield a total r-squared of 0.49. The variables resulting from the analysis are summarized in table 3.

Table 3. 25 Km radial averaged nitrate land use regression results.

		R-Squared = 0.49
Variable Name	**Buffer Size (Km)**	**Beta**
Intercept	**-**	**0.41**
Percent Agriculture	**50**	**0.72**
Percent Herbaceous	**50**	**2.14**
Percent Cretaceous Sedimentary	**50**	**0.26**
Percent Wetlands	**25**	**-0.85**
Percent Crops	**25**	**0.77**

3.3. SHORE Model of the Small Area Variability of Nitrate

We then run the SHORE model using point nitrate values (i.e. values of nitrate without any averaging). This resulted in the successful quantification of the effect of land use characteristics on groundwater nitrate at a point level scale. The results for the land use regression model are summarized in Table 4. The model results in a final r-squared of 0.18. The variable WTR has an exponential decay range of 2,000 meters which demonstrates that the WTR sites are a local source of nitrate pollution. Metaigneous bedrock is a basic geologic feature that is present in the central region of the state. The cubic trend represents a normalized, large area trend in nitrate across the entire state.

Table 4. Point level nitrate land use regression results

		R-Squared = 0.18
Variable Name	Buffer Size (m)	Beta
Intercept	-	0.5868
WTR	2000	0.0060
Percent Crops	100	0.3631
Metaigneous	1000	-0.5157
Percent Deciduous	20	-0.2483
Cubic Trend	-	0.0601
Percent Evergreen	100	-0.4630

3.4. Nitrate Space/Time Covariance Analysis

We investigate the spatial and temporal autocorrelation present in the nitrate data, and in the LUR residual nitrate data by modeling their covariance functions (Olea 2006; George Christakos 1990). The combination of nitrate data sources results in noisy estimated covariance values, thus we estimated and modeled covariance using the nitrate data from the private well dataset. These data represent approximately 2/3 of the data and lead to a smooth covariance estimation. The covariance model used is the sum of two space/time non-separable exponential/exponential covariance functions given by the following equation

$$c_X(r,\tau) = c_{01}\exp(-3r/a_{r1})\exp(-3\tau/a_{t1}) + c_{02}\exp(-3\tau/a_{r2})\exp(-3r/a_{t2}) \quad (5)$$

Table 5 shows the modeled covariance parameters for the g_1-transformed nitrate data Y_i=log(Z_i+1mg/L) used in the Kriging model ;and the LUR residual data $X_i = Y_i - LUR(\boldsymbol{s}_i)$ used in the LUR-Kriging model.

Table 5. Space/Time covariance model results for 1) private well g_1-transformed nitrate data Y_i=log(Nitrate+1*mg/L*) and 2) private well LUR-residual data $X_i = Y_i$–LUR(s_i). Model results were obtained by least square fitting of the covariance model (Eq 5) to experimental covariance values calculated from the data.

Covariance Model	C_{01} $(\log(mg/L))^2$	a_{r1} (m)	a_{t1} (days)	C_{02} $(\log(mg/L))^2$	a_{r2} (m)	a_{t2} (days)
g_1-transformed nitrate data from private wells	0.199	960	35,998	0.122	36,404	39,766
LUR-residual nitrate data from private wells	0.228	666	21,964	0.082	190,260	21,964

3.5. Nitrate Kriging Maps

Kriging maps of the median of nitrate concentrations were obtained using the g_1-transformed nitrate monitoring data and the covariance model shown in the first row of table 5. Figure 5 shows the Kriging median estimate of groundwater nitrate on 3/31/00. The Kriging map can be produced for any day during the study period; however, due to the long temporal covariance range there is relatively little change in the maps from day to day. Nonetheless, the link listed below Figure 5 shows an animated GIF movie of 4 different days across the study period.

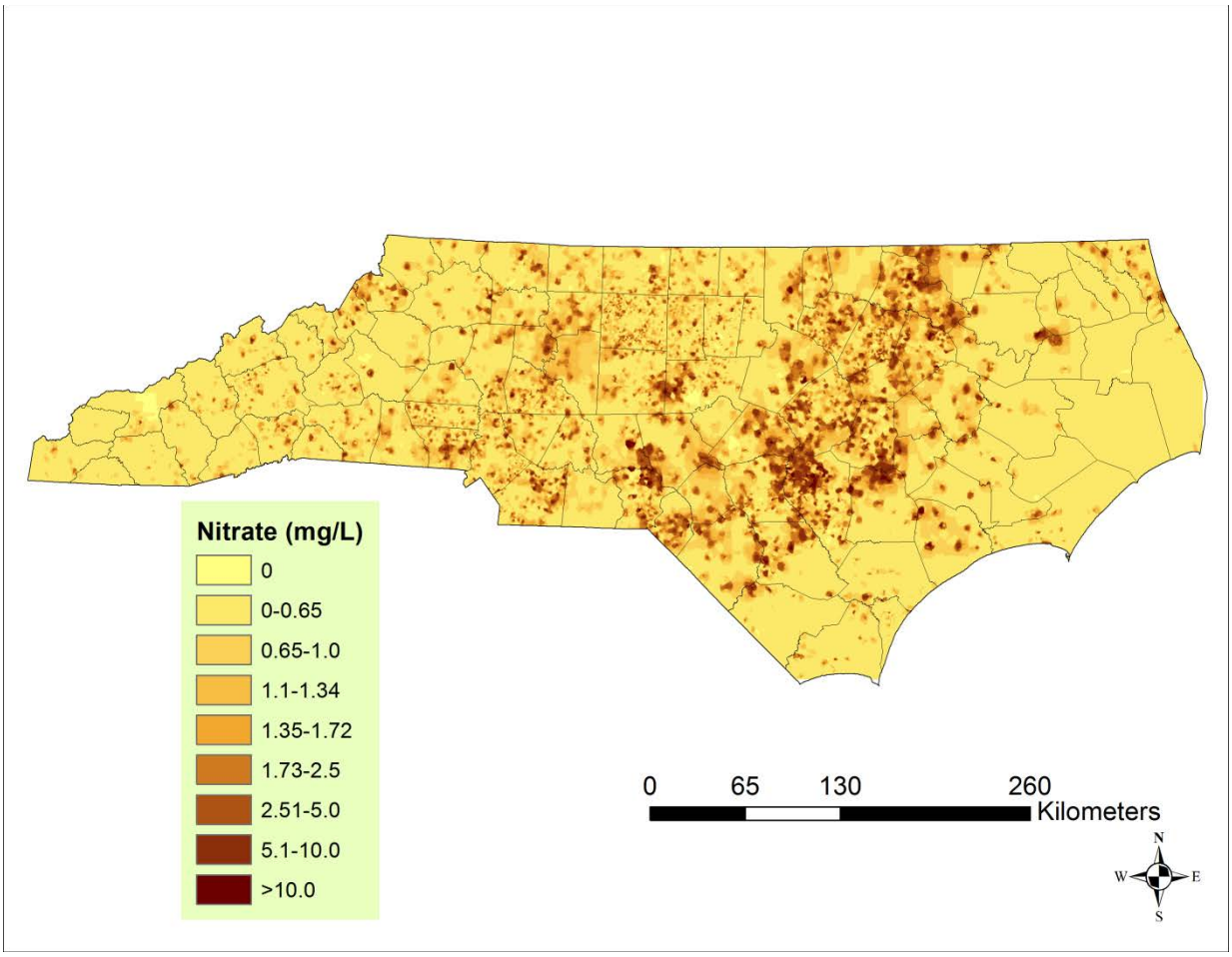


Figure 5. Kriging Median Estimate of Groundwater Nitrate on 3/31/00

http://www.unc.edu/depts/case/BMElab/studies/NO3-Rn_NC_WRRI/NO3_GW_Kriging_2000-2010_1.GIF

3.6. Nitrate LUR-Kriging Maps

LUR-Kriging maps were produced using the LUR-residual data and the covariance model shown in the second row of table 5. Figure 6 shows the LUR-Kriging median estimate of groundwater nitrate for the same day as figure 5 for comparison purposes. Moreover, there is also a link provided to see an animated GIF movie of the LUR-Kriging estimates for four days across the study period.

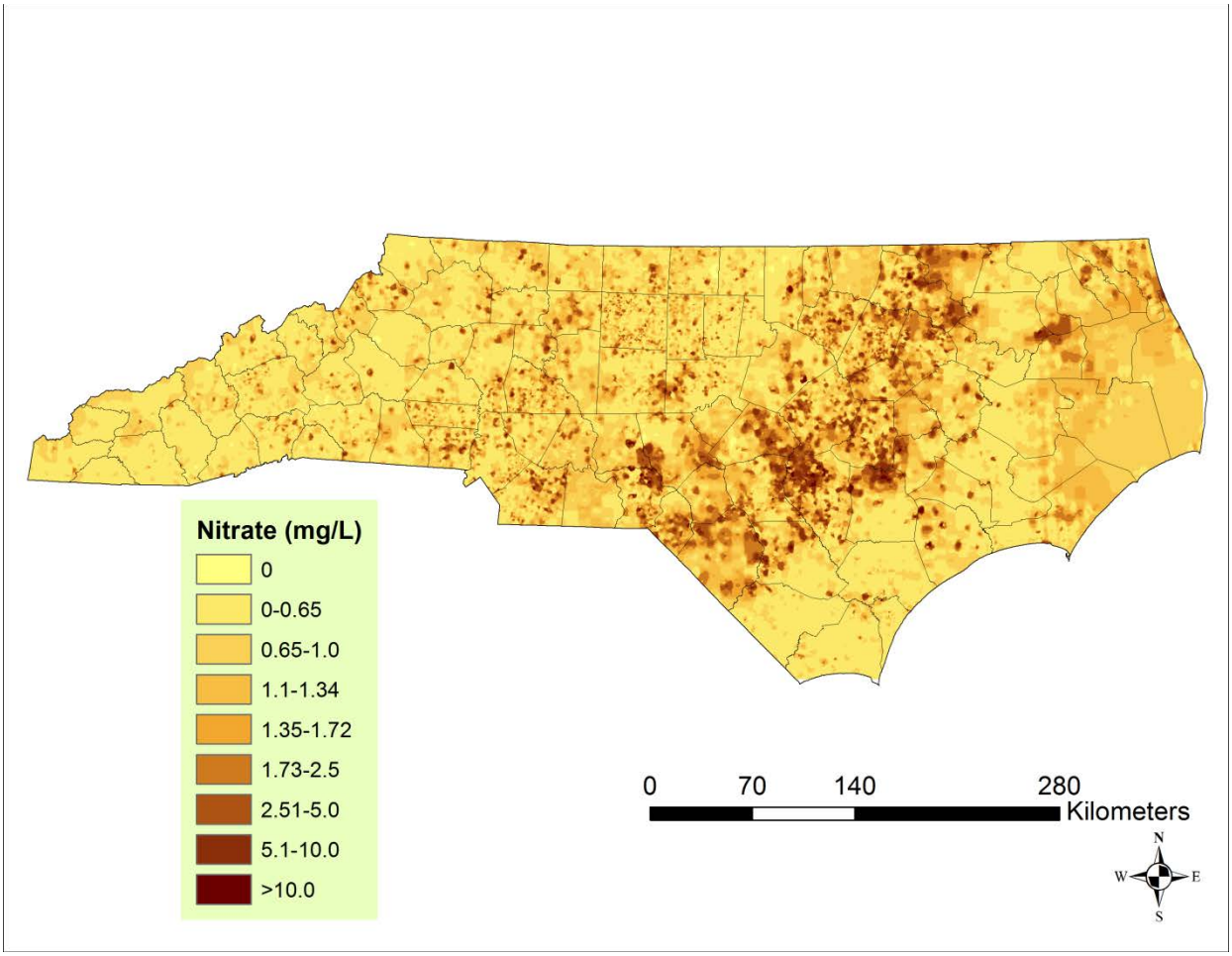


Figure 6. LUR-Kriging Median Estimate of Groundwater Nitrate on 3/31/00

http://www.unc.edu/depts/case/BMElab/studies/NO3-Rn_NC_WRRI/NO3_GW_LUR-Kriging_2000-2010_1.GIF

3.7. Nitrate Cross-Validation Analysis

The cross validation analysis consists in removing each detectable the g_1-transformed nitrate value Y_i in turn from the data, and using a given estimation method (k) to calculate its estimate $Y_i^{*(k)}$ based on the remaining data. The mean square error (MSE) calculated as

$$MSE^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{*(k)} - Y_i\right)^2 \qquad (6)$$

where *n* is the number of data points, provides a measure of the overall estimation error for method (k). Table 6 shows the results for the cross-validation analysis of the Kriging and LUR-Kriging models with all of the data present in the models, and with only above detect data

present. We also show the percent change when moving the Kriging to the LUR-Kriging case. It is clear that it both all data and above detect data cases, the LUR-Kriging leads to more accurate estimates of groundwater nitrate.

Table 6. Cross-Validation results for Kriging and LUR-Kriging methods of point level data.

	Kriging: All Data	LUR-Kriging : All Data	Kriging: Above Detect	LUR-Kriging: Above Detect
MSE $(\log(mg/L))^2$	**0.267**	**0.253**	**0.504**	**0.457**
Percent Change		**-5.2%**		**-9.3%**

4. Nitrate Discussion

4.1. Novel Contributions to Nitrate Spatial Modeling

This study resulted in a substantial increase in available groundwater nitrate data, provided large area averaged estimates of groundwater nitrate across North Carolina, and provided the first point level estimates of groundwater nitrate across North Carolina. Furthermore, this study was also the first to establish land use variables that affect point level groundwater nitrate state-wide for North Carolina.

There is substantial literature on the processes that affect the distribution of nitrate in groundwater; however, the analysis when conducting a statistical based model is based on data averaged over large buffer areas (B. T. Nolan & Hitt 2006; McLay et al. 2001; Gurdak & Qi 2012; Spalding & Exner 1993), and when the analysis is mechanistic the study area is usually a watershed (Kennedy et al. 2009; Tague & Band 2004) . This study provides a balance between the two sides by providing a point level estimate while using a statistical model which makes a study area such as North Carolina feasible. The results of this study will provide state personnel with maps to help better allocate scarce resources when it comes to nutrient management. It also provides the most accurate assessment of groundwater nitrate across North Carolina for epidemiologists studying potential health effects associated with groundwater nitrate exposure in North Carolina.

4.2. New Nitrate Data Sources

The largest new data source contribution from this study came from the geocoding of private well data. Over 22,000 private well samples were geocoded with a geocoding score of 80 or greater. Geocoding scores of 80 or greater (on a scale of 0-100) are usually considered good matches with the assigned spatial location assumed to be reliable (J. A. Mcelroy et al. 2011). Nonetheless, there is potential error that is difficult to discern in the geocoding process. Based on qualitative analysis of the difference between geocoded location and GPS reported coordinates, we believe that between ¼ of 1 percent and 1 percent of the geocoded data could be a significant distance from its actual location. We used a multistage geocoding process where addresses were tested against three reference databases, which increase the chances of an address to be geocoded. We also discarded any address that was geocoded with a score between 0 and 79. Reasons for addresses not being geocoded include complete or large percentage of address missing, address misspelled beyond recognition, and PO Box entered as address.

4.3. Large Area Variability of Nitrate

Figures 4a and 4b show the purely spatial large area averaged maps at the 50km and 25km radial scale respectively. A major difference between the two is that less high values are seen in the 50km map because the high values are smoothed out by the substantial averaging process. Both maps show that the area of largest concern is a large strip in the eastern portion of the state, but not necessarily all the way to the coastal area. This area is well-known to have nutrient contamination issues because of substantial agricultural practice.

Prior to this study, large area analysis of groundwater nitrate in North Carolina had been considered by studies whose study area was larger than North Carolina. Furthermore, the studies did not have the benefit of over 30,000 groundwater nitrate samples from 3 different data sources.

We conducted a SHORE analysis of Nitrate averaged at the at the 50km and 25km radial scales for two main reasons: 1) To give researchers and state personnel an idea of the general groundwater nitrate trends in North Carolina, and 2) to show that our land use regression modeling procedure can match previous studies in terms of total r-squared if we are on a level playing field, that is, modeling the large-area variability of groundwater nitrate.

The SHORE analysis at the 50 km radial averaged scale resulted in a total r-squared of 0.74. The variables selected were large area NLCD data variables representing land cover characteristics such as percent agriculture and forest. At this large area scale our point source variables were not significant in multivariate models.

As the area of the nitrate averaging was reduced from a 50km to a 25km radius, the SHORE procedure unsurprisingly resulted in a reduced r-squared of 0.49. It becomes clear that as the

observation scale or averaging area of nitrate is decreased, the corresponding r-squared also decreases. This is due to the fact that there are definitive large area trends of groundwater nitrate that can be explained by large-area scale land use predictors, but there is also significant local variability which is much more difficult to explain with small-area scale land use predictors. Similar results of the SHORE model were obtained in the 25km case as the 50km in terms of the variables selected. Variables selected represent NLCD land cover characteristics affecting groundwater nitrate.

4.4. Point Level Variability of Nitrate

The most significant contribution of this study is the land use regression and Kriging analyses of groundwater nitrate at a point level, or a scale in which the data has not been spatially or temporally averaged. The SHORE analysis resulted in a model r-squared of 0.18 with explanatory variables that all represent small area spatial scales. For instance, the NLCD variables present in the model, crops, deciduous, and evergreen, all have buffer sizes of 100 meters or less. Also of significance, is the selection of the treated wastewater or WTR variable. This point source variable has varying initial concentration values that are based on the size of the land application field in acres. The variable selected has an exponential decay range of 2,000 meters which suggest that the land application sites could be contributing to elevated levels of groundwater nitrate. Figure 7 shows how the r2 changes as a function of the decay range λ_l for the single predictor model using only the WTR variable (Eq. 2), where the initial concentration proxy C_{0j} is calculated either as a constant (dotted line), or as the field size (i.e. the acreage of the application field). It is clear that the decay range that is most predictive of groundwater nitrate is at a short range, between 2,000 and 4,000 meters with the max r-squared in the univariate case being 3,000 meters. One can also see that setting the initial concentration C_{0j} to the field size leads to a better model than setting it to a constant. This implies that field size could be a proxy for initial concentration in the groundwater below the application field, although actual amounts or rate applied would likely be the most accurate.

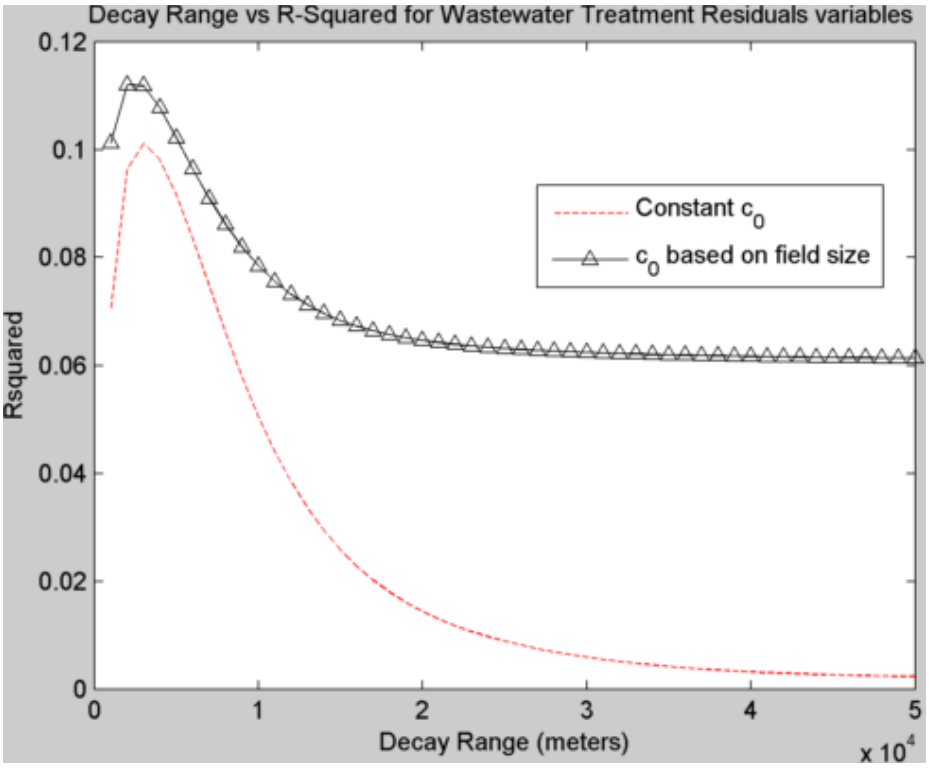


Figure 7. Land Application Residual exponential decay range vs. r-squared for a univariate model.

Overall, our model for point level nitrate suggests that WTR are a significant point source of nitrate within a range of about 2000m from the reported location of where WTR are applied. The size of that range characterizes the distance between the reported location and the location where WTR are actually applied, as well as the travel distance needed to achieve attenuation in nitrate. The implications of the WTR land application sites being a direct source of nitrate are significant in terms of management and regulation. However, given the fact that we lack comprehensive data on rates or amounts being applied, further study is needed before any formal policy recommendations are made.

4.5. Limitations and Recommendations for Nitrate

A limitation of our nitrate dataset is that it lacks information of well depth, and we therefore do not know what aquifer that the wells tap into. While our study has significantly increased the nitrate dataset for which geographical locations are known, more work is needed to retrieve well depth. However this limitation is not unusual for studies performed over domain sizes as large as ours, where it is not foreseeable that we can monitor multiple wells at multiple depths in order to obtain the data needed for a 3-dimensional characterization of aquifers over our whole study domain. Such 3-dimensional studies can be performed over relatively small watersheds that are

at risk of containing high nitrate levels, and these high risk watersheds can be identified using the map we produced.

The well depth changes primarily depending on the type of the well that is sampled. Monitoring wells near WTR sources tend to be shallow so that they capture the nitrate concentration at the top of the surficial aquifer being contaminated from the land application of WTR. Private wells on the other hand tend to be deeper because of the need to reach a depth providing sufficient flow. These two types of wells tend to be in areas that are geographically distinct (Figure 1). Therefore, our groundwater nitrate map should be interpreted as providing the groundwater nitrate concentration at a depth that varies depending on location: Starting at WTR sources, the map shows nitrate at the top of the surficial aquifer, and moving away from these sources, the map shows concentration at greater depths where private wells are installed.

This map can be used for two objectives: One is to provide information about levels that residents using private wells may be exposed to. The second is to quantify the groundwater discharge to streams.

It is important to also note that WTR application fields are almost always located in agricultural land use areas that can also receive nitrate from numerous other sources. WTR application fields are typically located on former agricultural fields which are likely to have been fertilized by inorganic fertilizers and/or animal waste prior to their use as WTR application fields. Since agricultural fields that receive unregulated fertilizers are not subject to groundwater monitoring, it is not possible to determine if WTR application fields may have pre-existing elevated groundwater nitrate levels that were present before the fields were used for WTR application.

Concerning the groundwater nitrate exposure of private well owners, our map provides an estimate of that exposure in areas where private wells are commonly used, but in areas near WTR sources our map is providing an estimate of nitrate concentration at the top of the surficial aquifer. Since that surficial aquifer is presumably being contaminated from the ground above, then we expect that nitrate at greater depth should be protected. However, while it is expected that most nitrate contamination would occur in surficial aquifers and that confining layers prevents most contamination from reaching the confined aquifer below, it is conceivable that in some cases the confining layer may actually be semipermeable or there may be geologic features that allow direct access to the confined aquifer. There are examples where acidic conditions in the unconfined aquifer have leached to the confining unit and created a small area where water and nitrate are more easily routed to the confined aquifer (Spruill 2004). Depth to the confined aquifer can also vary significantly making transport through the unconfined aquifer shorter. This can result in less time for denitrification to occur, which means there may be more nitrate in the unconfined aquifer and more potential for the confined aquifer to be recharged with nitrate. Because of this potential contamination of the confined aquifer, we recommend that our map be used to inform local and state regulatory agencies about WTR application fields with potentially high surficial nitrate levels, so that set backs can be recommended to private well owners living

in these areas. Our model indicates that prediction of surficial nitrate levels is maximized when the range for the WTR variable is 2000 m. This range comprises both the distance between the reported location and the location where WTR are actually applied, as well as the travel distance needed to achieve attenuation in nitrate. Hence this range is large and an overestimation of the actual surficial nitrate plume, because it also incorporates the locational error between the reported WTR location, and the actual location where WTR are applied.

Concerning the potential groundwater discharge to streams, our map could be used by combining it with maps of base flow index (BFI) to quantify nitrogen loading from groundwater. Tesoriero et al (http://pubs.acs.org/doi/abs/10.1021/es305026x) hypothesize that "a first approximation of stream vulnerability to legacy nutrients may be made by geospatial analysis of watersheds with high nitrogen inputs and a strong connection to groundwater (e.g., high BFI)." Hence we recommend that our map be used to provide an index of stream vulnerability to groundwater nitrate by multiplying our mapped estimates of nitrate concentration by mapped estimates of BFI.

Future work should be considered to address the need for more information on application rates of wastewater treatment residuals and on well depths. Of particular interest is the lack of well depth data, which could be addressed by separating the dataset and the LUR model into two categories representing confined and unconfined aquifers, respectively, similar to that of Nolan and Hitt (2006). The first category would be based on private well data, which would be assumed to represent mostly deep or confined aquifers. The second would be based on USGS and DENR land application data which would be assumed to represent shallow or unconfined aquifers that are most likely not directly used for drinking water. Alternatively the lack of well depth data could be addressed by conducting a smaller study with only a subset of the data, perhaps in one county where well depth data can be collected, which would allow to compare the results of a model with and without well depth.

5. Radon Introduction

Radon is a naturally occurring radioactive gas that is chemically inert, colorless, and odorless (WHO 2011). Radon is produced from the decay of uranium, which is found in rocks and soils worldwide. Outdoor radon levels are generally very low; however, when radon enters a residential home, its concentration can increase to levels that may lead to adverse health effects (WHO 2011). Several epidemiological studies revealed that residential exposure to radon in indoor air is associated with lung cancer (J H Lubin & Boice 1997; Krewski et al. 2005; Darby et al. 2004; Pavia et al. 2003) . Currently, exposure to radon is the second leading cause of lung cancer after smoking in the US. Even though radon gas directly escaping from soil and rock is an important route of exposure, radon can also degas from water under various water uses, such as showering, dishwashing, and clothes washing (Vinson et al. 2008; Fitzgerald & Hopke 1997).

Radon in groundwater is not only a concern because of its contribution to indoor air radon, but also due to the direct ingestion of drinking water with elevated radon. There is evidence that exposure to radon through drinking water and indoor air can lead to stomach cancer (Anssi Auvinen et al. 2005); however, this human health endpoint is understudied compared to lung cancer and there is not a consensus among the literature.

Several counties in western North Carolina are classified as EPA Zone 1 counties, with predicted indoor air radon concentrations above the action level of 4 picocuries per liter (pCi/L). Over 90 percent of wells sampled in that region exceed the EPA's proposed Maximum Contaminant Level of 300 pCi/L and a large number exceeded the alternate MCL of 4000 pCi/L(T. Campbell et al. 2011). Since monitoring radon concentration is not mandatory for private well owners, elucidating the spatial distribution of radon across the state is indispensable to informing people about exposure to waterborne radon.

Radon distributions are primarily influenced by geology (Campbell et al. 2011; Loomis 1987). About 25 percent of the Piedmont and mountains of North Carolina are underlain with rocks commonly associated with elevated radon in water, namely felsic intrusive rocks such as granites and granitic gneisses. Through water sampling Campbell et al.(2011) have found 19 counties in North Carolina that are particularly susceptible to elevated radon in water. In this study, we use the samples from Campbell et al. (2011) plus geocoded samples from private well sources and USGS to develop a linear regression model that can be used to model groundwater radon concentrations in the groundwater across North Carolina.

6. Radon Material and Methods

6.1. Radon Monitoring Data

Radon monitoring data was compiled from three data sources. The first data source corresponds to wells sampled and analyzed by the NC DWQ, including areas where radon levels were suspected to be elevated based on geology. This resulted in 655 samples of groundwater radon and their known spatial locations. The second data source was provided by two private radon testing companies that analyzed samples sent by private well owners. These private well owners are generally requesting their well to be tested because they want to know the radon level in their well. The private radon testing companies provided dissolved radon measured concentrations and addresses within a data use agreement that ensures that the data cannot be shared without their permission. These data were address geocoded giving an additional 1,167 samples. The third data source consisted of the USGS via their data website nwis.waterdata.usgs.gov, which yielded an additional 297 dissolved radon measurements. Figure 8 provides a view of the spatial distribution of the radon data stratified by the data source.

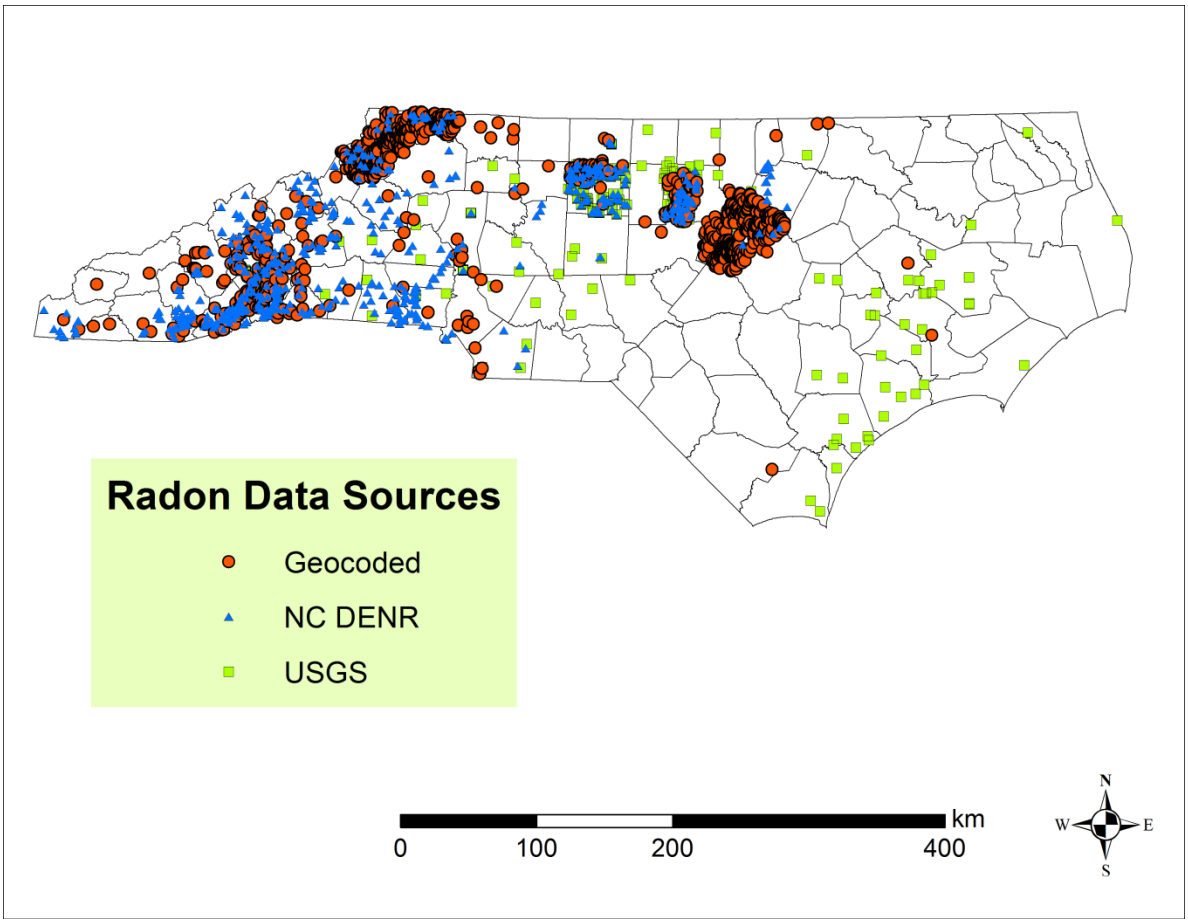


Figure 8. Radon Data by source. Geocoded radon data was obtained through a data use agreement with private companies. NC DENR data was collected by project collaborator Ted Campbell. USGS data was obtained on the internet via nwis.waterdata.usgs.gov.

For the NC DWQ data, groundwater radon samples were collected by NC DENR personnel from a plumbing fixture as close to the wellhead as possible, usually from the wellhead itself. The sample was collected after the pump had been operating for at least 20 minutes to ensure the water was not from a stagnant water column. Samples were collected using a special procedure to prevent aeration of the radon. Specifically, 60 milliliter glass radon vials were carefully submerged, filled, and sealed inside a 2 liter plastic beaker that had been filled with well water under laminar flow. The samples were then put on ice to maintain a cool temperature and shipped to a certified laboratory overnight. Most radon samples were analyzed using the analytical E-Perm ion electret de-emanation procedure (Kotrappa & Jesters 1993); a smaller number were analyzed using Standard Method 7500-Rn procedure (EPA, 1999).

The private well samples were analyzed using the Standard Method 7500-Rn procedure (EPA, 1999). Private home owners receive kits provided by private companies contracted to analyze the dissolved radon concentrations. The kits contain detailed instructions on how to sample, store, and ship according to EPA approved methods.

6.2. Geological Explanatory Variables

Geology formations were obtained from USGS/NCGS (USGS 2013; NCGICC 2013), which we used to create multiple hierarchal geological classifications. We also worked to create different hierarchal classifications of geology age. Tables 3 and 4 show the various levels of detail within each classification level for geology formations and age, respectively.

Table 7. The number of variables within each geological formation classification level. The number of variables refers to geology variable names only and does include the potential to create more variables with varying buffer sizes.

	Geology A	Geology B	Geology C	Geology D
Number of Variables	4	8	23	184

Table 8. The number of variables within each geological age classification level.

	Geology Age A	Geology Age B	Geology Age C
Number of Variables	6	9	14

The hierarchal classifications of geology formations and ages is designed to account for various spatial scales in geological formations and to protect against misclassification since there is inherent uncertainty in the geological formation maps created by the USGS and NCGS, which is discussed further in the next section.

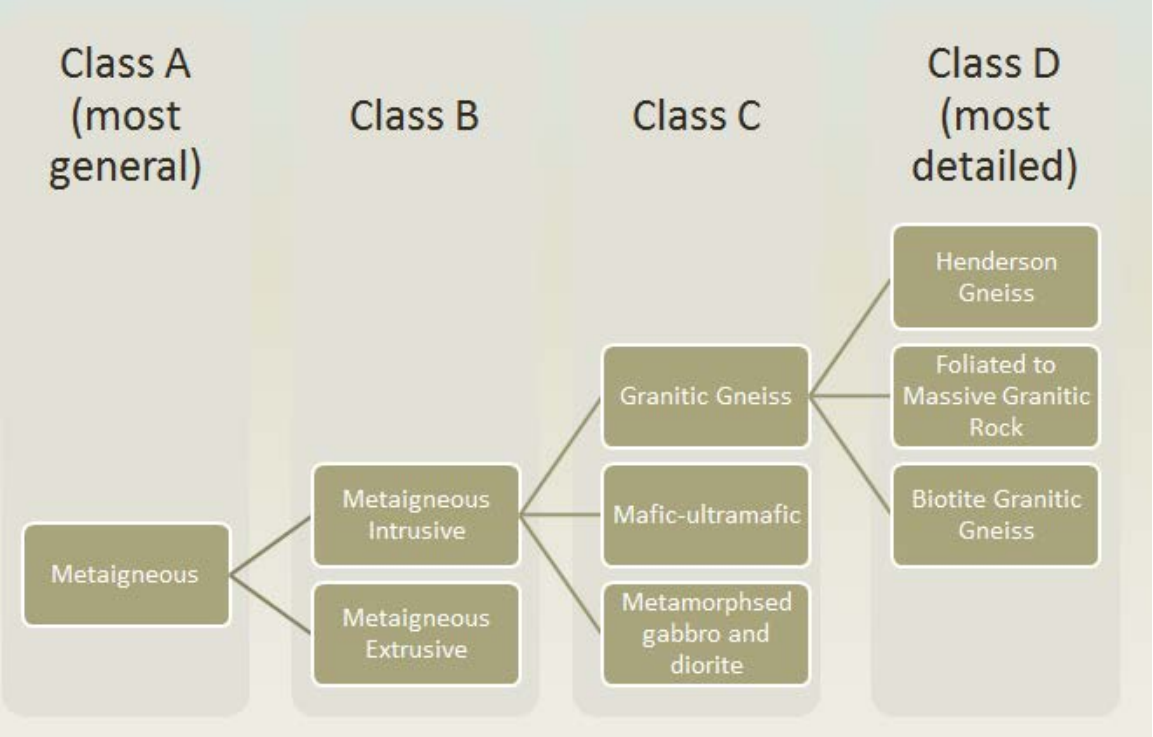


Figure 9. Example of the geology variable hierarchy.

We create variables for each geology formation and age type in the same manner we calculated land cover variables for nitrate with equation 1. We calculate the percentage of a geological formation of type l within a circular buffer of radius λ_l. Figure 9 shows an example of how the hierarchal nature of geological classifications lends itself to calculating multiple explanatory variables. In the most general case, we have metaigneous rock, which can then be classified into metaigneous intrusive or metaigneous extrusive, which makes up class B for the particular example. Metaigenous intrusive can then be further broken down into granitic gneiss, mafic-ultramafic, or metamorphosed gabbro and diorite, which represent class C in this example. Granitic gneiss can then be broken down into Henderson gneiss, foliated to massive granitic rock, or biotite granitic gneiss. For each block in Figure 9, we calculate multiple explanatory variables as the percent of the geological class within a buffer of length λ_l, which vary from 1 kilometer to 50 kilometers by 1 kilometer increments. Short buffer lengths correspond to potential direct sources of dissolved radon while the large buffer lengths (> 25km) correspond to broad geological and lithological trends. It should be noted that each branch in Figure 9 represents only one example and that the same process exists for metaigneous extrusive, and for the other most general classifications in class A.

6.3. Radon and Geology Data Limitations

Many of the wells selected by the NC DWQ to sample Radon are biased towards selecting sites where dissolved radon is thought to be elevated because of the desire to protect citizens' health. This could potentially lead to sampling sites having a very narrow geographical coverage;

however, with the addition of geocoded and USGS dissolved radon data, we improve the geographical coverage of our dataset (Figure 8). Nonetheless, there are still substantial areas of central and eastern NC that are underrepresented in sampling. Additionally, the USGS sample locations are also spatially biased because certain counties have had specific studies leading to some areas with a high density of samples.

Geocoded private well samples are generally collected by well owners who might not have an extensive experience in sampling techniques. Though sampling instructions are provided, it is possible that an inexperienced person samples a well in a way that allows for the off-gassing of radon from the water sample. However this can only lead to a potential bias of the geocoded private well data towards low values, but not toward high values. After sampling the samples are analyzed in a lab that follows similar protocols as those followed by the NC DWQ and USGS, thus we do not expect that samples from private well owners have a lab measurement error that is substantially different than that of NC DWQ and USGS samples.

The geology data used for explanatory variables also brings in its own inherent data limitations. The geology is mapped at the 1:250,000 scale, which translates to a minimum grid cell size of 125 meters. Therefore the minimum buffer size for geology variables is also a 125 meter radius. Furthermore the assigning of geological rock types is a process that can lead to potential misclassification. The creation of multiple variables at different levels in the geological hierarchy is designed to minimize the impacts of this misclassification since misclassification errors are averaged out.

6.4. SHORE Analysis for Radon

We start off with all possible geological formation and age variables with buffer hyperparameters as possible explanatory variables. The spatial regression equation is given by Eq. 3, where the independent variable is set to the g_1-transform of radon concentrations, i.e. $Y_i = g_1(Z_i)$, where $g_1()$ is the log-transform, and Z_i are observed radon concentration values. We then use all the radon samples to implement the SHORE procedure described in Figure 3. This results in the selection of the best possible explanatory variables for point level groundwater radon.

To help determine how well the radon model predicts at unmonitored locations we also run the SHORE analysis with split training and validation sets. We randomly select 90 percent of the data to be in the training set, run the SHORE procedure using only the training set, and then use the selected model parameters to estimate at the validation set. We calculate the cross-validation correlation, $R_*^2 = r^2(Y_{obs}^{val}, \hat{Y}^{val})$, and the *shrinkage on cross-validation*, which is the percent difference in r^2 from the training set to the validation set.

6.5. Radon Probability Maps

While the Kriging and LUR-Kriging estimation methods can be used to produce maps of the median estimates of groundwater radon concentrations, these methods can also be used to produce maps of the probability that radon exceeds a given cutoff value, i.e. Prob(Radon>Z_c), where Z_c can be any cutoff value of interest. In this work we produce a map of the probability that groundwater Radon exceeds a cutoff value of Z_c=10,000pCi/L, the concentration at which the NC Radon in Water Advisory Committee (T. Campbell et al, 2011) recommends treatment of radon in water

7. Radon Results

7.1. SHORE Model for Radon

We conducted a SHORE analysis of all of the variables created. Table 9 shows the result for the best model obtained using all the radon samples. We see that 3 variables from class C (2nd most detail) variables are picked, one class D, and one Age class B. All of the variables are highly significant, and the maximum variance inflation factor is only 1.45 indicating the SHORE procedure is successful in avoiding over-fitting the model or producing high muliticollinearity.

Table 9. SHORE analysis results giving the best point level radon model.

Variable Name	Variable Class and Hyperparameter	Estimate (Beta)	pValue	R2 = 0.26 Max VIF = 1.45
Granitic Gneiss	Lithology C 28km	3.37	1e-92	
Granodiorite	Lithology C 5km	1.78	4.1e-34	
Late Cenozoic	Age B 46km	-1.73	2.1e-17	
Garnet Mica Schist	Lithology D 5km	32.00	7.7e-15	
Gneiss Schist	Lithology C 16km	3.47	9.1e-10	
Intercept	-	5.70	0	

Figure 10 displays the scatterplot of the observed versus LUR predicted radon. As expected with an r2 of 26%, this plot indicates that some areas have relatively large under and over predictions, but the plot also indicates that model is approximately equal in terms of the frequency at which it under and over predicts at sampled wells.

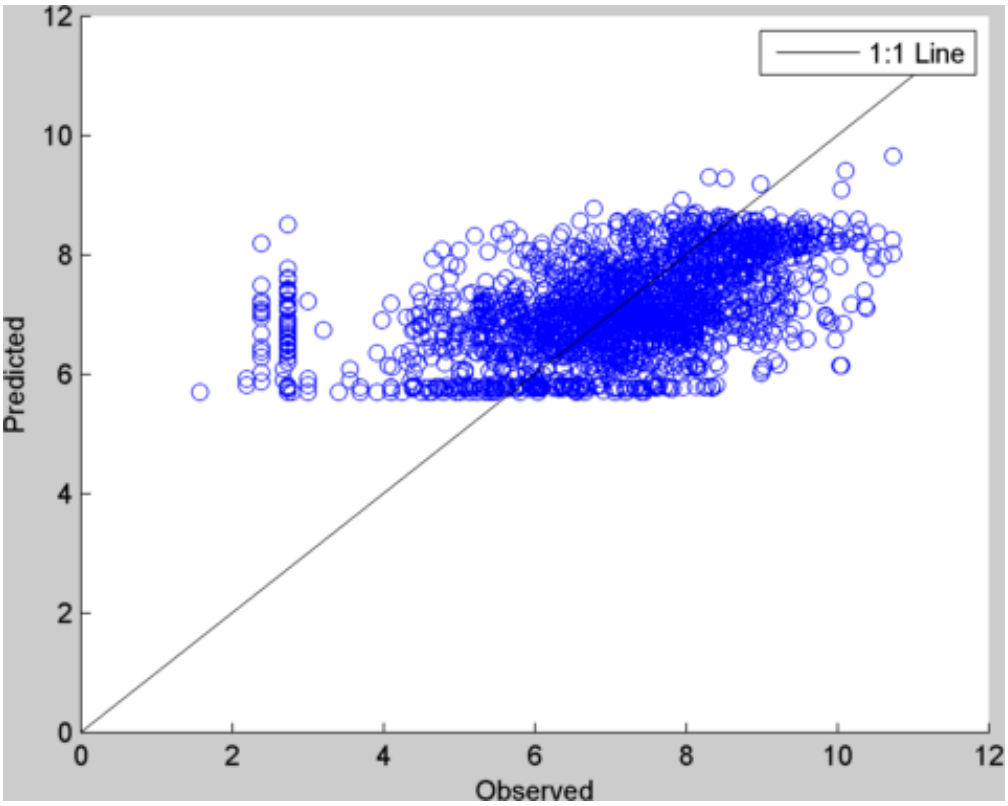


Figure 10. Scatterplot of observed versus LUR predicted log dissolved groundwater radon concentration.

7.2. SHORE Maps of Radon

We created a map of the modeled median groundwater radon based on the SHORE results, which is shown in Figure 11. It is clear that the majority of high groundwater radon is in the western portion of the state, however, there are geological features leading to potentially high groundwater radon in the piedmont region, particularly in the eastern portion of Wake County and in selected areas within Gaston, Cleveland, and Mecklenburg Counties.

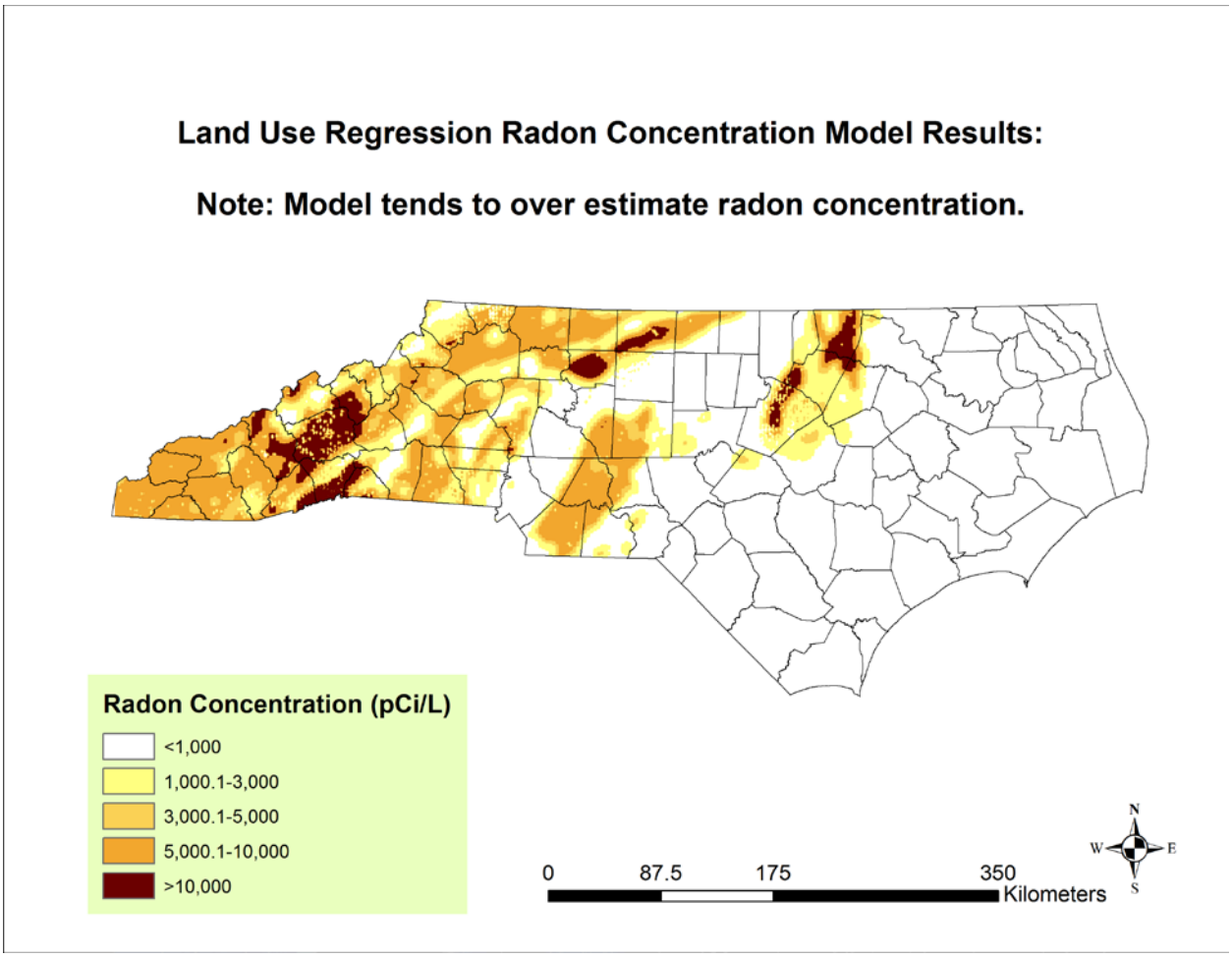


Figure 11. Land Use Regression Model Results for Groundwater Radon Median Across North Carolina. Note that this model over-estimates radon concentrations.

7.3. Radon LUR-Kriging Map

We also created the LUR-Kriging median estimate of groundwater radon concentration, which is shown in Figure 12. It is similar to Figure 11, however there are differences where there is monitoring data showing high values that are not represented in a map based purely on the LUR prediction. For instance, in Orange and Guilford counties there are sporadic high values that are not associated with the geological features that normally give rise to elevated radon, such as the variables in our final LUR model for radon.

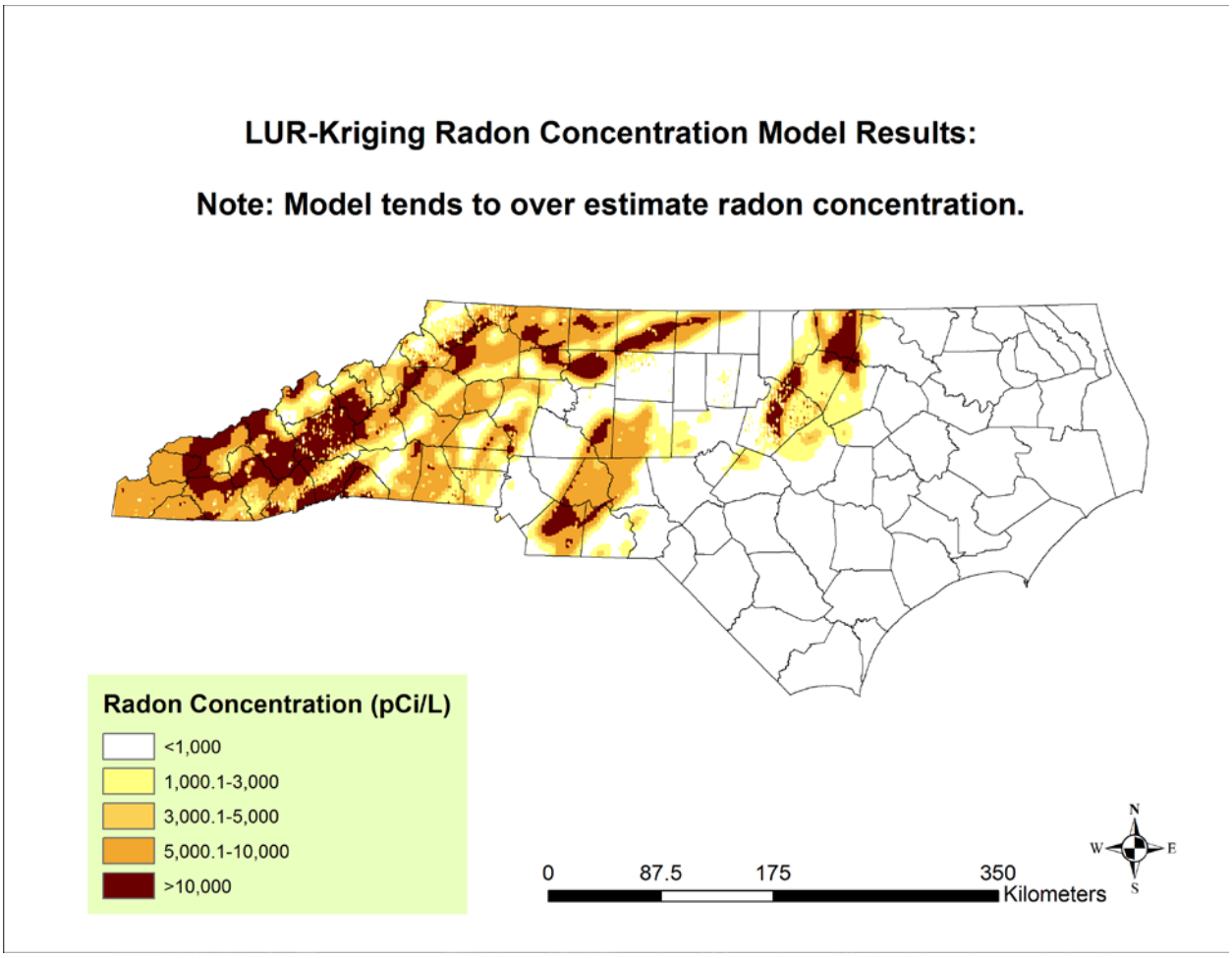


Figure 12. LUR-Kriging Model Results for Groundwater Radon Median Across North Carolina. Estimates are error prone, please see the error variance map.

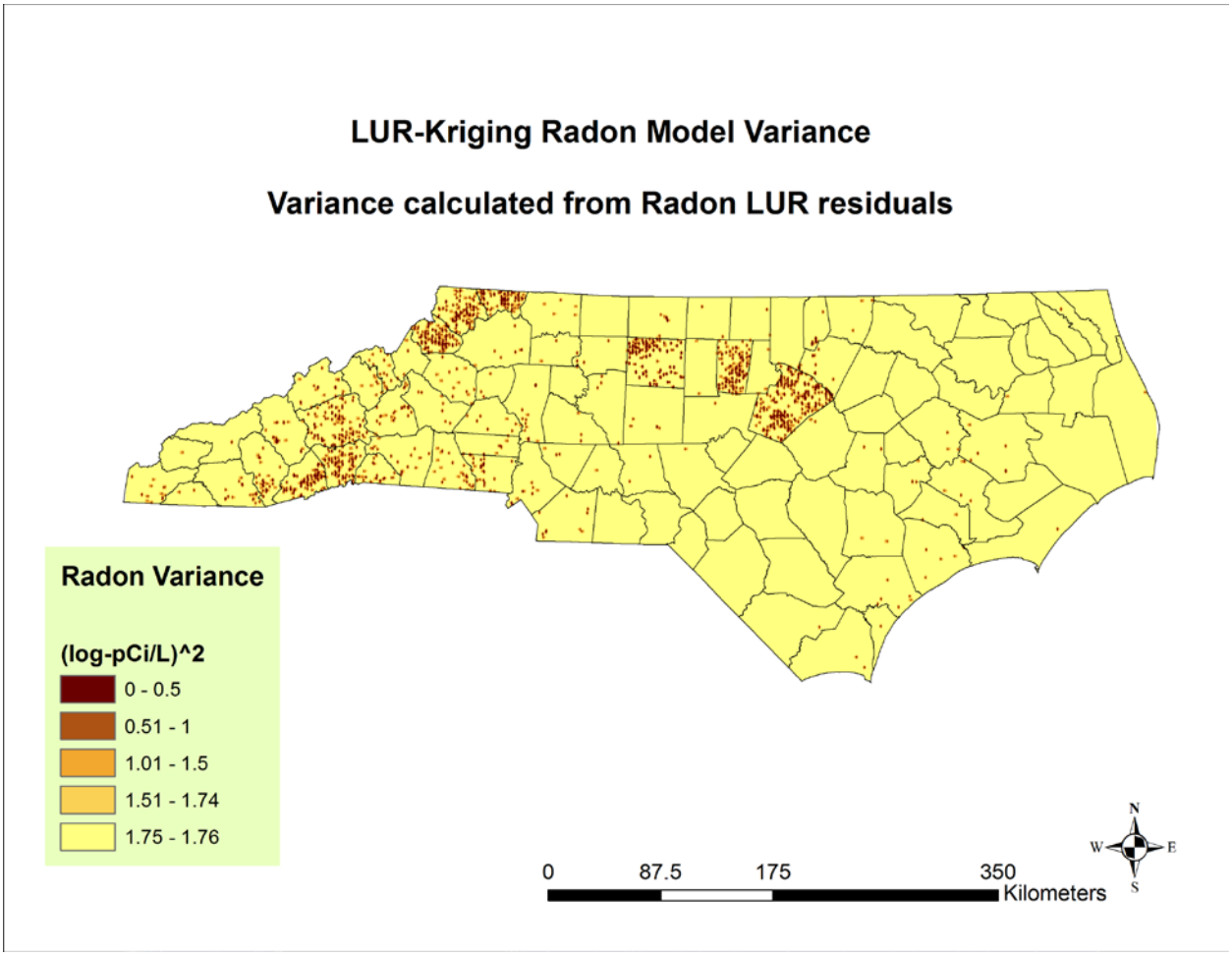


Figure 13. LUR-Kriging Radon Model Variance. The majority of estimations outside areas without monitoring result in a large error variance, thus are error prone estimates. The variance is calculated from the LUR residual, thus the modeled error variance is small in comparison to measured and model radon values.

Figure 13 displays the LUR-Kriging model variance which helps determine where the model estimates are reliable. Due to the relatively low R-squared value in our land use regression model and the sparseness of the monitoring data, a large portion of the state is in a high error area. All estimates where the radon residual variance is greater than or equal to 1.75 $(\text{log-pCi/L})^2$ are considered unreliable.

7.4. Radon Probability Map

Using the final LUR-Kriging estimates for the mean and variance from the SHORE model, we calculate the probability that groundwater radon will exceed 10,000 pCi/L. The results are shown in a map of the probability in Figure 14, which represents our best estimate of the probability dissolved groundwater radon exceeding 10,000 pCi/L. Please note that this represents the modeled estimate and not a true probability.

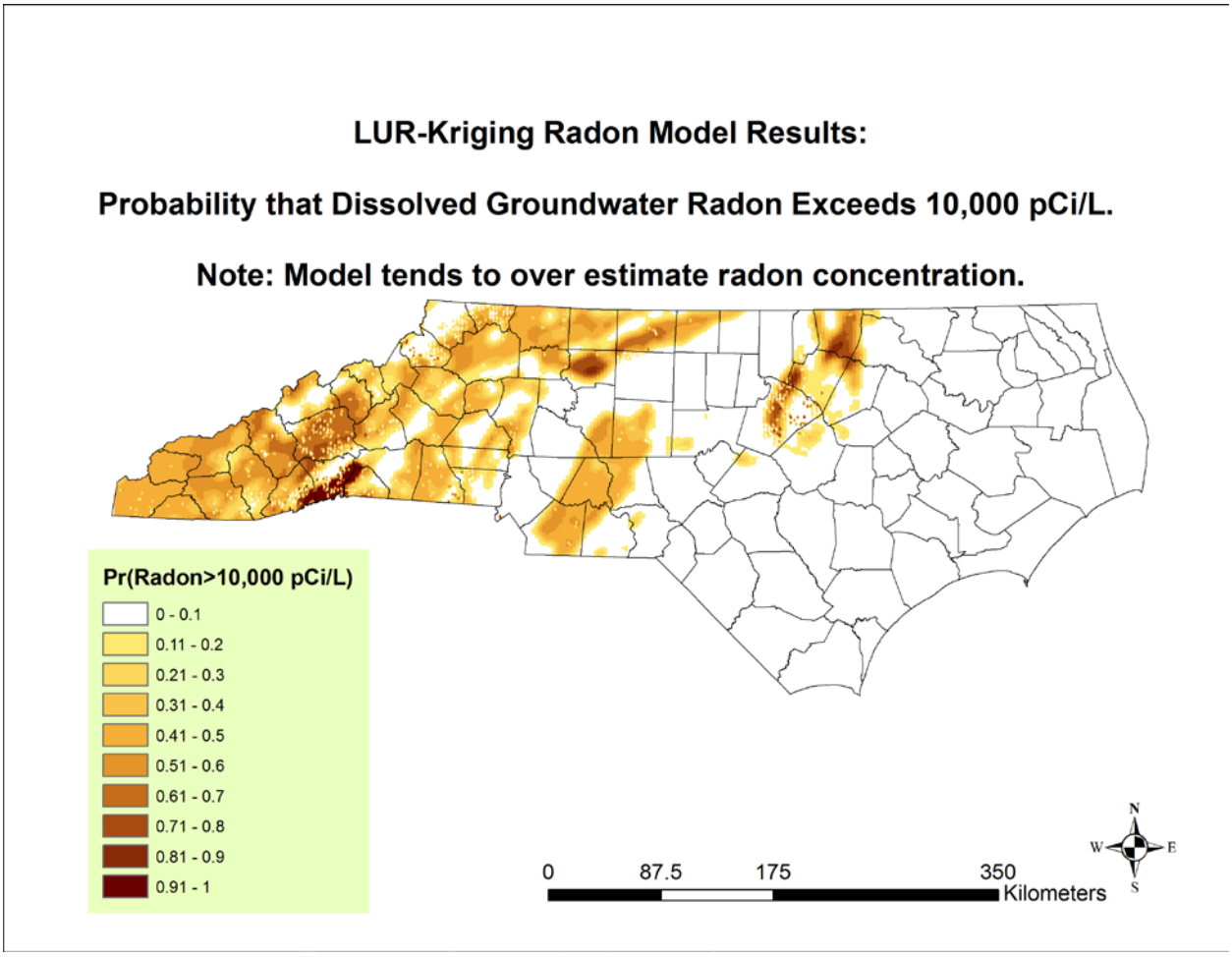


Figure 14. Probability of groundwater radon exceeding 10,000 pCi/L based on the LUR-Kriging model results.

7.5. Radon Cross-Validation Analysis

A leave one out cross validation analysis was conducted on the Kriging, land use regression, and LUR-Kriging models and the mean squared error was calculated. In this cross validation analysis, the land use regression model is calibrated using all the data, while for Kriging and LUR-Kriging, each observed value is removed in turn and estimated using the remaining data. The results are summarized in Table 10 along with percent difference. In the case of groundwater radon, Kriging alone performs the worst with a MSE of 4.25 $(\text{log-pCi/L})^2$. LUR alone changes the MSE by -59% (i.e a 59% reduction) compared to Kriging, so as to yield a MSE of 1.76 $(\text{log-pCi/L})^2$. LUR-Kriging further changes the MSE by -15% relative to LUR alone resulting in a reduced MSE of 1.49 $(\text{log-pCi/L})^2$.

Table 10. Cross-Validation results for Kriging, Land Use Regression, and LUR-Kriging methods.

	Kriging	Land Use Regression	LUR-Kriging
Mean Squared Error (log-pCi/L)^2	4.25	1.76	1.49
Percent change		-59%	-15%

In order to refine our analysis of the performance of land use regression alone, we then performed the validation analysis consisting in randomly selecting 90% of the data as the training set, and using the remaining 10% of the data as the validation set. Three iterations of this analysis were performed. For each of these iterations, the training set was used by the SHORE procedure to select a model, and that model was used to estimate Radon at the validation set, which was not used for calibration of the model. For each iteration, we calculate the cross-validation correlation and shrinkage on cross-validation. The results for each run are summarized in Table 11. A scatterplot of the third iteration validation set actual versus predicted values are shown in Figure 15.

Table 11. Training and Validation SHORE results for 3 iterations.

Run	Variable Name	Variable Class and Hyperparameter	Training R-Squared	Cross-Validation Correlation (Validation R-Squared)	Shrinkage on Cross Validation (% reduction)
1	Granitic Gneiss	C , 28km	0.21	0.16	25%
	Gneiss-Schist	C, 49km			
2	Granitic Gneiss	C, 28km	0.23	0.17	24%
	Gneiss-Schist	C, 49km			
	Late Proterozoic-Early Paleozoic	Age B, 15km			
3	Granitic Gneiss	C, 28km	0.23	0.24	-3%
	Gneiss-Schist	C, 49km			
	Late Proterozoic-Early Paleozoic	B, 24km			
	Felsic mica gneiss	D, 24km			

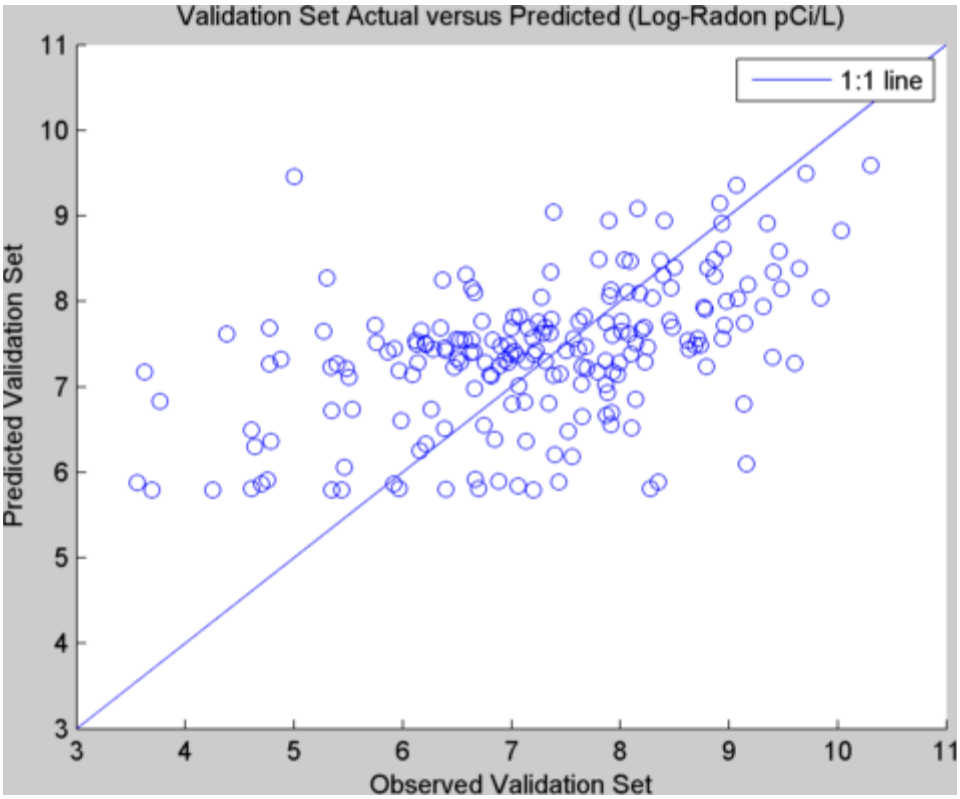


Figure 15. Scatterplot of the third iteration training and validation SHORE iteration. Validation set (n=200, 10%) actual versus predicted.

8. Discussion

8.1. Novel Contributions to Radon Spatial Modeling

This study of groundwater radon in North Carolina resulted in a substantial increase in the coverage of radon data and related geological predictors across the State, and provided the first state-wide estimates of the groundwater radon median and probability of exceeding 10,000 pCi/L. Furthermore, it also provided the first model that incorporates geology with positive results allowing for predictions state-wide. Given the results of the land use regression model and the scale of the geology data, our model predictions represent radon at approximately a 5 kilometer scale.

The maps of the median estimate groundwater radon will be useful to epidemiological researchers for assessing exposure to groundwater radon in North Carolina, and for risk assessors for quantifying the burden of disease attributable to radon exposure in North Carolina. For example investigators of this project have obtained data from the North Carolina Central Cancer Registry on lung and stomach cancer in North Carolina in a proposal to investigate the association between these cancers and radon exposure in North Carolina.

Furthermore the map showing the probability that radon exceeds 10,000 pCi/L will be useful to state personnel in determining where more monitoring is needed and where to allocate resources in community outreach. The probability P(Radon>10,000 pCi/L) generally increases in areas with rocks commonly associated with elevated radon and additional monitoring is needed in these and other areas. Hence the probability map provides a useful metric to guide where new groundwater samples should be collected and tested for radon. We recommend that the map be used in conjunction with previous DWQ maps to help guide further sampling for research.

8.2. New Data Sources

By obtaining groundwater radon monitoring data from two private companies, we almost tripled the geocoded groundwater radon monitoring data, which considerably widened the geographical coverage of sampled locations (Figure 8). This in turn increased the range of geological formations that were sampled, which improved the power of SHORE to select geological variables that can predict radon across the whole state. Another new data source contributed by this work comes from the calculation across the whole state of these predictive geological variables. The calculated values of these geological variables can be shared with the scientific community and interested stakeholders for future research.

8.3. Point Level Radon Analysis

The association between groundwater radon and the underlying geology was well established; however, this is the first study to quantify that association in a way that explains small area variability of groundwater radon. The results of this study are useful for assessing exposure at an individual level over a large study area, which is needed for a state-wide exposure assessment. As mentioned in section 8.1, based on the results of the LUR model, and the map scales of the geology data, our analysis leads to estimates that are approximately 5 kilometers in scale. This is due to the fact that the minimum differential distance from the geology data is 250 meters (1:125,000 scale), and the minimum hyperparameter distance in model results is 5 kilometers.

Results from the land regression analysis were obtained using data that had some limitations. As described earlier, a limitation of the private well data is that sampling by inexperienced private well owner could lead to off-gassing, which would result in concentrations biased toward low values, and not high values. Another noted limitation is the possibility for geology misclassification. However; these two potential errors would lead to land use regression toward the null hypothesis. In other words, in the presence of such errors, then the model might lead to a type II error (regression false negative) and not type 1 error (regression false positive). Therefore the associations found by the model are robust in the sense that any associations that are found to be statistically significant under the presence of these errors would presumably be even more statistically significant (i.e. have a lower p-value) without such errors.

The SHORE procedure leads to the selection of some geological variables that are consistent with previous literature and expertise of field scientists, however it also contains some variables that are not expected and are selected potentially due to data and model limitations. For instance, Garnet Mica Schist is not normally associated with elevated dissolved radon, but it was likely selected due to its proximity to high values overlaying geology more commonly associated with elevated radon. Nonetheless, the model results lead to a better understanding of the possible scales at which the geology is impacting the groundwater radon concentration. Henderson Gneiss geological formation is a known hot spot for high groundwater radon concentration in western North Carolina. Henderson Gneiss was available as a potential explanatory variable as a class D variable, the variable class that is most detailed. However, granitic gneiss, a class C variable, was favored ahead of Henderson Gneiss, which is a more generalized classification of Henderson Gneiss. Furthermore, the granitic gneiss was selected with a buffer of 28km which means large area variability is perhaps over powering any small area variability in groundwater radon.

The validation analysis using separate training and validations sets shows that the model produces results in the validation set that are consistent with those obtained in the training set. The shrinkage on cross-validation is usually equal to about 25%, though there was an iteration that produced a validation r-squared slightly greater than training set (Figure 15), resulting in a negative shrinkage. The random iterations of the training sets produced different models, but the most consistent variable selected was the Granitic Gneiss.

The LUR model is relatively unbiased in predicting radon concentrations at measured sample locations as evidenced by the Figure 10 scatterplot . Furthermore, the validation set cross-validation correlation demonstrates that near current monitoring locations, the model is producing unbiased prediction. However, we believe that the LUR and LUR-Kriging models could be systematically overpredicting in most cases in areas far away from measured values where the prediction is completely reliant on the LUR model. By examining the median of observed values by county versus the median of LUR-Kriging estimates by county (Figure 16) we see that there are more counties with the predicted median higher than the observed median, although the r-squared statistic for observed versus LUR-Kriging predicted medians by county is 0.53. Although this comparison is not a truly fair comparison since there are many more LUR model estimates in each county than measured values, the results give us some reasons for caution in the current model results. The finding that the model predicts higher values in unsampled areas than what was observed in the sampled wells goes against our a priori belief that most of the radon hot spots in North Carolina have already been sampled. The LUR model is either selecting erroneous variables through the SHORE procedure, a potential drawback of any stepwise selection method, or there are in fact many unsampled areas across western North Carolina with high dissolved groundwater radon, an unlikely scenario under our current belief. We therefore believe that the presence of the gneiss-schist formation in the final LUR model is driving up high values in areas away from wells that were sampled. Gneiss-schist is a general enough geological type that it is widespread throughout western and piedmont North Carolina.

Further research is needed to refine the results such that a geology type that is less prominent is selected in the SHORE model while still maintaining a reasonable r-squared.

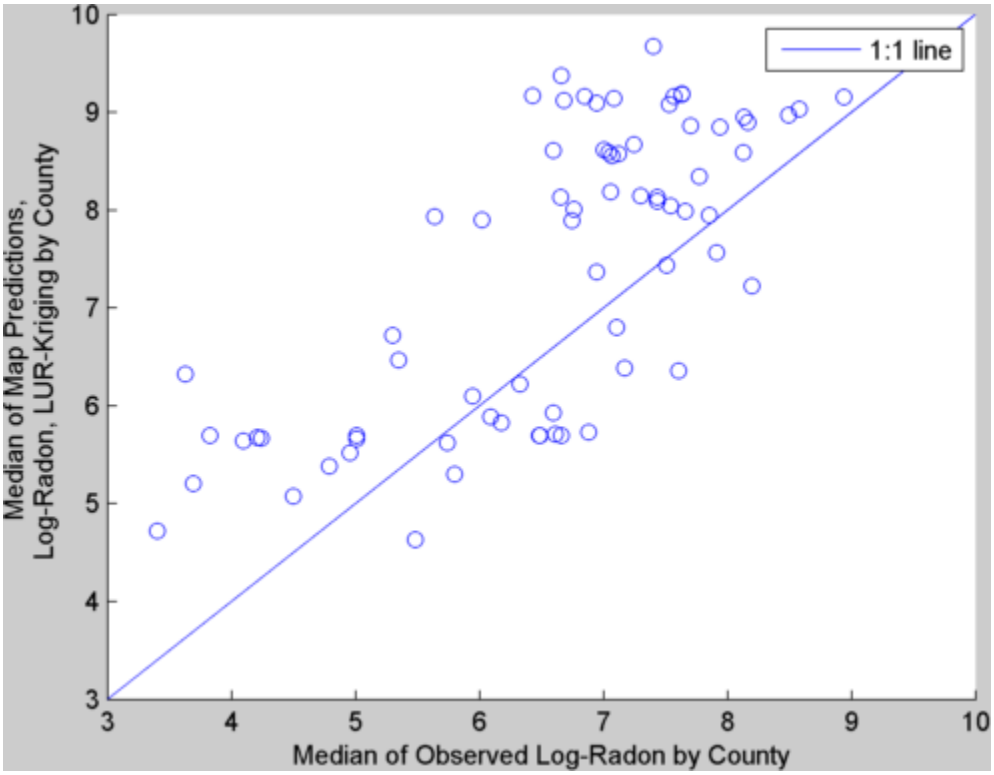


Figure 16. Scatterplot of median observed radon and median predicted radon (All map locations).

Further research will include accounting for anisotropy in the geological variables by creating variables that are the percent of a geological feature within an ellipse as opposed to a circle. The ellipse will be allowed to vary in both the major and minor radii and the angle. Since most of North Carolina's geological features run on an angle from the southwest to the north east, we anticipate these variables will better be able to capture the physical reality of the underlying geology. Further research will also investigate potential nonlinear regression models and varying nonlinear models in SHORE.

Overall, the SHORE model selected some variables consistent with literature and the knowledge of field experts while offering novel insights into the length scales of significance. Furthermore, the LUR-Kriging procedure allows for a significant improvement over that of LUR or Kriging alone in the accuracy of groundwater radon estimation across North Carolina.

9. Summary

This study created a substantial increase in available data for groundwater nitrate and radon and modeled them at a point level scale using a stepwise hyperperameter optimization regression procedure combined with a Kriging analysis. The results led to the first maps of point level estimates for both groundwater contaminants across the entire state.

The first objective of the study was to determine all of the available data sources for groundwater nitrate and radon in North Carolina. In both cases, our collaborative efforts with NC DENR led to the inclusion of high quality monitoring data. We further added to both nitrate and radon datasets by implementing a multi-stage address geocoding procedure. In the nitrate case, we address geocoded private well samples, and in the radon case we geocoded private well samples of houses.

The second objective of the study was to create a land use regression model for both nitrate and radon. The land use regression model for point level nitrate resulted in a total r-squared of 0.18 with the following explanatory variables: The sum of exponentially decaying contributions from treated sewage sludge field application sites with amount of fertilizer estimated by the application field acreage and an exponential decay range of 2000 meters, the percent of crops within a 100 meter buffer, the percent of metaigneous geology within a 1000 meter buffer, the percent of deciduous land within a 20 meter buffer, a generalized cubic trend of nitrate, and the percent of evergreen forest within a 100 meter buffer. The land use regression for point level radon resulted in a total r-squared of 0.26 with the following explanatory variables: The percent of granitic gneiss within 28km buffer, the percent of granodiorite within a 5km buffer, the percent of late Cenozoic within a 46km buffer, the percent of garnet mica schist within a 5km buffer, and the percent of gneiss schist within a 16km buffer.

The final objective of the study was to integrate the LUR models for each contaminant into the Kriging method of geostatistics and create maps of the estimates across North Carolina. The LUR-Kriging results were mapped and displayed in the results section of the study. Furthermore, the LUR-Kriging procedure resulted in reduced estimation error in terms of the cross-validation mean squared error for both nitrate and radon. For nitrate the Kriging MSE was 0.267 (log-mg/L)^2 while the LUR-Kriging MSE was 0.253 (log-mg/L)^{2}which is a percent difference of -5.2%. For above detect data only in the cross-validation, the Kriging MSE was 0.504 (log-mg/L)^2 while the LUR-Kriging was 0.457 for a percent difference of -9.3%. For radon the Kriging mean squared was 4.25 (log-pCi/L)^2 and the LUR model only was 1.76 (log-pCi/L)^2 which is a percent difference of -59%. The mean squared error is further improved in the LUR-Kriging with a MSE of 1.49 (log-pCi/L)^2 , which is a percent change over the LUR model of -15%.

10. Conclusions

The SHORE procedure created a land use regression model describing the small area variability of point level groundwater nitrate that is a significant improvement over models that rely on spatially and temporally averaged nitrate data. The results are successfully integrated into the Kriging method resulting in a significant improvement in cross-validation mean squared error indicating a more accurate model.

Similarly, the SHORE procedure created a land use regression model for point level groundwater radon that results in the first model of groundwater radon across the state using geological information. The results are also successfully integrated into the Kriging method producing more accurate results.

11. Recommendations

We recommend for our maps of groundwater nitrate and radon to be used by staff at the NC DENR in determining where additional resources are allocated for monitoring, community outreach, and policy-making decisions. The nitrate map can be used to better help protect communities in potentially high nitrate areas. Since the nitrate map cannot distinguish between confined and unconfined aquifers at a given location, the maps represent a worst-case scenario. Thus the maps can be used to determine where possible groundwater nitrate pollution problems exist regardless of depth.

The map of radon mean concentration and the probability map of exceeding 10,000 $(pCi/L)^2$ can be used in conjunction with current DWQ maps to help determine future sampling and educational outreach and to further refine our understanding of dissolved radon across NC.

12. References

Agency, U.S.E.P., 1999. Standard Method 7500-Rn. *Standard Methods for the Examination of Water and Wastewater*, 19th editi.

Auvinen, Anssi et al., 2005. Radon and other natural radionuclides in drinking water and risk of stomach cancer: a case-cohort study in Finland. *International journal of cancer*, 114(1), pp.109–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15523702 [Accessed March 3, 2013].

Campbell, T. et al., 2011. North Carolina Radon-in-Water Advisory Committee Report. *North Carolina Department of Environment and Natural Resources, Division of Water Quality*.

Christakos, G., Bogaert, P. & Serre, M. L., 2002. *Temporal GIS: Advanced Function for Field-Based Applications*, New York, NY: Springer.

Christakos, George, 1990. A Bayesian/maximum-entropy view to the spatial estimation problem. *Mathematical Geology*, 22(7), pp.763–777. Available at: http://www.springerlink.com/index/10.1007/BF00890661.

Cressie, N. & Majure, J.J., 2012. Spatio-Temporal Statistical Modeling of Livestock Waste in Streams. , 2(1), pp.24–47.

Darby, S. et al., 2004. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ (Clinical research ed.)*, 330(7485), p.223. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=546066&tool=pmcentrez&rendertype=abstract [Accessed July 27, 2011].

Fan, a M. & Steinberg, V.E., 1996. Health implications of nitrate and nitrite in drinking water: an update on methemoglobinemia occurrence and reproductive and developmental toxicity. *Regulatory toxicology and pharmacology : RTP*, 23(1 Pt 1), pp.35–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8628918.

Fitzgerald, B. & Hopke, P.K., 1997. Experimental Assessment of the Short- and Long-Term Effects of Rn from Domestic Shower Water on the Dose Burden Incurred in Normally Occupied Homes. , 31(6), pp.1822–1829.

Giese, G.I., Eimers, J.L. & Coble, R.W., 1993. Simulation of ground-water flow in the Coastal Plain system of North Carolina. *US Geological Survey Professional Paper 1404-M*, p.142.

Gurdak, J.J. & Qi, S.L., 2012. Vulnerability of Recently Recharged Groundwater in Principle Aquifers of the United States To Nitrate Contamination. *Environmental science & technology*, (3). Available at: http://www.ncbi.nlm.nih.gov/pubmed/22582987.

Hengl, T., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120(1-2), pp.75–93. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0016706103002787 [Accessed October 8, 2010].

Howarth, R.W. et al., 1996. Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: Natural and human influences. *Biogeochemistry*, 35(1), pp.75–139. Available at: http://www.springerlink.com/index/10.1007/BF02179825.

Keil, A., Wing, S. & Lowman, A., 2011. Suitability of Public Records for Evaludating Health Effects of Treated Sewage Sludge in North Carolina. , 72(2), pp.98–104.

Kennedy, C.D. et al., 2009. Spatial and temporal dynamics of coupled groundwater and nitrogen fluxes through a streambed in an agricultural watershed. *Water Resources Research*, 45(9),

pp.1–18. Available at: http://www.agu.org/pubs/crossref/2009/2008WR007397.shtml [Accessed May 1, 2012].

Kenny, J. et al., 2005. Estimated Use of Water in the United States in 2005. *USGS Circular 1344*.

Kotrappa, P. & Jesters, W.A., 1993. Electret ion chamber radon monitors measure dissolved 222Rn in water. *Health Physiscs*, 64.4, pp.397–405.

Krewski, D. et al., 2005. Residential Radon and Risk of Lung Cancer. *Epidemiology*, 16(2), pp.137–145. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00001648-200503000-00001 [Accessed January 11, 2012].

Loomis, D.P., 1987. Radon-222 Concentration and Aquifer Lithology in North Carolina. *Groundwater Monitoring and Remediation*.

Lubin, J H & Boice, J.D., 1997. Lung cancer risk from residential radon: meta-analysis of eight epidemiologic studies. *Journal of the National Cancer Institute*, 89(1), pp.49–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8978406.

Mcelroy, J.A. et al., 2011. Geocoding Addresses from a Large Population-Based Study : Lessons Learned Published by : Lippincott Williams & Wilkins Stable URL : http://www.jstor.org/stable/3703788 . Your use of the JSTOR archive indicates your acceptance of JSTOR ' s Terms and Condit. *North*, 14(4), pp.399–407.

McLay, C.D. et al., 2001. Predicting groundwater nitrate concentrations in a region of mixed agricultural land use: a comparison of three approaches. *Environmental pollution (Barking, Essex : 1987)*, 115(2), pp.191–204. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11706792.

Messier, K.P., Akita, Y. & Serre, Marc L, 2012. Integrating address geocoding, land use regression, and spatiotemporal geostatistical estimation for groundwater tetrachloroethylene. *Environmental science & technology*, 46(5), pp.2772–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22264162.

De Nazelle, A., Arunachalam, S. & Serre, Marc L, 2010. Bayesian maximum entropy integration of ozone observations and model predictions: an application for attainment demonstration in North Carolina. *Environmental science & technology*, 44(15), pp.5707–13. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2912419&tool=pmcentrez&rendertype=abstract.

NCGICC, 2013. ncONEmap. Available at: www.nconemap.com [Accessed January 3, 2012].

Nolan, B. & Stoner, J., 2000. Nutrients in Groundwaters of the Conterminous. , pp.1992–1995.

Nolan, B.T. & Hitt, K.J., 2006. Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States. *Environmental science & technology*, 40(24), pp.7834–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17256535.

Olea, R. a., 2006. A six-step practical approach to semivariogram modeling. *Stochastic Environmental Research and Risk Assessment*, 20(5), pp.307–318. Available at: http://www.springerlink.com/index/10.1007/s00477-005-0026-1 [Accessed May 8, 2011].

Pavia, M. et al., 2003. Meta-analysis of residential exposure to radon gas and lung cancer. *Bulletin of the World Health Organization*, 81(10), pp.732–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2572329&tool=pmcentrez&rendertype=abstract.

Qian, S.S., 2005. Nonlinear regression modeling of nutrient loads in streams: A Bayesian approach. *Water Resources Research*, 41(7), pp.1–10. Available at: http://www.agu.org/pubs/crossref/2005/2005WR003986.shtml [Accessed May 25, 2011].

Serre, M. L. & Christakos, G., 1999. Modern geostatistics: computational BME analysis in the light of uncertain physical knowledge - the Equus Beds study. *Stochastic Environmental Research and Risk Assessment (SERRA)*, 13(1-2), pp.1–26. Available at: http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/s004770050029.

Smith, R.A., Schwarz, G.E. & Alexander, R.B., 1997. Regional interpretation of water-quality monitoring data. , 33(12), pp.2781–2798.

Spalding, R.F. & Exner, M.E., 1993. Occurrence of Nitrate in Groundwater—A Review. *Journal of Environment Quality*, 22(3), p.392. Available at: https://www.agronomy.org/publications/jeq/abstracts/22/3/JEQ0220030392.

Spruill, T., 2004. Geochemistry and Characteristics of Nitrogen Transport at a Confined Animal Feeding Operation in a Coastal Plain Agricultural Watershed , and Implications for Nutrient Loading in the Neuse River Basin , North Carolina , 1999-2002. , pp.1999–2002.

Su, J.G., Jerrett, M. & Beckerman, B., 2009. A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. *The Science of the total environment*, 407(12), pp.3890–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19304313.

Tague, C.L. & Band, L.E., 2004. RHESSys: Regional Hydro-Ecologic Simulation System—An Object-Oriented Approach to Spatially Distributed Modeling of Carbon, Water, and Nutrient Cycling. *Earth Interactions*, 8(19), pp.1–42. Available at: http://journals.ametsoc.org/doi/abs/10.1175/1087-3562(2004)8<1:RRHSSO>2.0.CO;2.

Tesoriero, A.J. et al., 2013. Vulnerability of Streams to Legacy Nitrate Sources. *Environmental science & technology*.

USGS, 2013. Mineral Resources On-Line Spatial Data. *mrdata*. Available at: www.mrdata.usgs.gov/geology/state [Accessed January 3, 2012].

Vinson, D.S., Campbell, T.R. & Vengosh, A., 2008. Radon transfer from groundwater used in showers to indoor air. *Applied Geochemistry*, 23(9), pp.2676–2685. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883292708002114 [Accessed March 7, 2013].

WHO, 2011. *WHO guidelines for drinking-water quality.*, Available at: http://www.ncbi.nlm.nih.gov/pubmed/15806952.

13. Appendix 1: List of Abbreviations and Symbols

BME – Bayesian Maximum Entropy

CAFO – Concentrated Animal Feeding Operation

DENR – Department of Environment and Natural Resources

DHHS – Department of Health and Human Services

DWR – Division of Water Resources

LUR – Land Use Regression

mg/L- Milligrams per Liter

NCOD – National Contaminant Occurrence Database

NCGS – North Carolina Geological Survey

NLCD – National Land Cover Database

NO3 – Nitrate

pCi/L- Picocuries per Liter

USGS – United States Geological Survey

14. Appendix 2

1) There are no publications in referred journals concerning this work to date. We intend to write and submit a paper for both the nitrate and radon sections of this study. We have already discussed matters with our NC DENR collaborators to participate as co-authors on the papers.

2) Conference Publications:

Messier, K.P.; Akita, Y.;Campbell, T.; Serre, M.L.; You're too Gneiss, You take me for Granite: Preliminary Geology-based Land Use Regression and Kriging Analysis of Groundwater Radon Across North Carolina. In: North Carolina Water Resource Research Institute Annual Conference, March 20-21, 2013 . OralPresentation.

Messier, K.P.; Akita, Y.; Bolich, R.; Kane, E.; Serre, M.L.; Preliminary Results of Land Use Regression and Kriging Analysis of Groundwate Nitrate Across North Carolina. In: North Carolina Water Resource Research Institute Annual Conference, March 20-21, 2013 . Poster Presentation.

Messier, K.P.; Akita, Y.; Bolich, R.; Kane, E.;Campbell, T.; Serre, M.L.; Building a North Carolina groundwater nitrate database using multiple data sources and land use regression modeling. In: North Carolina Water Resource Research Institute Annual Conference, March 27-28, 2012. Oral Presentation.

Messier, K.P.; Akita, Y.; Bolich, R.; Kane, E.;Campbell, T.; Serre, M.L.; Multiple North Carolina groundwater radon data sources and correlations with rock types. In: North Carolina Water Resource Research Institute Annual Conference, March 27-28, 2012. Oral Presentation.

3) After publication of our results in peer-reviewed journals, we plan on making the nitrate data available to researchers with permission of the PI for this project. It is not in our data use agreement to allow release of the geocoded radon data. We intend to maintain communication with our collaborators at NC DENR to allow communication of results to NC DENR stake holders and North Carolina citizens.