Spatio-Temporal Modeling of Nitrate Levels in Iowa Municipal Drinking Water Supplies

By: Marina Antillon

Mentor: Brian Smith

Introduction

- Farming activity helps raise the level of nitrate in the groundwater.
- Nitrate is broken down in the body in this way:

$$NO_3 \xrightarrow{\text{Bacterial Flora}} NO_2^- \xrightarrow{\text{Digestive Processes}} \text{N-nitroso compounds}$$

- N-nitroso compounds have caused liver damage, internal bleeding, and cancer in experimental organisms.

Introduction

- There is interest in studying the potential carcinogenic effects of nitrate exposure from drinking in human populations.
- Estimates of nitrate levels in water are needed for retrospective epidemiologic cancer studies.
- Finding reliable models to estimate the exposure of Iowans to nitrate is an important biostatistical problem.

Previous Study

- Weyer et al. analyzed Iowa alluvial groundwater to examine the use of raw water nitrate levels of retrospective exposure estimates.
- They found that

  NO_3 from raw water = NO_3 from finished water

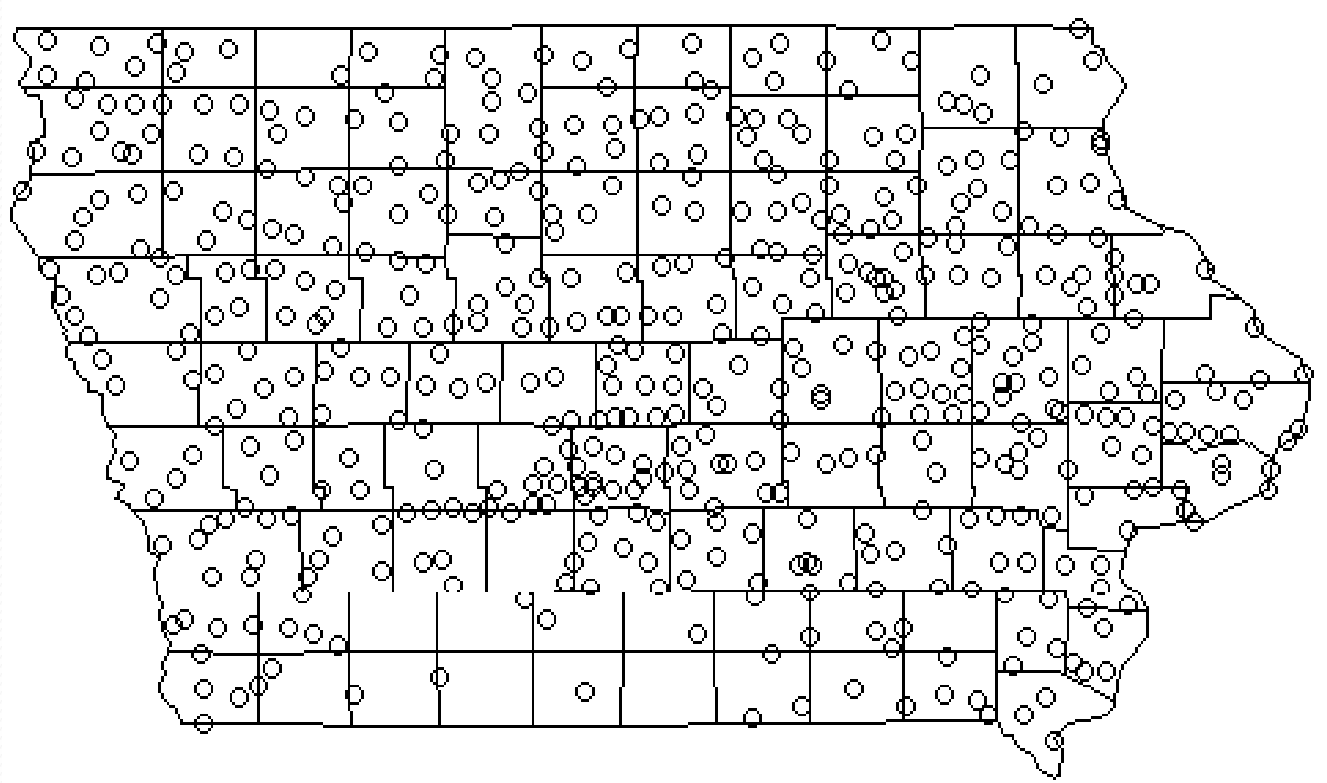
  in separate municipalities.
- However, if you aggregate these sites into a region, only some regions have nitrate raw water measurements that can supplement for finished water measurements.

Let's Expand the Model

- In this project, we want to examine the effect on the level of nitrate in water based on:
  - Decade
  - Depth
  - Season
  - Treatment
- The dataset analyzed by Weyer et al. is a smaller subset of the one that we use.

Data

- From the University of Iowa's Center for Health Effects of Environmental Contamination (CHEEC)
- Data for 500 different locations around Iowa
- 25,390 observations.



Model

- Let (NO_3) correspond to the measurement of milligrams of nitrate per liter of water at geographic location *s*. Then our analytic model is

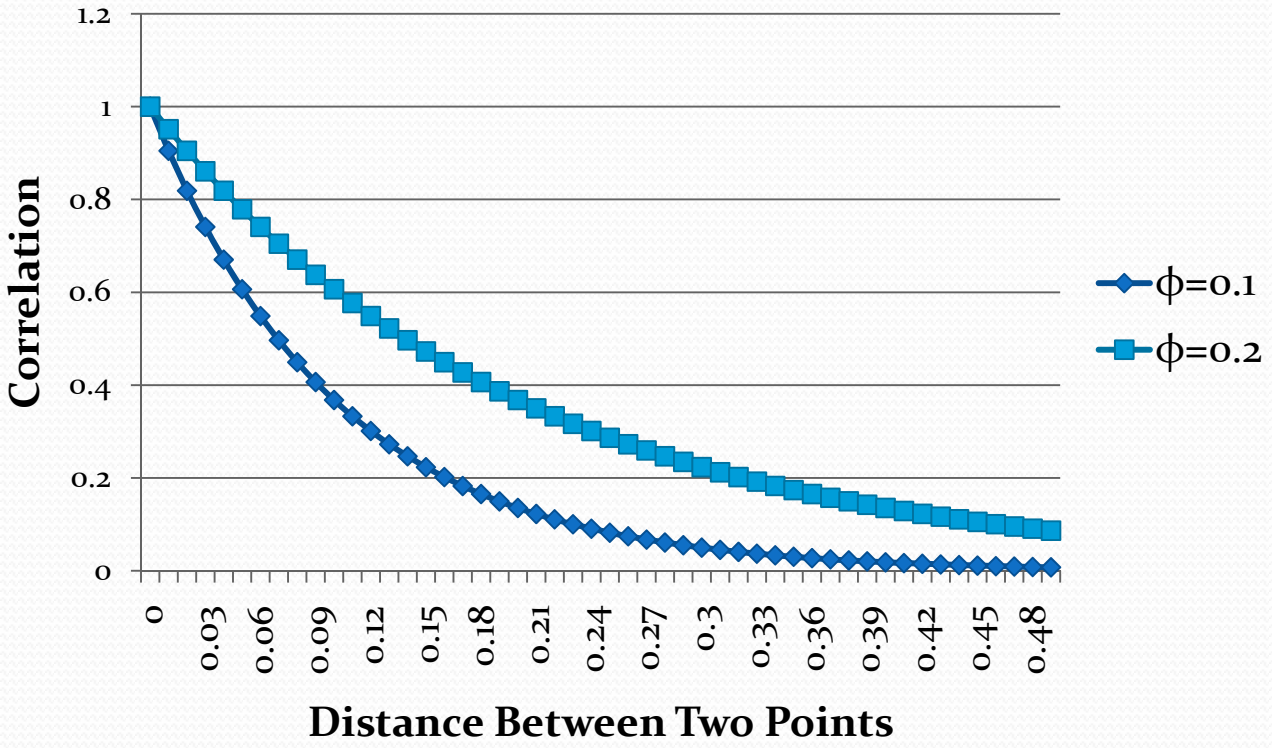
$$\ln(NO_3) = X^T \beta + Z(s) + \varepsilon$$

  - where X is a vector of indicator variables for decade, depth, season, and treatment.
  - β is a vector of mean parameters for each indicator variable to be estimated from the data.
  - $Z(s)$ is a normal random variable that accounts for spatial correlation between nitrate concentrations.
  - ε is an independent normal error term with σ_ε^2 variance. It accounts for errors in measurement and other random sources of variability.

Spatial Correlation

- The assumed exponential spatial correlation depends on two factors: the rate of decay and the distance.

$$\text{corr}(Z(s_i), Z(s_j)) = \exp\{-d_{ij}/\phi\}$$



RAMPS

- We used ramps, a geostatistical package for R statistical software developed by Smith, Yan, and Cowles.
- The package can fit a Bayesian model, provide posterior distributions for the model parameters, predict the amount of nitrate in the water across the state.
- ramps is designed for the analysis of large spatial data sets.

Bayesian Approach

- Goal is to estimate the joint posterior distribution:

$$P(\theta|NO_3) = \frac{P(NO_3|\theta)\ P(\theta)}{P(NO_3)}$$

Where θ is the vector of parameters that we want to estimate (β, σ_ε^2, σ_Z^2, ϕ).

- The software uses a Markov chain Monte Carlo (MCMC) computational method to repeatedly draw samples from the posterior distribution.
- The sampled parameter values are used to obtain posterior summary statistics (means, s.d., etc).

Prior Distributions

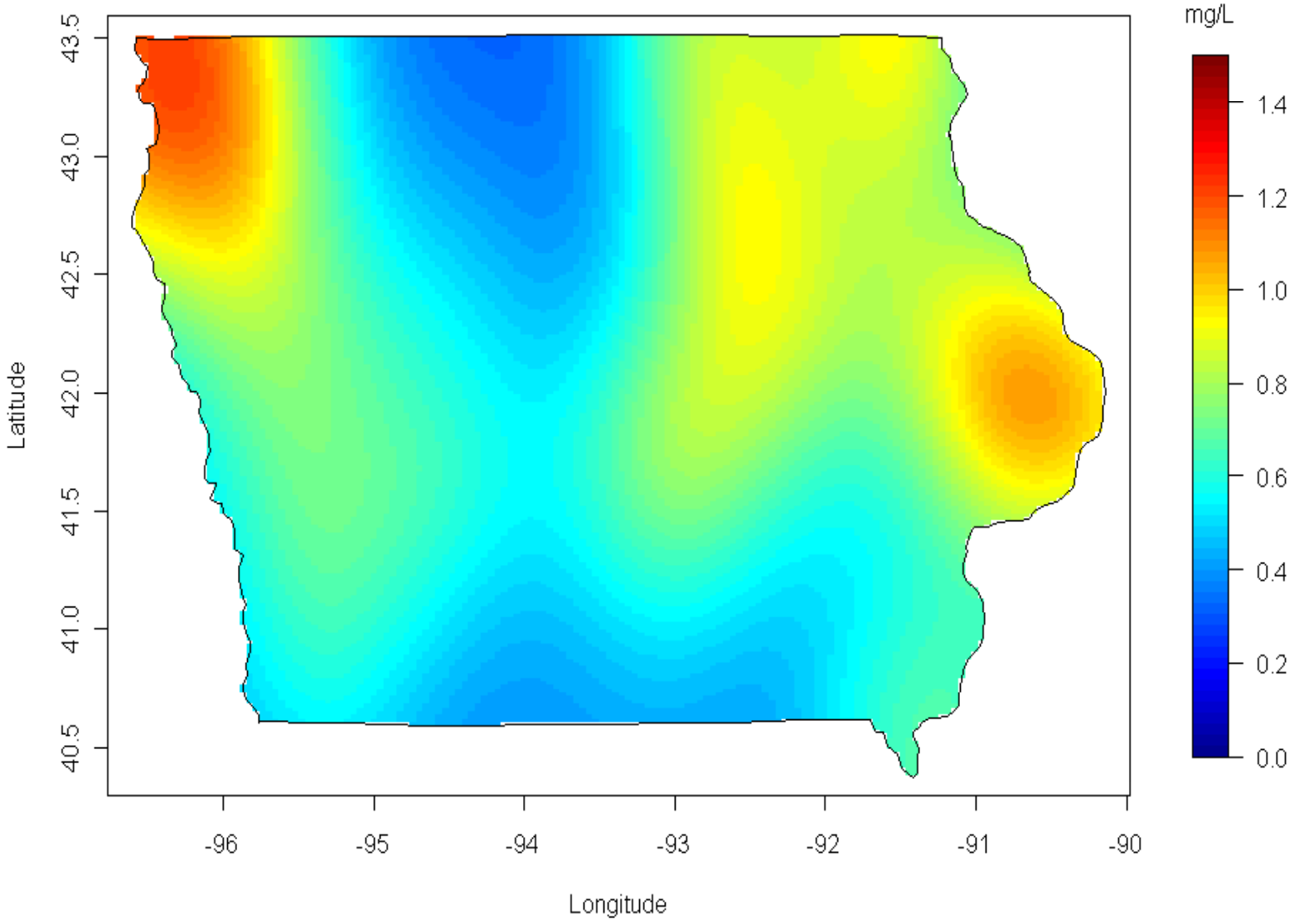
• Mean	β	"Flat" (Noninformative)
• Error Variance	$1/\sigma_\varepsilon^2$	$\Gamma(2, 10)$
• Spatial Decay	ϕ	Uniform (0,35)
• Spatial Variance	$1/\sigma_Z^2$	$\Gamma(2, 10)$

- We generated 1,000 MCMC draws from the posterior distribution.

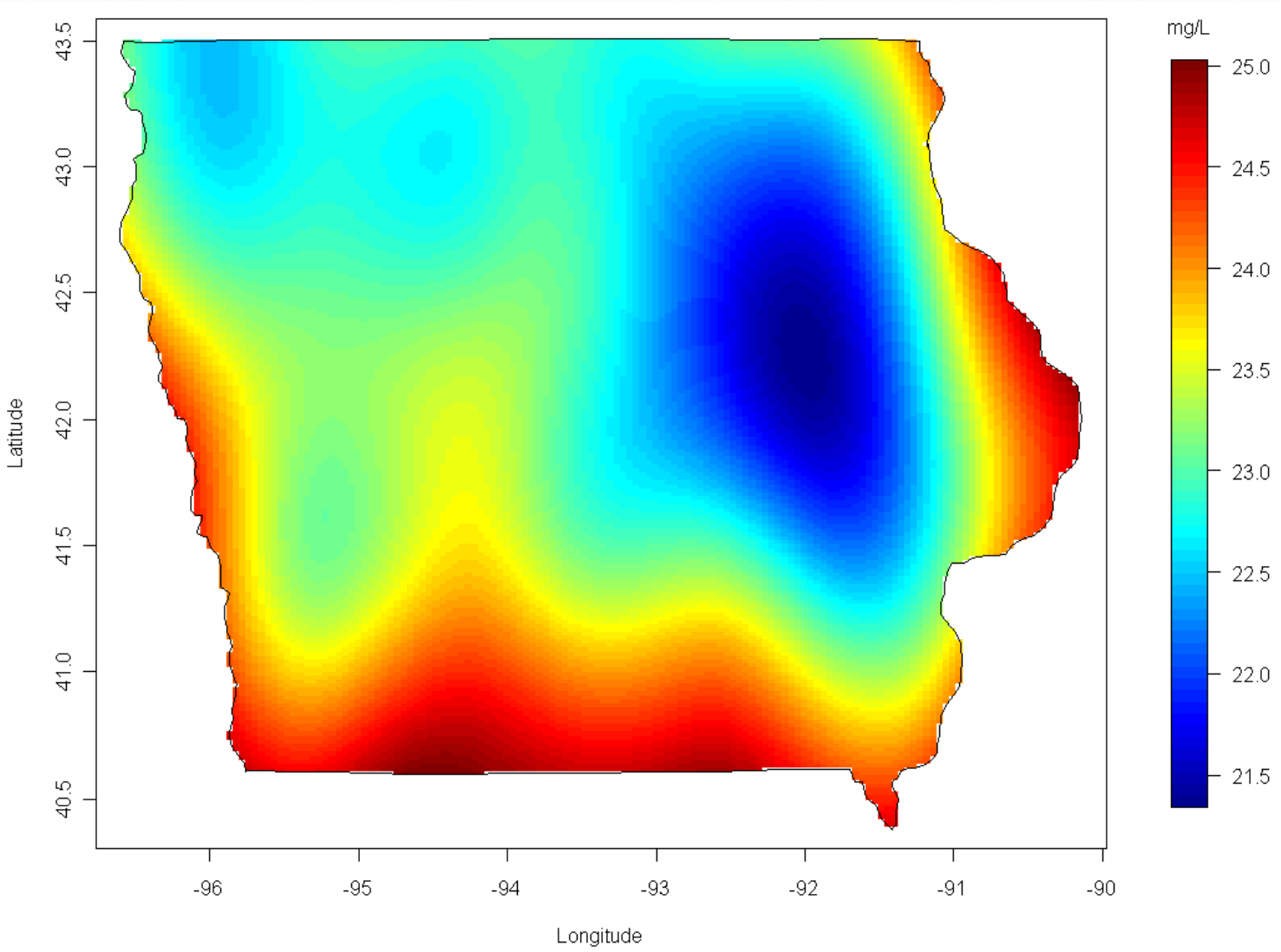
Posterior Parameters' Distributions

	ϕ	4.62 (3.33, 6.00)
	σ_ε^2	5.97 (5.87, 6.08)
	σ_Z^2	4.75 (4.10, 5.50)
Decade	$\beta_{<1960}$	0.78 (0.68, 0.91)
	β_{1960}	1.03 (0.90, 1.19)
	β_{1970}	1.48 (1.32, 1.66)
	β_{1980}	1
	β_{1990}	0.75 (0.69, 0.82)
	β_{2000}	0.82 (0.71, 0.94)
Depth	β_1	1 (0-50 feet)
	β_2	1.02 (0.74, 1.41) (50-150 feet)
	β_3	0.23 (0.16, 0.33) (150-300 feet)
	β_4	0.07 (0.05, 0.58) (300-500 feet)
	β_5	0.05 (0.04, 0.07) (500 + feet)
Treatment	β_{Raw}	0.54 (0.49, 0.58)
	$\beta_{Finished}$	1
Season	β_{Dry}	1
	β_{Wet}	1.11 (1.05, 1.19)

Posterior Geometric Mean



Posterior GSD



Discussion

- The model accounts for important sources of error.
- Although the model uses Bayesian statistics, the flat priors allows the data to shape the parameter summaries. This reflects our lack of prior information.
- It would improve accuracy of the model if there was more data included about the southern counties.
- In our model, finished water seems to have about twice the nitrate in raw water. Perhaps in diluting to lower the concentration of other pollutants, they increase the NO_3 levels.

Bibliography

- Division of Public Health. 2007. *Frequently Asked Questions: N-nitroso Compounds.* Dover, DE.: Delaware Health and Social Services.
- Smith, B. J., Yan, Y., Cowles, M.K.: 2008, 'Unified Geostatistical Modeling for Data Fusion and Spatial Heteroskedasticity with R Package ramps', *J. Stat. Software* **25**(10), 1-21.
- Weyer, P.J., Smith, B. J., Feng, Z.F., Kantamneni, J.R., Riley, D.G.: 2006, 'Comparison of Nitrate Levels in Raw Water and Finished Water from Historical Monitoring Data on Iowa Municipal Drinking Water Supplies', *Environmental Monitoring and Assessment,* **116:** 81-90.