Modeling the Distribution of Environmental Radon Levels in Iowa: Combining Multiple Sources of Spatially Misaligned Data

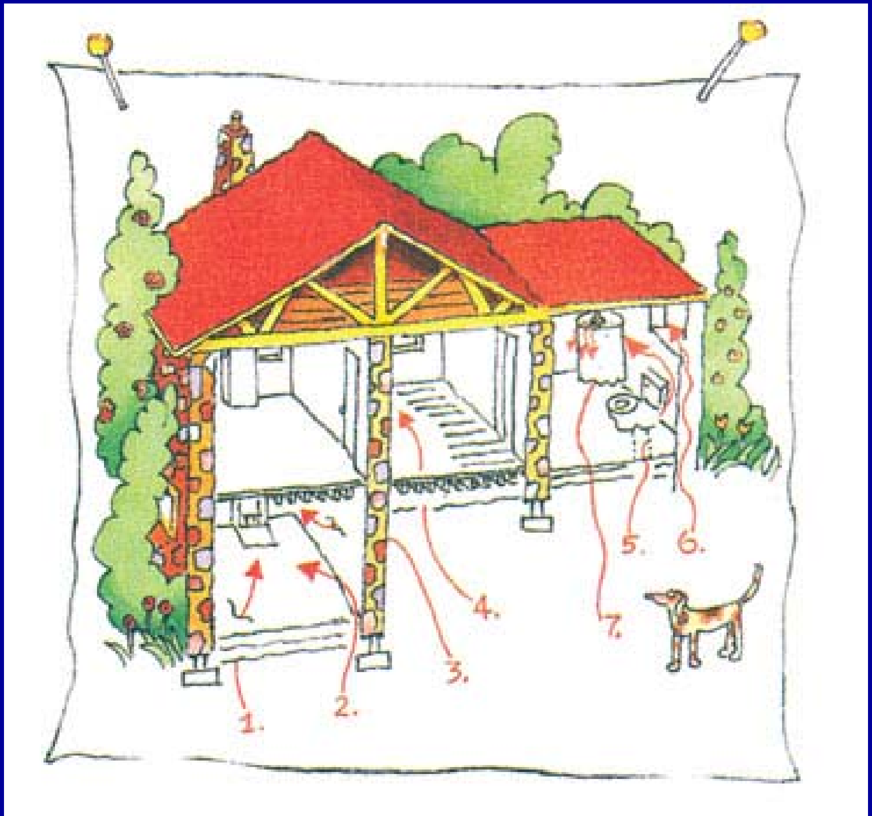
Brian J. Smith, Ph.D.
The University of Iowa
Joint Statistical Meetings
August 10, 2005

Environmental Radon

- Radon is a radioactive gas that originates from uranium in rocks and soil.
- Present to some extent in all dry-land surface air.
- Decay products of radon emit alpha particles which are potentially harmful to lung tissue.
- 18,600 annual lung cancer deaths attributed to radon.
- EPA recommends remediation if home levels exceed 4 pCi/L.

Radon in the Home

1. Cracks in solid floors
2. Construction joints
3. Cracks in walls
4. Gaps in suspended floors
5. Gaps around service pipes
6. Cavities inside walls
7. Water supply



Outline

- Epidemiologic study of residential radon and lung cancer.
- Radon and uranium measurements in Iowa.
- Geostatistical model specification.
- Posterior estimates from Bayesian analysis.

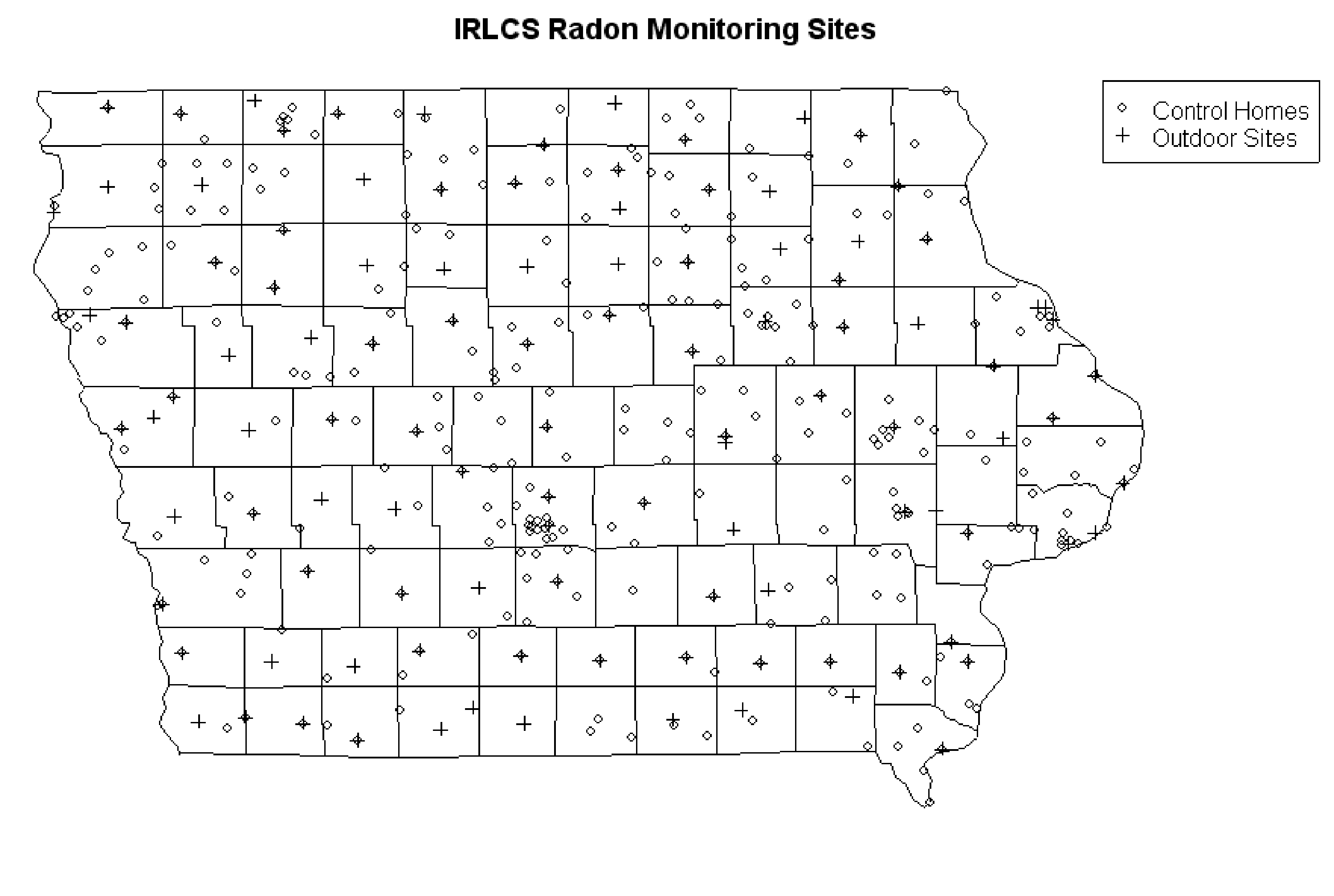
Iowa Radon Lung Cancer Study

- Retrospective study to estimate the lung cancer risk associated with residential radon exposure.
- 413 lung cancer cases; 614 population-based controls.
- Multivariate regression used to estimate the effect of radon exposure on the odds of lung cancer.
- Excess odds estimates for 15-year exposure to 4 pCi/L
  - All subjects: 0.24 (95% CI -0.05 - 0.92)
  - Live subjects: 0.49 (95% CI 0.03 - 1.84)

Exposure Assessment

- Ambient Radon Measurements: Home, outdoors, other buildings
- Subject Mobility: Time spent in different locations within and outside the home
- Cumulative Exposure = Sum of ambient concentrations times subject mobility over the past 20 years.
- Risk estimates attenuated due to random error in the exposure assessment.

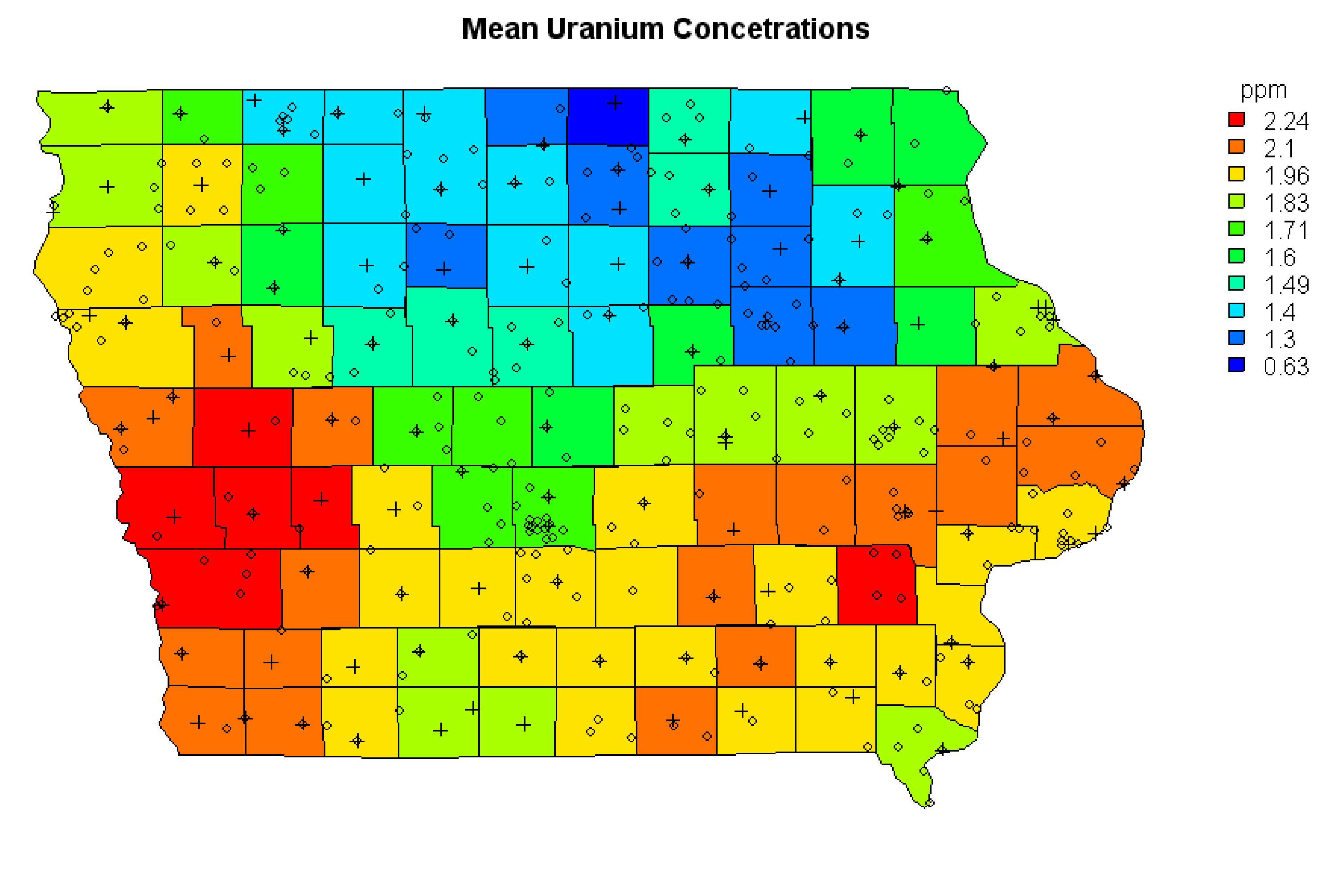
IRLCS Radon Monitoring Sites



National Uranium Resource Evaluation

- Database of surficial uranium concentrations for the contiguous U.S.
- Surficial uranium concentrations serve as an indicator of soil radium concentrations, and hence of ambient radon.
- Surficial radium concentrations have been shown to be useful in explaining variation in indoor radon concentrations.
- Duval, J.S., Jones, W.J., Riggle, F.R., and Pitkin, J.A., 1989, Equivalent uranium map of the conterminous United States: U.S. Geological Survey Open-File Report 89-478, 1989.

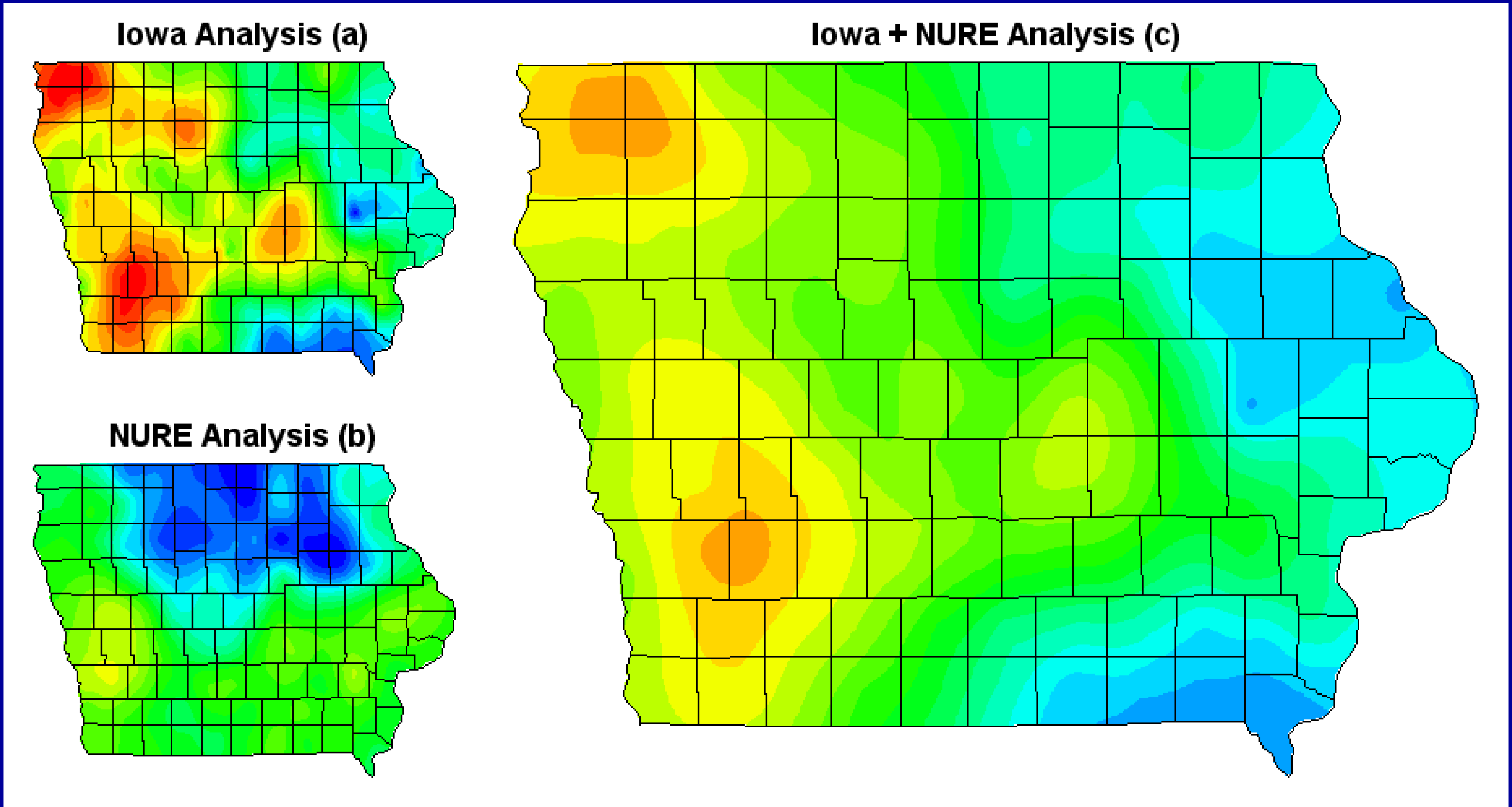
NURE Mean Concentrations



Spatial Modeling

- Assume (point-source) radon measurements and (county average) uranium data arises from a common spatial process.
- Prediction at any point in the geographic region.
- Estimate the mean outdoor radon levels and characterize the distribution of indoor radon levels.
- Develop a unified statistical approach that accounts for the sources of variability.
  - Detector measurement error, housing effect, choice of spatial correlation structure.

Underlying Spatial Process



Latent Spatial Process

- Assume an underlying Gaussian random process that accounts for the spatial correlation in the data.
- Let z_s denote a random draw at site s from a Gaussian process with the following properties:

$$E[z_s] = 0$$

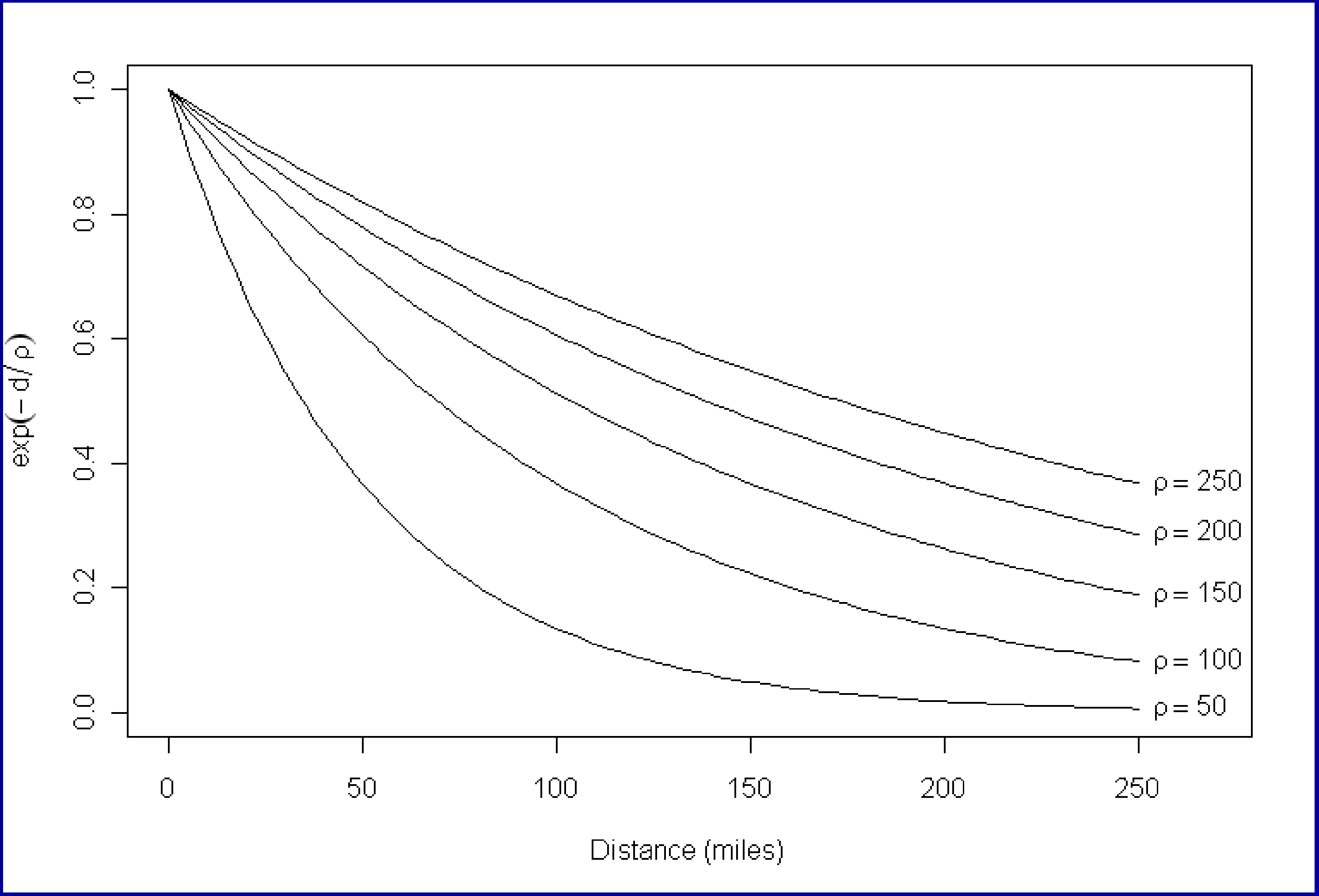
$$\text{cov}[z_s, z_{s'}] = \sigma_S^2 c(s - s'; \theta)$$

$$c(s - s'; \theta) = \exp\{-\|s - s'\| / \rho_S\}$$

where

- $c(s - s'; \theta)$ is an exponential correlation function
- $\|s - s'\|$ is the arc distance between geographic sites
- ρ_s controls the range of decay

Exponential Correlation Function



IRLCS Indoor Radon Measurements

- 2,590 radon measurements were taken at 614 control subject homes (at least one measurement per floor).
- Mean levels differ between floors and potentially as a function of housing characteristics.
- Measurements are subject to independent detector measurement error.

$$\ln y_{H,ij} = \underset{\sim}{\beta}_H^T \underset{\sim}{x}_{H,ij} + \gamma_i + z_{s,i} + \varepsilon_{H,ij}$$

$$\gamma_i \sim N\left(0, \sigma_{BH}^2\right)$$

$$\varepsilon_{H,ij} \sim N\left(0, \sigma_{WH}^2\right)$$

where

$y_{H,ij}$	- j^{th} measurement at the i^{th} home
$\underset{\sim}{x}_{H,ij}$	- housing covariates
$\underset{\sim}{\beta}_H$	- mean covariate parameters
γ_i	- random effect for home
$\varepsilon_{H,ij}$	- independent measurement error

IRLCS Outdoor Radon Measurements

- 136 Outdoor measurements were taken across Iowa (approximately one per county).
- Outdoor levels tend to be lower than indoor levels.
- Measurements are subject to independent detector measurement error.

$$\ln y_{OS,i} = \beta_{OS} + z_{s,i} + \varepsilon_{OS,i}$$

$$\varepsilon_{OS,i} \sim N\left(0, \sigma_{OS}^2\right)$$

where

$y_{OS,i}$ - i^{th} outdoor measurement

β_{OS} - mean concentration

$\varepsilon_{OS,i}$ - independent measurement error

NURE Uranium Measurements

- Assume a model for a potential uranium measurement at site *s*.
- Allow the spatial variance parameter to differ between the uranium and radon measurements.

$$\ln y_U = \beta_U + z_s + \varepsilon_U$$
$$\varepsilon_U \sim N\left(0, \sigma_U^2\right)$$

where

y_U - uranium measurement for a site
β_U - mean concentration
ε_U - independent measurement error

- Measures of the mean surficial uranium concentrations are available for each of the 99 Iowa counties.

$$\ln y_{U,k} = \frac{1}{|B_k|}\int_{B_k}\left(\beta_U + z_s + \varepsilon_U\right)ds = \beta_U + z_{B,k} + \varepsilon_{U,k}$$
$$\varepsilon_{U,k} \sim N\left(0, \sigma_U^2 / |B_k|\right)$$

Spatial Distribution

- The latent spatial variables are distributed as

$$\begin{pmatrix} z_s \\ z_B \end{pmatrix} \Bigg| \sigma_S^2, \phi_S, \theta \sim N\left(0, \sigma_S^2 \begin{pmatrix} H_s(\theta) & \phi_S H_{s,B}(\theta) \\ \phi_S H_{s,B}^T(\theta) & H_B(\theta) \end{pmatrix}\right)$$

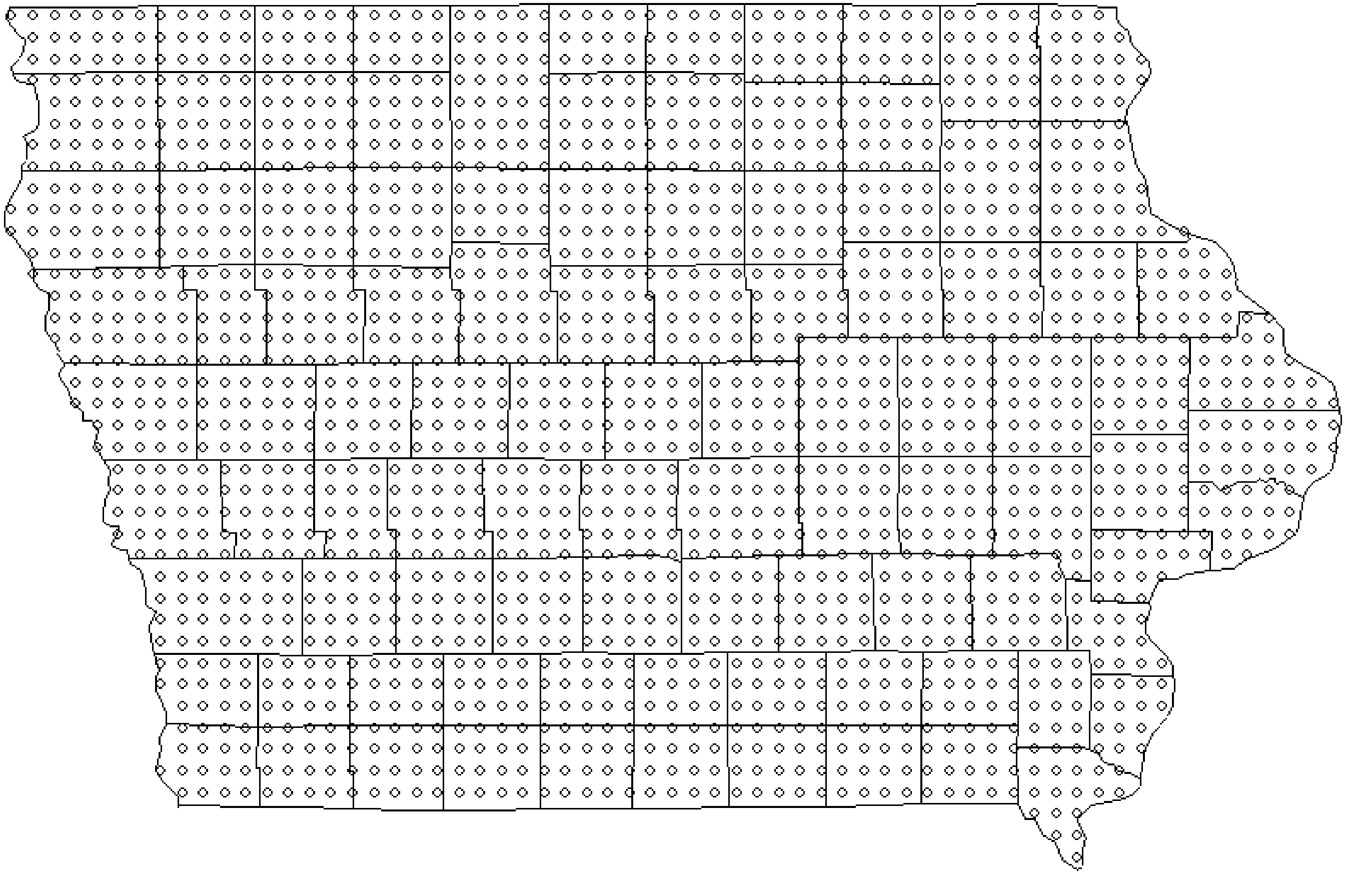
where

$$(H_s(\theta))_{ii'} = c(s_i - s_{i'};\theta)$$

$$(H_{s,B}(\theta))_{ik} = \frac{1}{|B_k|}\int_{B_k} c(s_i - s;\theta)\,ds$$

$$(H_B(\theta))_{kk'} = \frac{1}{|B_k||B_{k'}|}\int_{B_k}\int_{B_{k'}} c(s - s';\theta)\,ds'ds$$

Grid for Monte Carlo Integration



Monte Carlo Integration

- The spatial covariance for the county uranium data can be approximated via Monte Carlo integration.
- Integration over county *k* is replaced with summation over a uniform grid of L_k discrete geographic points.

$$\left(H_s(\theta)\right)_{ii'} = c\left(s_i - s_{i'};\theta\right)$$

$$\left(\hat{H}_{s,B}(\theta)\right)_{ik} = \frac{1}{L_k}\sum_{j=1}^{L_k} c\left(s_i - s_j;\theta\right)$$

$$\left(\hat{H}_B(\theta)\right)_{kk'} = \frac{1}{L_k L_{k'}}\sum_{j=1}^{L_k}\sum_{j'=1}^{L_{k'}} c\left(s_j - s_{j'};\theta\right)$$

Bayesian Modeling Approach

- Conceptually straight-forward method for fitting the proposed hierarchical spatial model.
- Provides an estimate of the posterior distribution for all model parameters.
- Markov chain Monte Carlo methods (Gibbs and Slice Sampling) used to draw from the posterior.
- Uncertainty regarding all components of the model is accounted for and results in more realistic prediction errors.
- Allows for the specification of prior information about the model parameters.

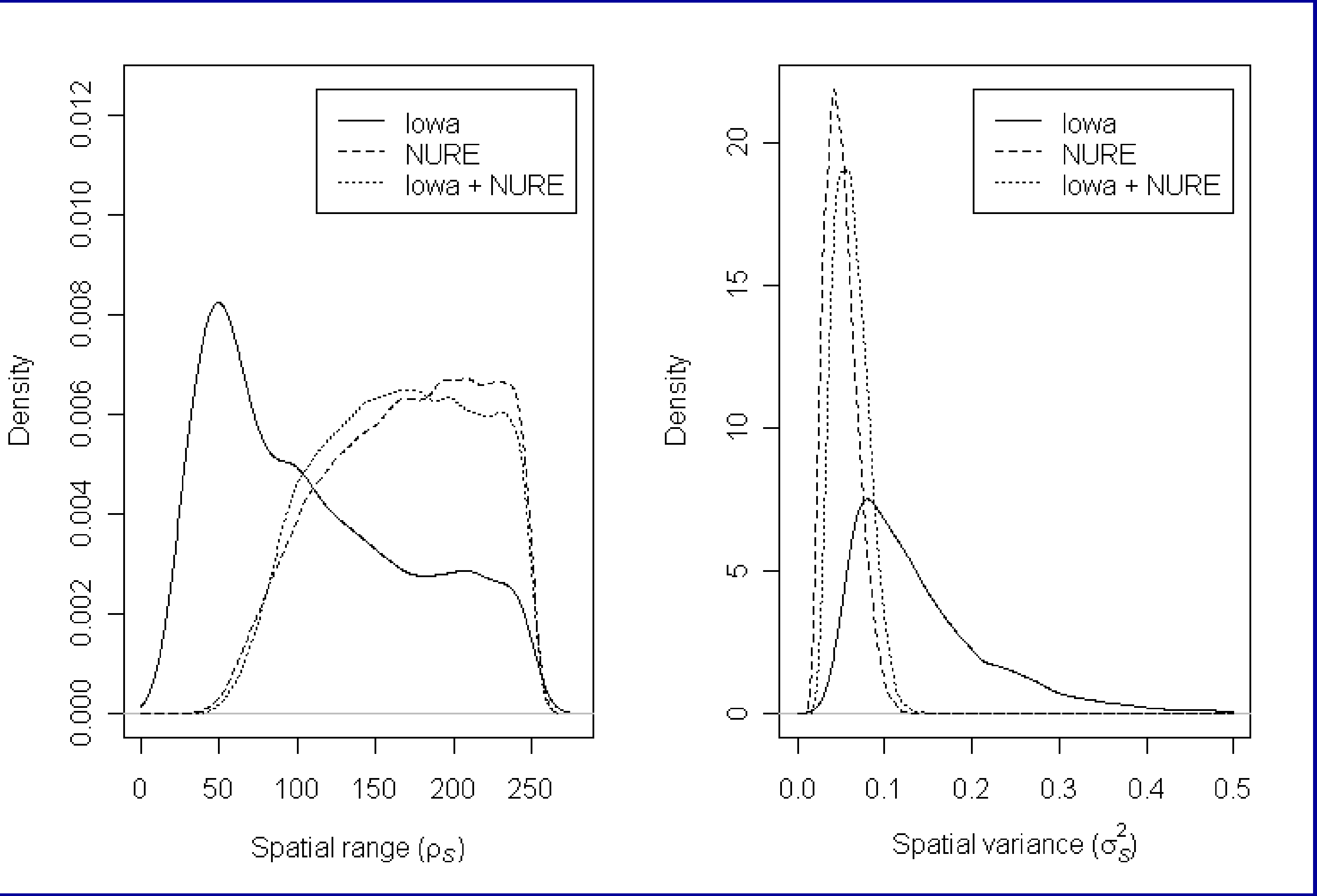
Posterior Estimates

Parameter	Mean	SD	95% HPD
β_{OS}	-0.331	0.137	(-0.582, -0.036)
σ^2_{OS}	0.081	0.012	(0.059, 0.103)
$\beta_{H,0}$	1.472	0.138	(1.214, 1.759)
$\beta_{H,1}$	0.836	0.138	(0.573, 1.117)
$\beta_{H,2}$	0.752	0.138	(0.495, 1.041)
σ^2_{BH}	0.52	0.032	(0.458, 0.582)
σ^2_{WH}	0.064	0.002	(0.060, 0.068)
β_U	0.616	0.119	(0.405, 0.878)
σ^2_U	0.099	0.137	(0.000, 0.388)
ϕ_S	0.336	0.147	(0.036, 0.596)
ρ_S	166.9	48.6	(86.6, 249.8)
σ^2_S	0.061	0.02	(0.028, 0.099)

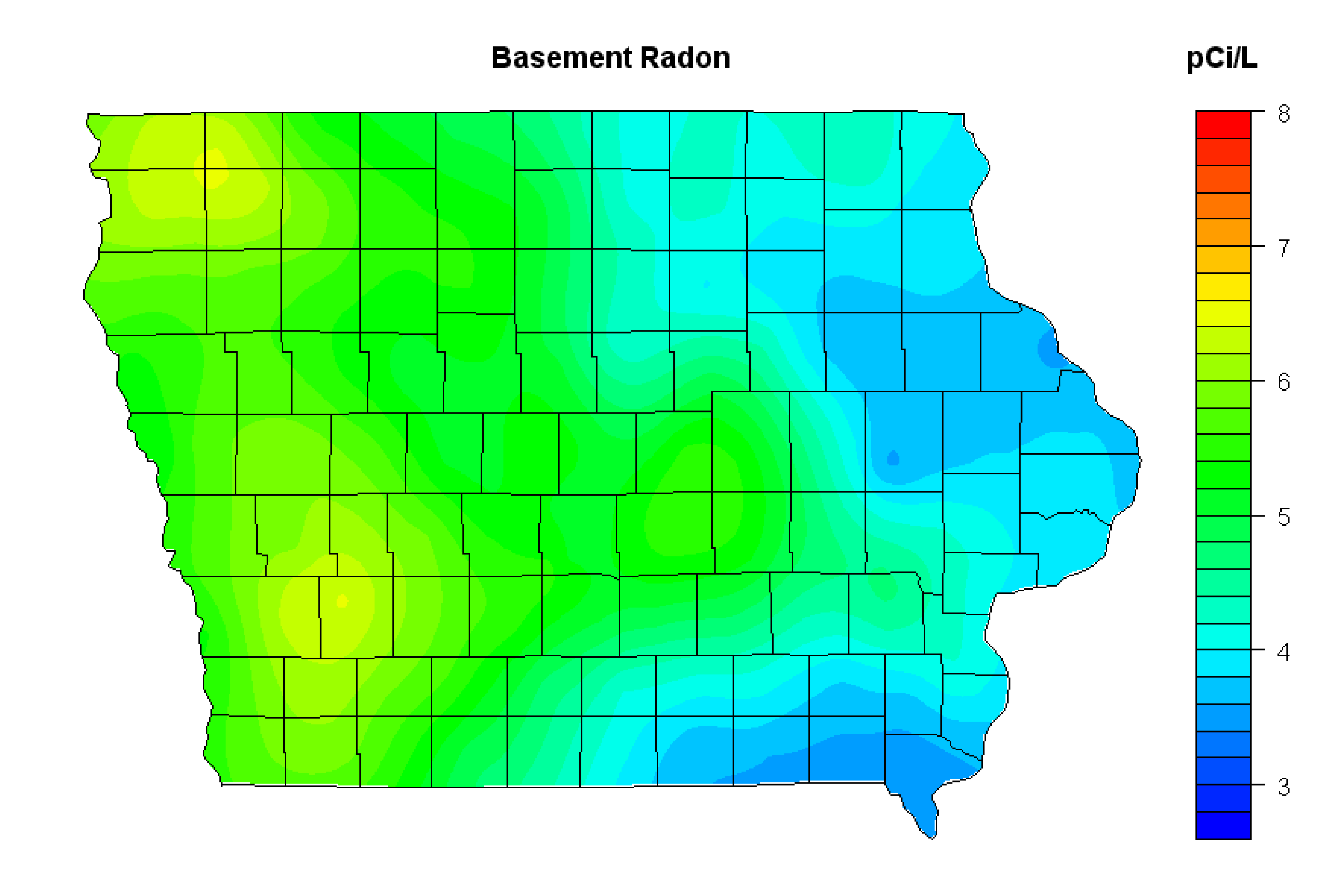
Posterior Estimates

- 80.6% (95% HPD 75.3 – 85.3) of the variability in the indoor radon measurements is due to the random variation between homes.
- Mean outdoor and basement radon levels are 0.73 pCi/L (0.54 - 0.94) and 4.40 pCi/L (3.32 – 5.74), respectively
- First floor radon levels are 0.53 (0.52-0.54) times that of basement levels.
- Levels on the second floor and above are 0.49 (0.47 - 0.50) times those in the basement.
- Mean uranium concentrations are 1.87 ppm (95% HPD 1.50 – 2.40ppm).

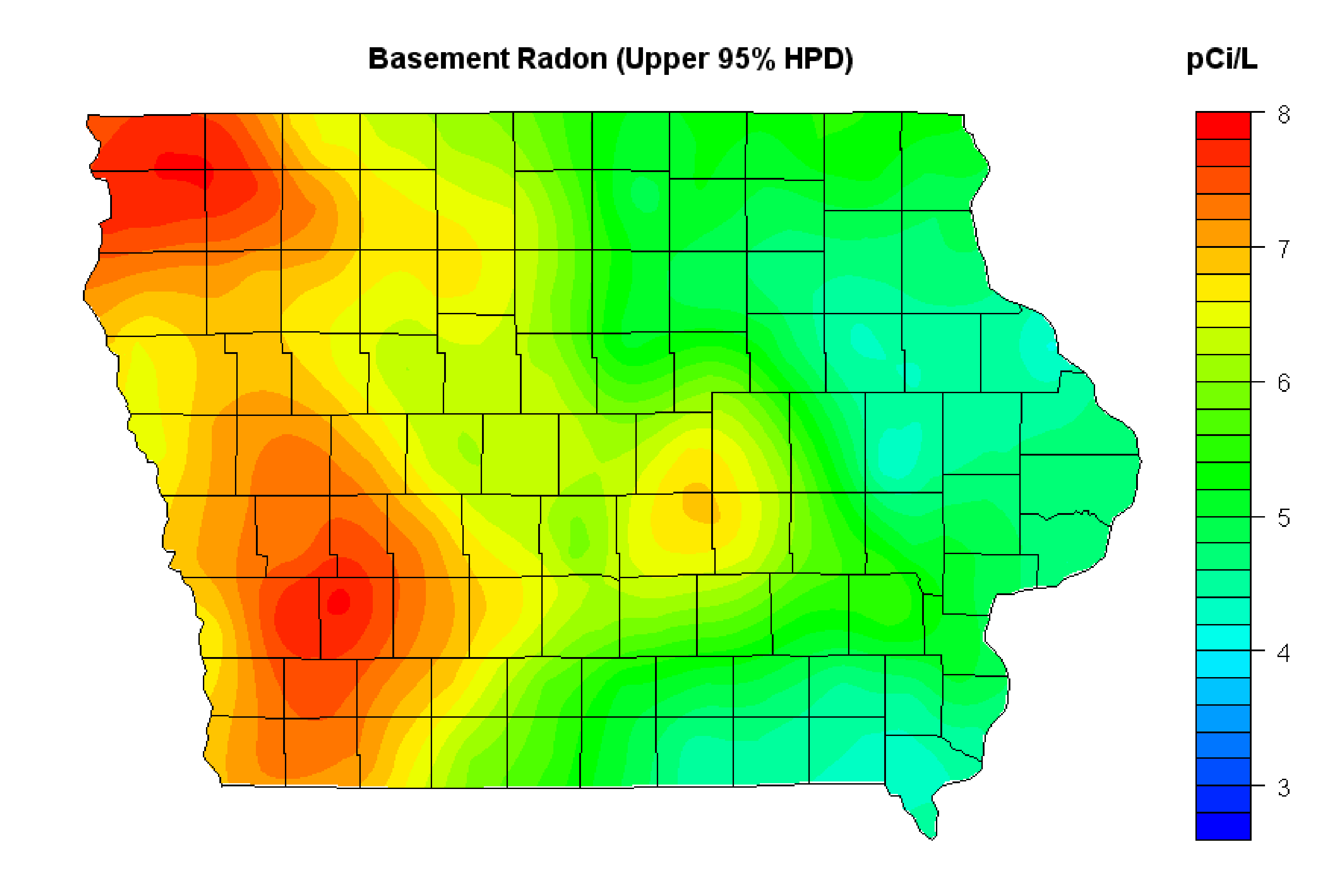
Posteriors Spatial Parameters



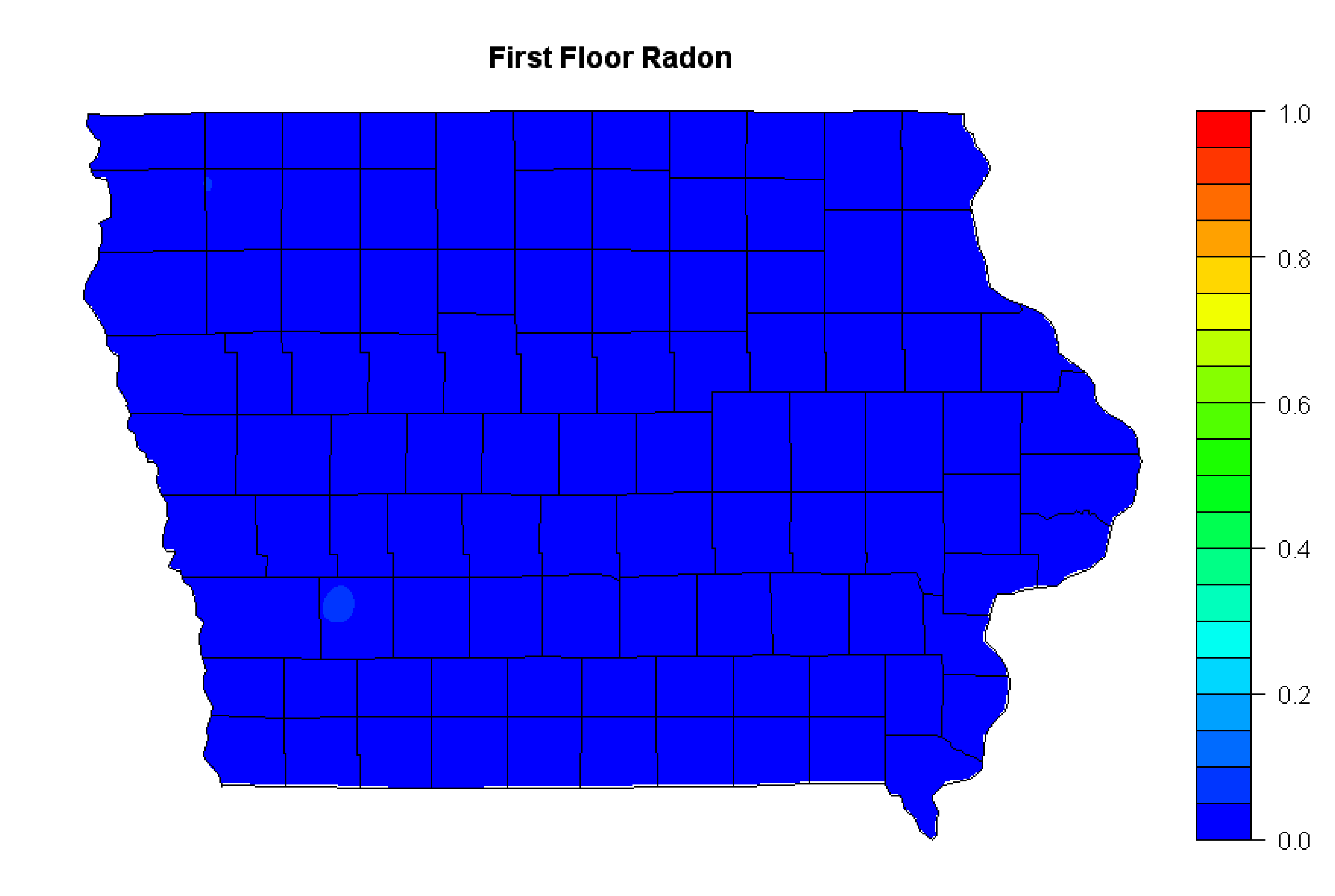
Posterior Mean Predicted Radon



Posterior HPD Predicted Radon



Probability Radon Levels > 4 pCi/L



Comments

- Combination of point-source and aggregate data based on
  - Gelfand AE, Zhu L, and Carlin BP. On the change of support problem for spatio-temporal data. *Biostatistics* 2(1):31-45, 2001.
- MCMC algorithms implemented in R; computationally intensive
- Extension to include temporal variability:
  - Correlation = $c_S(s - s') \times c_T(t - t')$