Computing for Bayesian Spatial Estimation and Prediction with Application to Residential Radon

Kate Cowles[1,2] Brian Smith[2] Jun Yan[1]

[1]Department of Statistics and Actuarial Science
The University of Iowa

[2]Department of Biostatistics
The University of Iowa

Statistics Spring Research Conference
May 22, 2007

Outline

The University of Iowa

Outline

The University of Iowa

Outline

The University of Iowa

Outline

The University of Iowa

Outline

Outline

Radon

- exposure to radon gas in buildings is a likely risk factor for lung cancer
- many public health studies have attempted to estimate residential radon levels
- naturally-occurring radioactive gas (^{222}Rn)
- found to some extent in all dry-land surface air
- a decay product of the uranium in rocks and soil

Our goals

- scientific goal: to improve understanding of the geographic distribution of residential radon levels in Iowa by combining two available datasets
  - *estimation* of parameters describing the spatial correlation and average levels
  - *prediction* of radon levels at unmeasured locations
  - *mapping* of surface of average residential radon over Iowa
- computing goal: to develop computing algorithms and strategies that would enable use of much larger datasets of this kind

Two data sources for this study

- Iowa Radon Lung Cancer Study (IRLCS)
- EPA/State Residential Radon Survey for Iowa (SRRS)

Iowa Radon Lung Cancer Study

- purpose: to estimate the effect of residential radon on lung cancer risk
- began in 1993
- 4626 radon measurements in 1027 homes in Iowa
  - 413 incident lung cancer cases
  - 614 population-based, disease-free controls
  - at least one radon detector placed on each floor of each home
  - locations of homes representative of geographic distribution of Iowa population
- measurements were *annual averages* taken using alpha-track detectors
- *point-referenced* data
- units are picocuries per liter (pCi/l)

More on IRLCS

- Iowa – ideal region for study of residential radon exposure
- we used data only from control homes
  - 598 homes
  - 2802 measurements

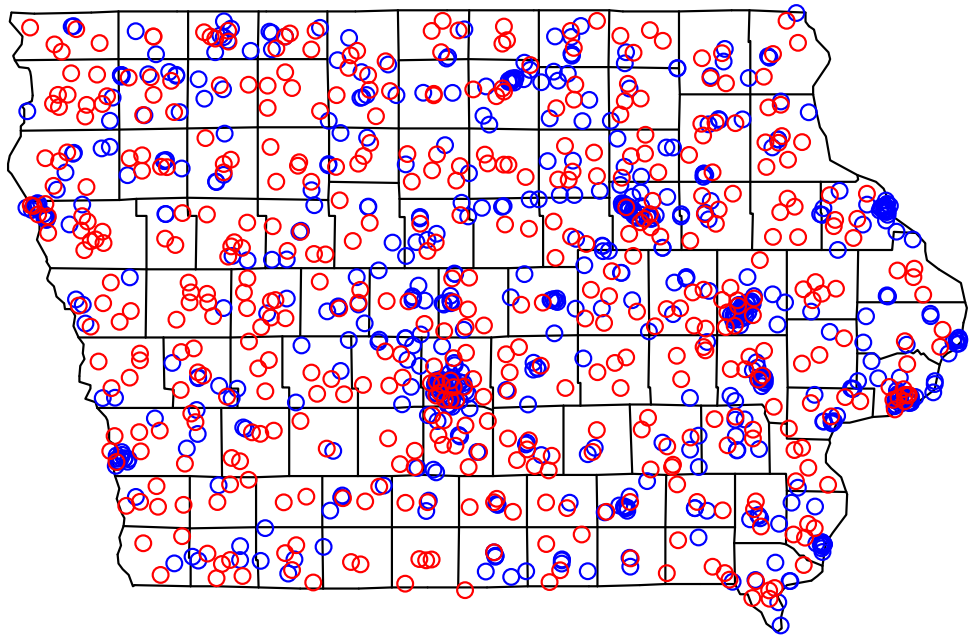
EPA/State Residential Radon Surveys

- EPA began in winter of 1986/87 to assist states in measuring home radon
- two objectives
  - to estimate the *maximum* indoor radon potential
    - short-term measurements were take with charcoal canisters placed in the lowest livable areas of homes for 2–7 winter days
  - to identify geographic regions of elevated indoor radon
    - random-digit dialing to sample geographically dispersed homes

SRRS data for Iowa

- 1381 short-term radon measurements
- addresses not available, so geocoding to latitude/longitude coordinates not possible
- zipcodes and county identifiers provided
  - zipcode sometimes not considered reliable
- 486 distinct zipcodes

IRLCS homes (blue) and SRRS zipcode centroids (red)



Facts about radon in homes

- radon concentrations highest in basements; decreasing gradient as go up through floors
- characteristics of home and residents affect radon levels

Outline

Bayesian geostatistical models

- geostatistical models
  - natural and interpretable way to model spatial correlation for data measured at irregularly-spaced point sites
  - correlation is a function of the distance, and possibly orientation, between sites
- Bayesian approach
  - leads to more accurate quantification of uncertainty in estimation and prediction than frequentist kriging

Parametric correlation functions

Function	$corr(\phi, d)$
Exponential	$exp(-\phi d)$
Gaussian	$exp(-(\phi d)^2)$
Spherical	$\frac{1}{2}(\phi^3 d^3 - 3\phi d + 2)$, $d \leq \frac{1}{\phi}$
	0, $d > \frac{1}{\phi}$

- ϕ is a parameter controlling the rate of decay of correlation with increasing distance
- $corr(\phi, d)$ is the correlation between residuals at two sites separated by distance d.

Isotropic and anisotropic spatial correlation

- **isotropic**: correlation decays with increasing distance at the same rate in all directions
- **geometrically anisotropic**: correlation is a function of *orientation* as well as distance between points
  - requires two additional parameters in correlation function

Simplest geostatistical model for point-source data with spatial correlation and additive measurement error

$$Y = X\beta + Z + \varepsilon, \tag{1}$$

$$Z \sim N\left(0, \sigma_z^2 \Omega(\phi)\right), \quad \varepsilon \sim N(0, \sigma_e^2 I), \tag{2}$$

- **X** is a matrix of covariates
- β is a vector of coefficients to be estimated
- $\Omega(\phi)$ is spatial correlation matrix
  - entries are calculated from correlation function
- σ_z^2 is spatial variance
- σ_e^2 is random variance (measurement error variance)
- I is identity matrix
- Bayesian model completed by specification of prior distributions on ϕ, σ_z^2, σ_e^2, and β

Specifying prior distributions for model parameters

$$
\begin{aligned}
p(\beta) &\propto 1 \\
\sigma_e^2 &\sim IG(a_e, b_e) \\
\sigma_z^2 &\sim IG(a_z, b_z) \\
\phi &\sim U(l_\phi, u_\phi)
\end{aligned}
$$

- improper flat prior on β
- semi-conjugate inverse gamma prior distributions on variances
  - infinite variance; centered at 0.5
- uniform prior on ϕ chosen to include all possible values

Fitting the model using standard MCMC algorithms

- MCMC methods construct a Markov chain whose stationary distribution is the posterior joint distribution of interest:

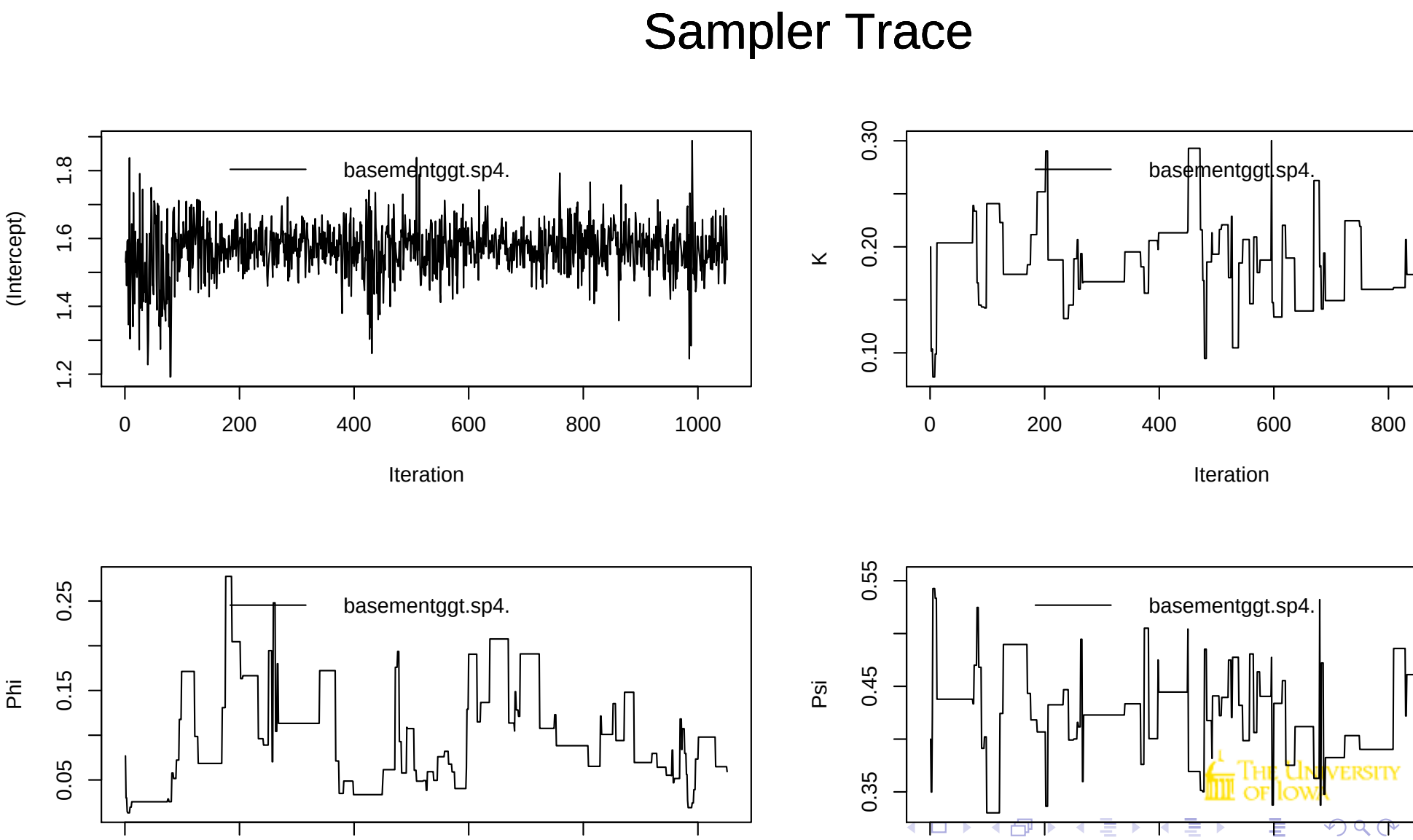
$$p(\beta, \sigma_e^2, \sigma_z^2, \phi|\mathbf{y})$$

- WinBUGS?
- `ggt.sp` function in R package `spBayes`
  - developed by Andrew Finley, Brad Carlin, and Sudipto Banerjee at U of Minn Biostats
  - factor posterior distribution into

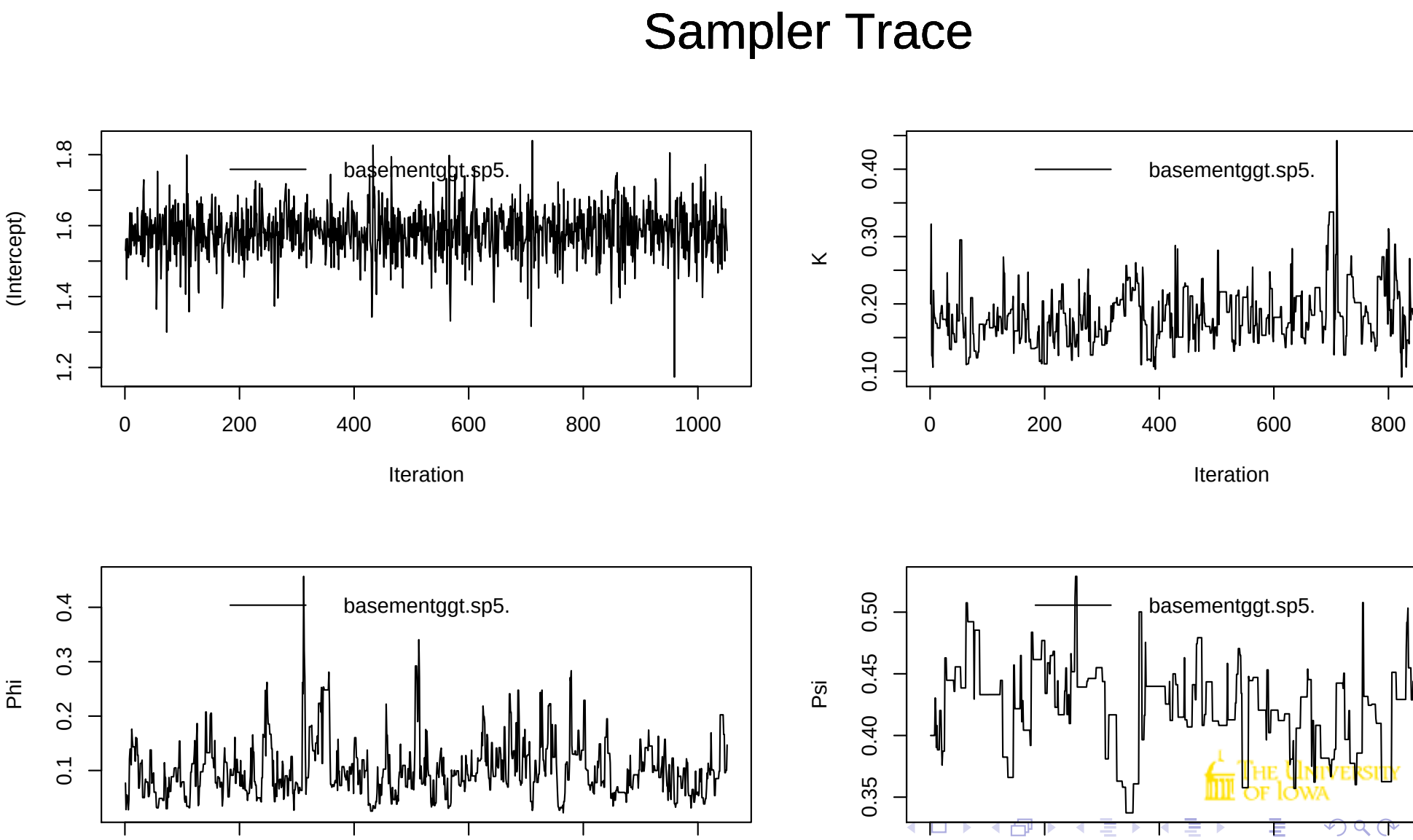
$$p(\sigma_e^2, \sigma_z^2, \phi|\mathbf{y})p(\beta|\sigma_e^2, \sigma_z^2, \phi, \mathbf{y})$$

  - use Metropolis-Hastings algorithm to draw from $p(\sigma_e^2, \sigma_z^2, \phi|\mathbf{y})$

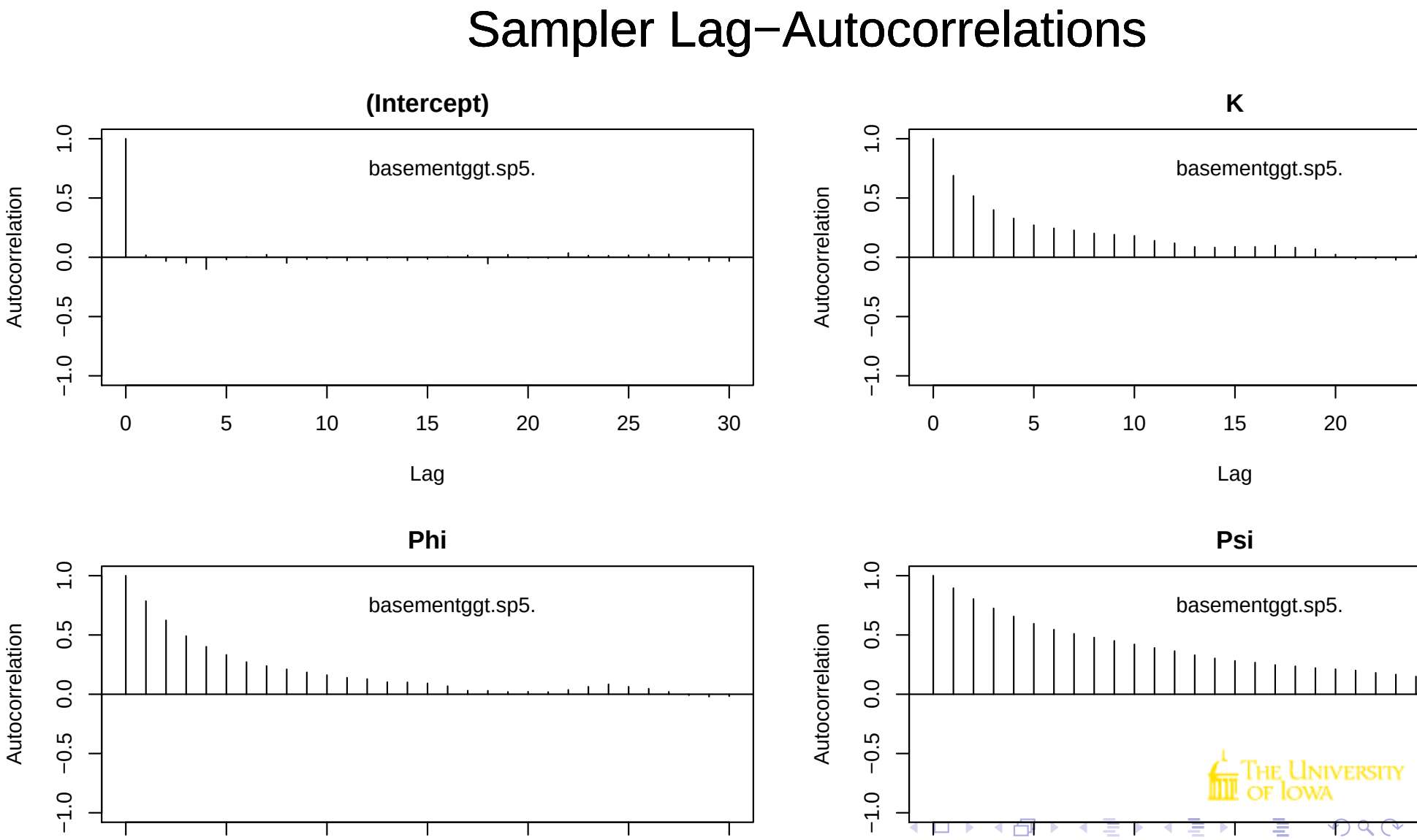
Trace plots: ggt.sp algorithm 1



Trace plots: ggt.sp algorithm 2



Autocorrelation plots: ggt.sp algorithm 2



Our alternative reparameterization

- facilitates prior specification and computing algorithm
- reparameterized covariance matrix

$$Y \sim N\left(X\beta, \sigma_z^2\Omega(\phi) + \sigma_e^2 I\right)$$

$$\sigma_z^2\Omega(\phi) + \sigma_e^2 I \;=\; \sigma_{tot}^2\left[(1-\kappa)\,\Omega(\phi) \;+\kappa\, I\right]$$

where

$$\begin{aligned}\sigma_{tot}^2 &= \sigma_z^2 + \sigma_e^2\\ \kappa &= \frac{\sigma_e^2}{\sigma_z^2 + \sigma_e^2}\end{aligned}$$

- κ is “shrinkage factor" used in uniform shrinkage prior on variances in hierarchical models

Prior distributions for σ^2_{tot} and κ

If

$$\begin{aligned} \sigma^2_z &\sim IG(a_z,\, b_z) \\ \sigma^2_e &\sim IG(a_e,\, b_e) \end{aligned}$$

then standard multivariate change-of-variable methods can be used to show that the joint prior distribution induced on (σ^2_{tot}, κ) is

$$f(\kappa; a_z, b_z, a_e, b_e) = \frac{\Gamma(a_e + a_z) b_e^{a_e} b_z^{a_z}}{\Gamma(a_e)\Gamma(a_z)} \frac{\kappa^{a_z - 1}(1-\kappa)^{a_e - 1}}{[b_z \kappa + b_e(1-\kappa)]^{a_e + a_z}}, \quad \kappa \in (0,1)$$

$$\sigma^2_{tot}|\kappa \sim IG\left(a_e + a_z,\, \frac{b_z}{1-\kappa} + \frac{b_e}{\kappa}\right).$$

When $b_z = b_e$, the marginal density of κ is a Beta density with shape parameters a_z and a_e.

Our MCMC algorithm for estimation in this model

- based on factoring posterior

$$p(\boldsymbol{\beta}, \phi, \sigma^2_{tot}, \kappa|\mathbf{y}) = p(\phi, \kappa|\mathbf{y}) \times p(\sigma^2_{tot}|\phi, \kappa, \mathbf{y}) \times p(\boldsymbol{\beta}|\phi, \sigma^2_{tot}, \kappa, \mathbf{y})$$

- each MCMC iteration m
  - generate $(\phi^m, \kappa^m,)$ from continuous joint marginal $p(\phi, \kappa|\mathbf{y})$ using slice sampling
  - generate $\sigma^{2\,m}_{tot}$ from $p(\sigma^2_{tot}|\phi^m, \kappa^m, \mathbf{y})$ — inverse gamma
  - generate $\boldsymbol{\beta}$ from $p(\boldsymbol{\beta}|\phi^m, \sigma^{2\,m}_{tot}, \kappa^m, \mathbf{y})$ — multivariate normal
- all parameters are blocked – would be i.i.d. sampling if there were a way to obtain independent draws from joint posterior marginal of ϕ and κ

Drawing from joint posterior marginal distribution of ϕ and κ

$$p(\phi,\kappa \mid \mathbf{y}) \propto$$

$$\frac{1}{\left(\frac{b_z}{1-\kappa}+\frac{b_e}{\kappa}+\frac{\left[\mathbf{y}-\mathbf{X}\,\hat{\boldsymbol{\beta}}(\phi,\kappa)\right]^T\Omega(\phi,\kappa)^{-1}\left[\mathbf{y}-\mathbf{X}\,\hat{\boldsymbol{\beta}}(\phi,\kappa)\right]}{2}\right)^{a_e+a_z+\frac{n-p}{2}}}$$

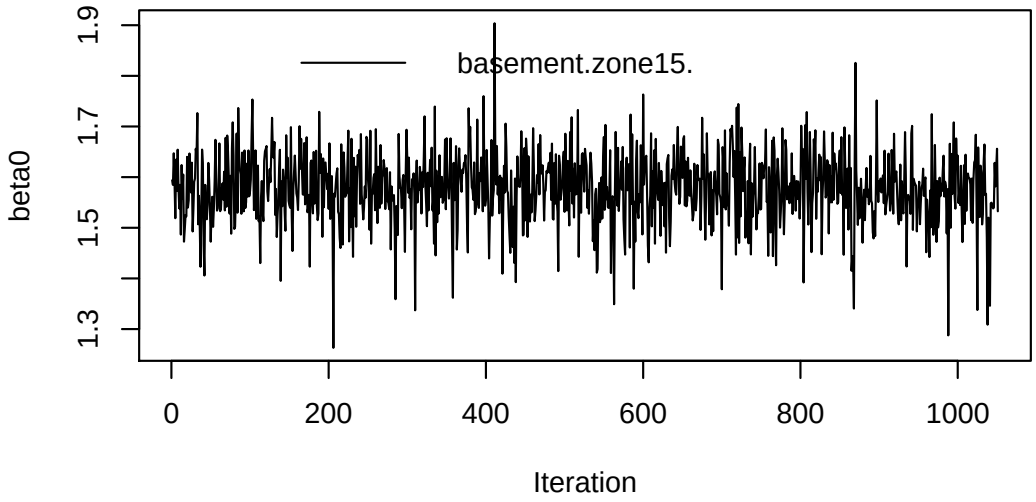
$$\times\frac{1}{|\Omega(\phi,\kappa)|^{1/2}}\,\frac{1}{|\mathbf{X}^T\Omega(\phi,\kappa)^{-1}\mathbf{X}|^{1/2}}$$

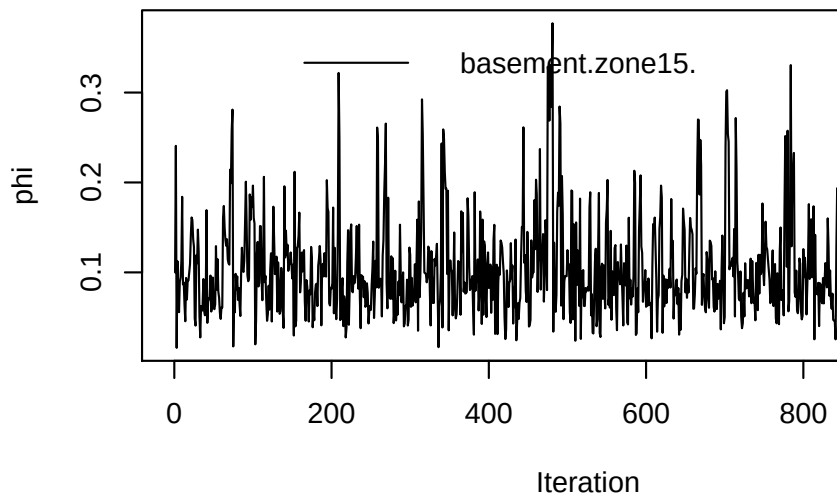
$$\times\kappa^{-a_e-1}(1-\kappa)^{-a_z-1}\times I_{(0,1)}(\kappa)\times I_{l,r}(\phi)$$

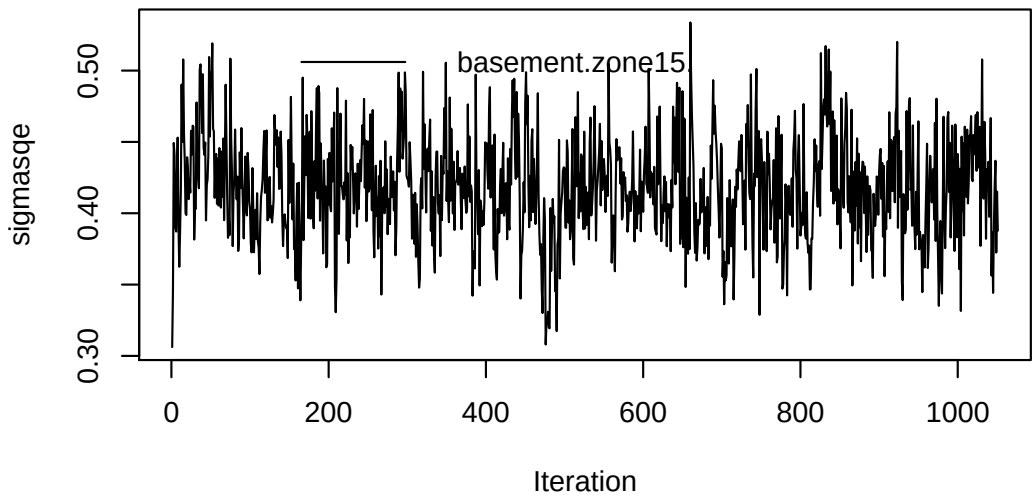
Drawing from joint posterior marginal distribution of ϕ and κ

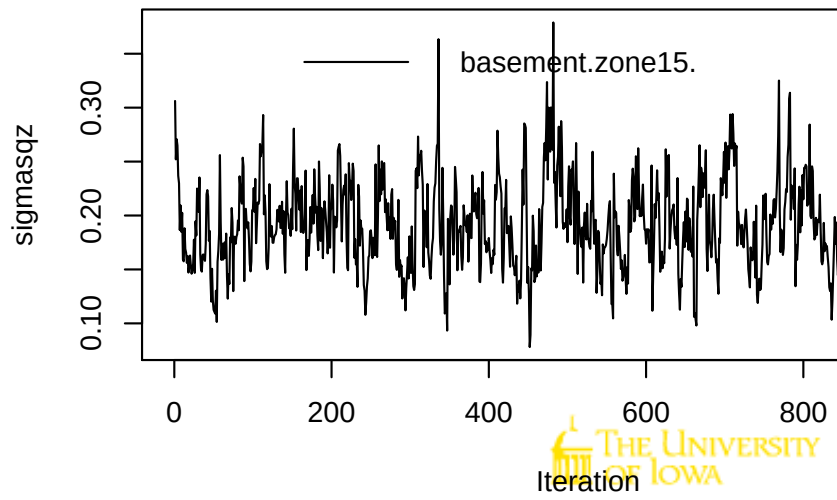
- Cholesky decomposition of $\Omega(\phi, \kappa)$ enormously reduces computation involved for obtaining required determinants and quadratic form
- bivariate slice sampling (Neal, 2003) attractive due to finite support of both parameters
- in our experience, compared to Metropolis updating, slice sampling
  - results in lower autocorrelation in sampler output since new values are drawn at every iteration
  - at cost of requiring more computationally-expensive evaluations within each iteration
  - results in more “effective samples per second"

Sampler Trace

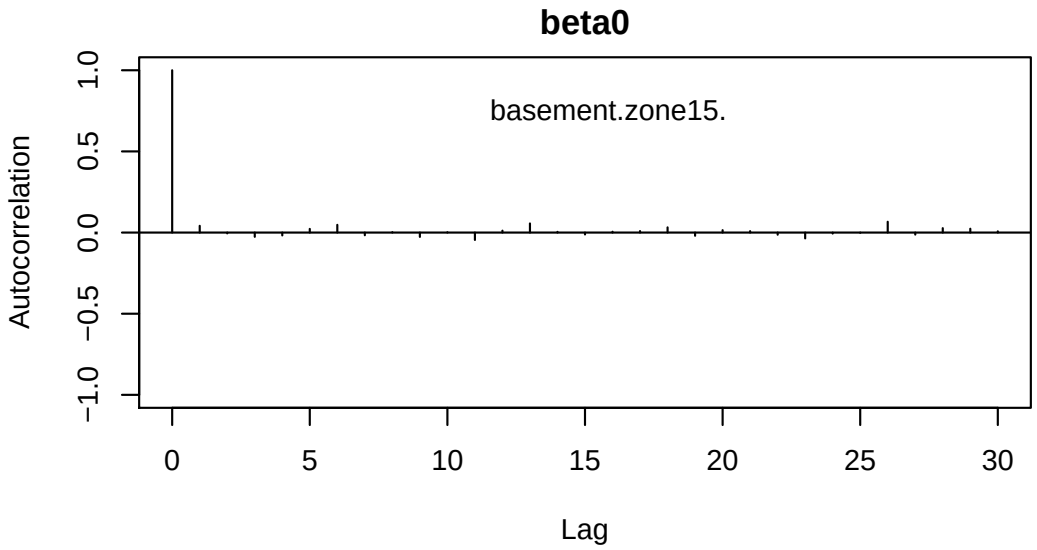


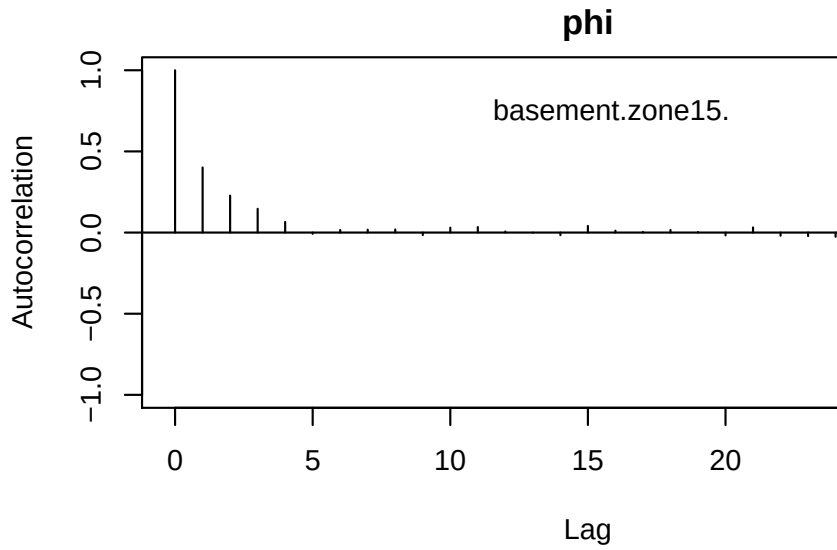


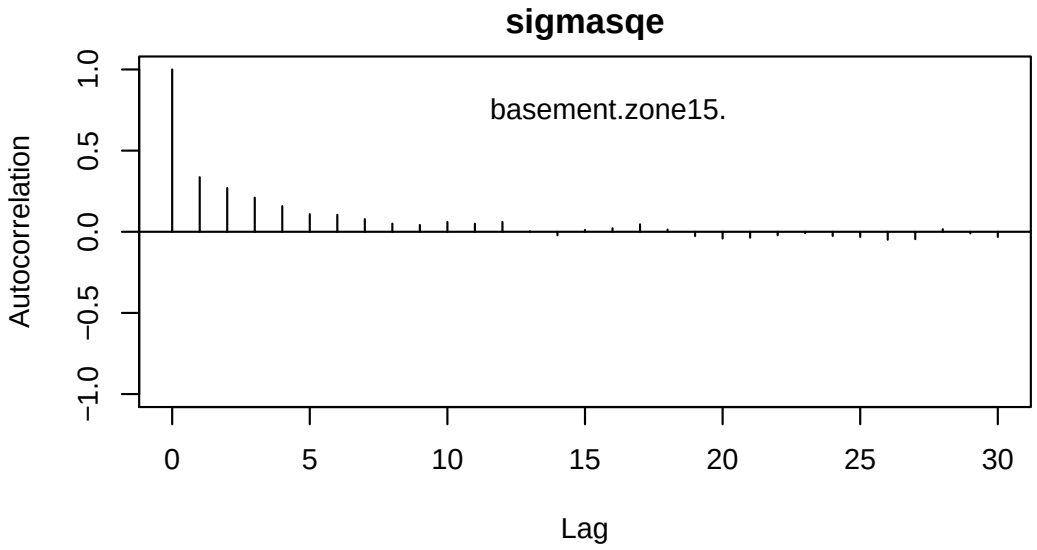


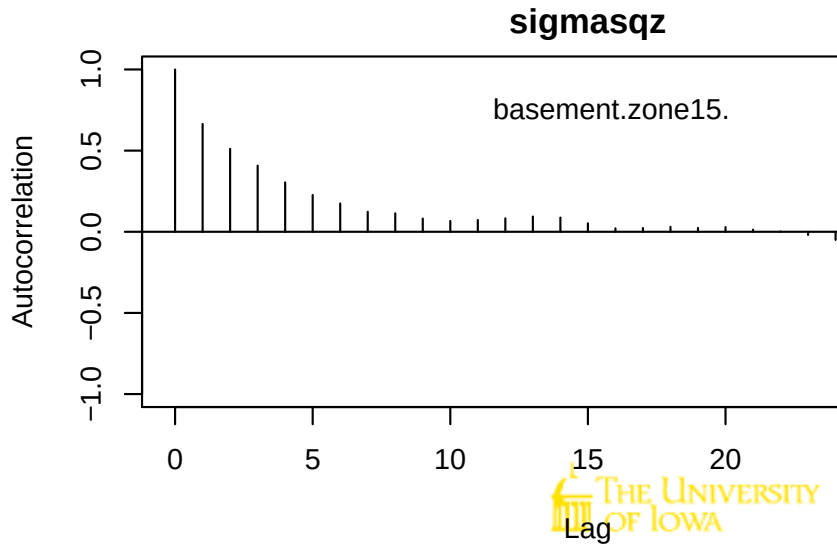


Sampler Lag−Autocorrelations









Comparison of RAMPS and standard MCMC

Table: Efficiency comparison of the RAMPS algorithm and the Gibbs sampler for a sample dataset of 437 spatial observations. ARL is the loss of information due to autocorrelation measured by the expected lag at which autocorrelation drops to zero. ESS is the effective sample size. ESS/s is ESS per second. Both algorithms were run with 10,000 iterations on a 2.40GHz CPU Linux machine with the first 1,000 discarded. The RAMPS algorithm took 9390s and the Gibbs sampler took 6228s.

Parameter	Gibbs Sampler			RAMPS Algorithm			ESS/s Ratio
	ARL	ESS	ESS/s	ARL	ESS	ESS/s	RAMPS/Gibbs
β_0	1.00	9000.0	1.445	1.00	9000.0	0.958	0.663
σ_z^2	37.28	241.4	0.039	1.30	6913.9	0.736	18.872
σ_e^2	23.27	386.8	0.062	1.74	5172.1	0.551	8.887
ϕ	39.74	226.5	0.036	5.22	1725.7	0.184	5.111

Aside: Parallelizing RAMPS algorithm

Yan, J., Cowles, M.K., Wang, S., Armstrong, M. “Parallelizing MCMC for Bayesian Spatiotemporal Geostatistical Models." to appear, *Statistics and Computing*.

Outline

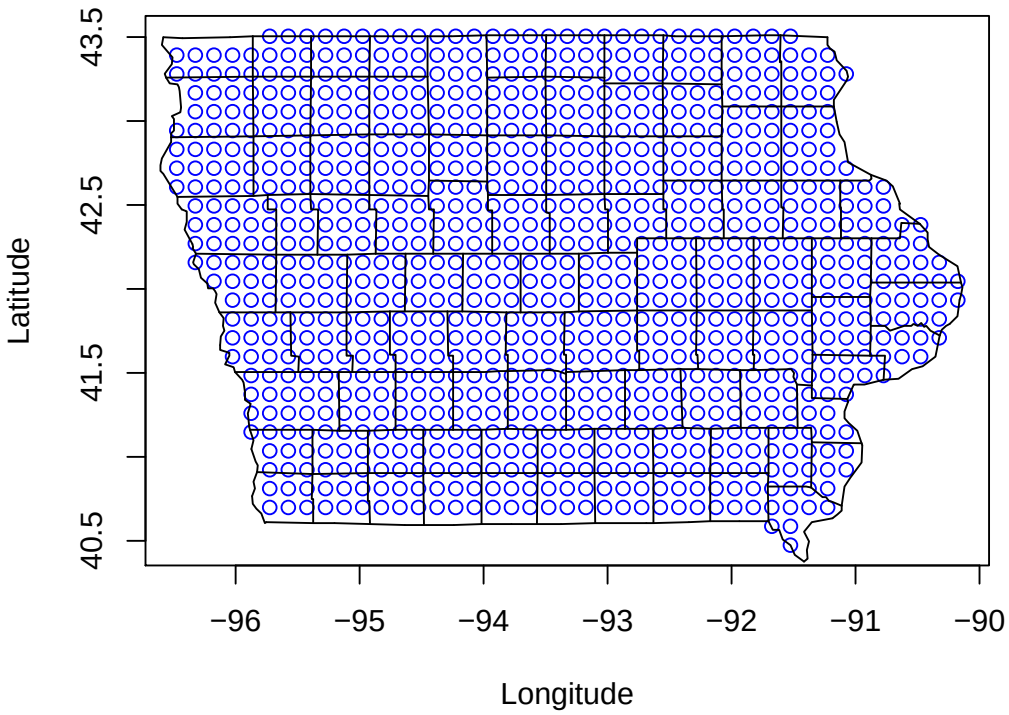
Why preceding model is inadequate

- IRLCS data
  - multiple observations at same home
  - need home-specific random effects to capture effects of characteristics of each home and its residents
- SRRS data
  - don't know exact locations

Possible ways to handle SRRS data

- attribute all measurements in a given zipcode to zipcode centroid
  - Young and Gotway show this can give wrong inference regarding regression coefficients
  - obviously would not allow for correct inference regarding spatial correlation
- work with areal county averages; for each county:
  - average radon measurements
  - average covariates
  - keep track of how many actual observations contributed to each average
- work with areal zipcode averages

Grid of points for areal averages and prediction



Geostatistical model actually used

$$
\begin{aligned}
Y &= X\beta + W\gamma + KZ + \epsilon, \\
\gamma &\sim N(0, \Omega_\gamma), \quad Z \sim N(0, \Omega_Z), \quad \epsilon \sim N(0, \Omega_\epsilon),
\end{aligned}
\tag{3}
$$

where

- Y is concatenation of 2802 log-transformed radon values from IRLCS homes and 99 county averages obtained from SRRS data
- β is a 5×1 vector of fixed-effect coefficients
  - $\beta_{IRLCS,0}$, $\beta_{IRLCS,1}$, $\beta_{IRLCS,2}$, $\beta_{SRRS,0}$, $\beta_{SRRS,1}$
- γ is a 598×1 vector of non-spatial random effects for homes
- Z is an $(598 + 932) \times 1$ vector of spatial random effects
  - one for each IRLCS home location, plus 932 grid points for areal data
- ϵ is an 2901×1 vector of measurement error,
- matrices X, W, and K are design matrices for fixed effects, non-spatial random effects, and spatial random effects.

Matrix K

$$K_{ij} = \begin{cases} 1, & Y_i \text{ is a point source datum measured at site } j, \\ \frac{1}{N_i}, & \text{site } j \text{ is one of } N_i \text{ sites contributing to areal average } Y_i, \\ 0, & \text{otherwise.} \end{cases}$$

If Y_i is a point-source measurement, then $N_i = 1$. For Y_i's that are areal averages, the corresponding N_i's are roughly proportional to the areas of the regions over which the measurements are averaged. The finer the grid of sites used, the closer the proportionality will be.

Variance/covariance matrices

- 4 variances
  - $\sigma^2_{e,IRLCS}$ measurement error variance, long term
  - $\sigma^2_{e,SRRS}$ measurement error variance, short term
  - σ^2_{re} between-home variability
  - σ^2_{z} variance of underlying spatial process
- exploratory analysis of IRLCS data indicated isotropic exponential correlation function had best combination of fit and parsimony

Prior parameters

β	improper, $\propto 1$ on real line
$\sigma^2_{e,IRLCS}$	IG(0.01, 0.01)
$\sigma^2_{e,SRRS}$	IG(0.01, 0.01)
σ^2_{re}	IG(0.01, 0.01)
σ^2_z	IG(0.01, 0.01)
ϕ	U(0.009, 1.5)

Prior distribution on ϕ implies that distance at which spatial correlation drops to 0.05 is somewhere between 2 and 340 miles

Outline

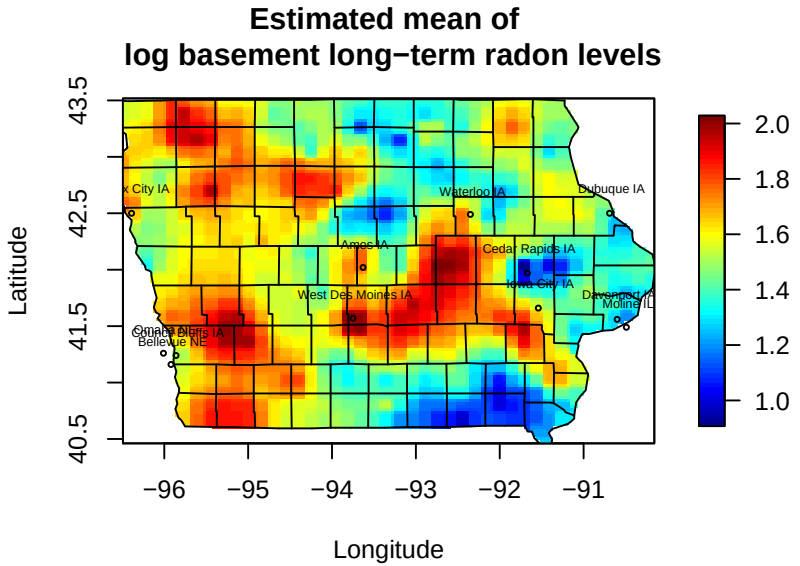
Computing

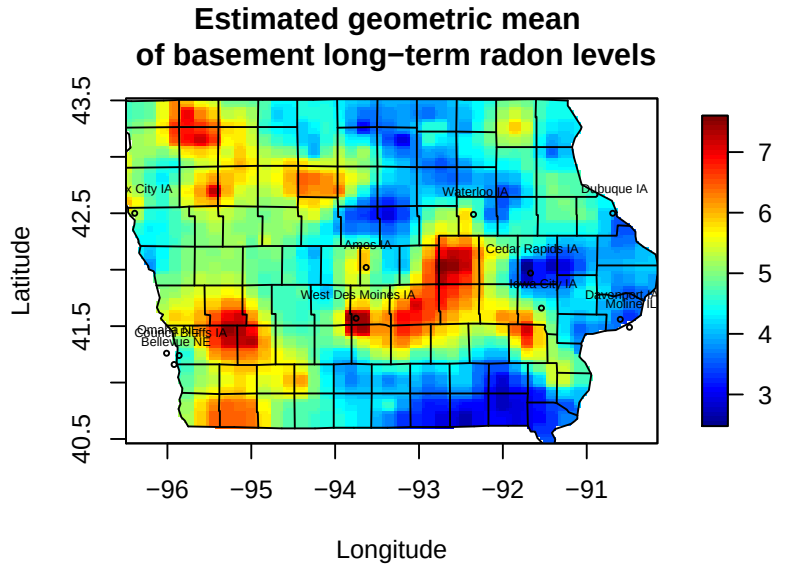
- fitted model to IRLCS data alone and to combined IRLCS and SRRS
- for each analysis, 2 parallel chains run for 1000 iterations
- trace plots and Brooks, Gelman, and Rubin diagnostic in R package boa suggested convergence within 200 iterations
- reported results are based on combining last 500 iterations of both chains for each analysis

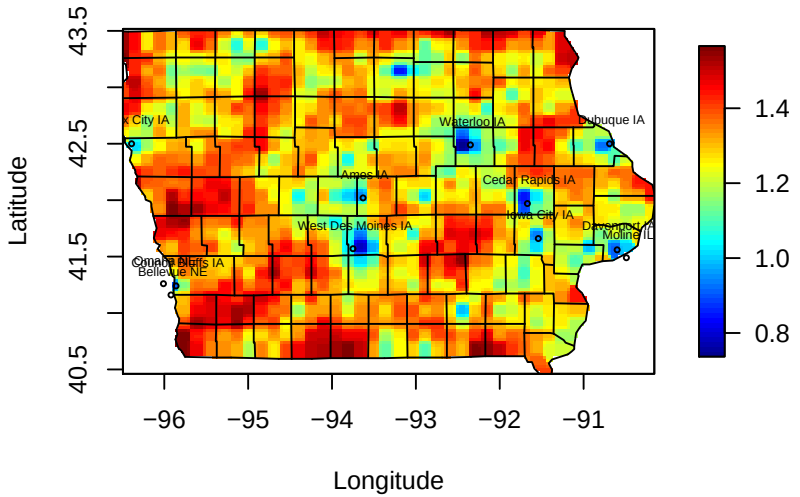
Estimation results

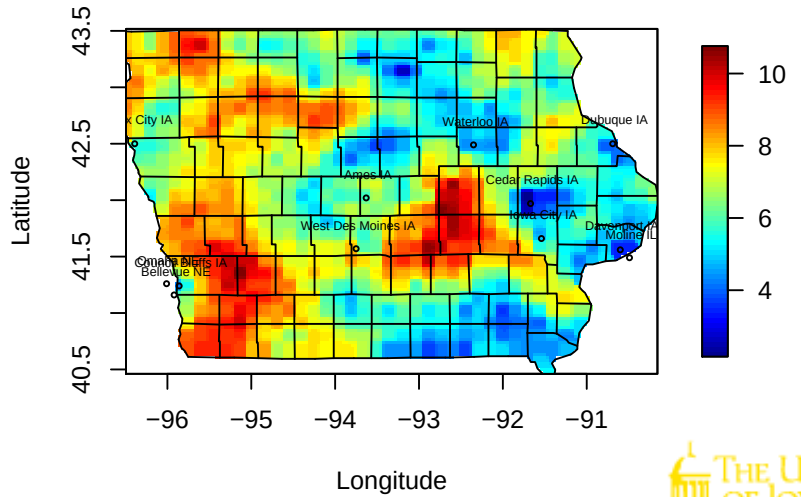
Parameter	IRLCS and SRRS		IRLCS only	
	post mean	95% c.s.	post mean	95% c.s.
$\beta_{0,IRLCS}$	1.533	(1.331, 1.720)	1.533	(1.364, 1.684)
$\beta_{1,IRLCS}$	0.909	(0.707,1.093)	0.908	(0.744, 1.054)
$\beta_{2,IRLCS}$	0.811	(0.607,1.003)	0.809	(0.635, 0.956)
$\beta_{0,SRRS}$	1.825	(1.609,2.023)		
$\beta_{1,SRRS}$	1.683	(1.207,2.184)		
ϕ	29.177	(12.004,74.832)	16.504	(5.596, 40.777)
σ_z^2	0.127	(0.076,0.213)	0.159	(0.091, 0.239)
$\sigma_{e,IRLCS}^2$	0.093	(0.087,0.098)	0.093	(0.087, 0.099)
$\sigma_{e,SRRS}^2$	0.638	(0.425,0.888)		
$\sigma_{re,IRLCS}^2$	0.436	(0.376,0.501)	0.403	(0.336, 0.474)

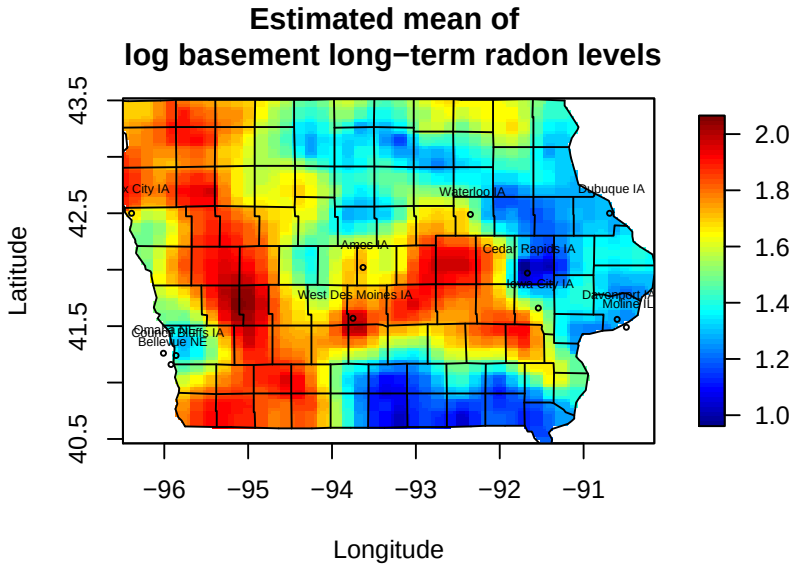
Table: Parameter estimates from Bayesian analyses

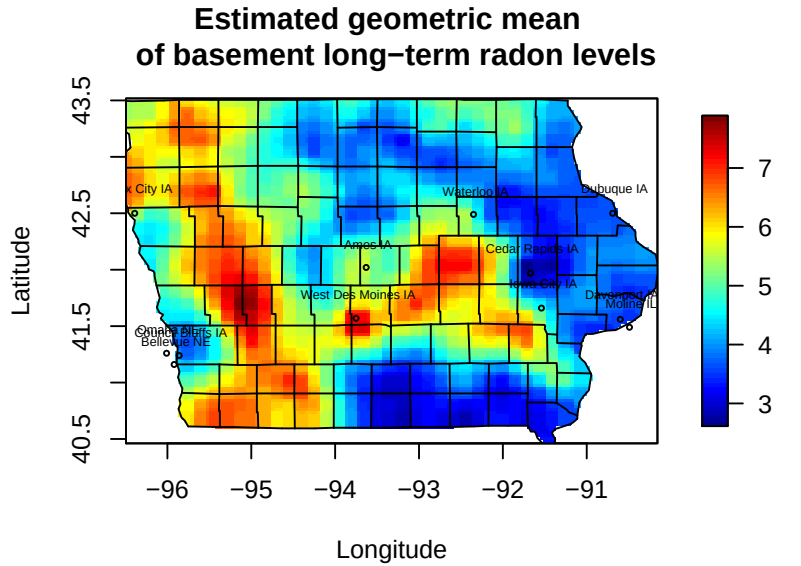

Estimated mean of
log basement long−term radon levels
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
2.0
1.8
1.6
1.4
1.2
1.0
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Omaha Bluffs IA
Bellevue NE

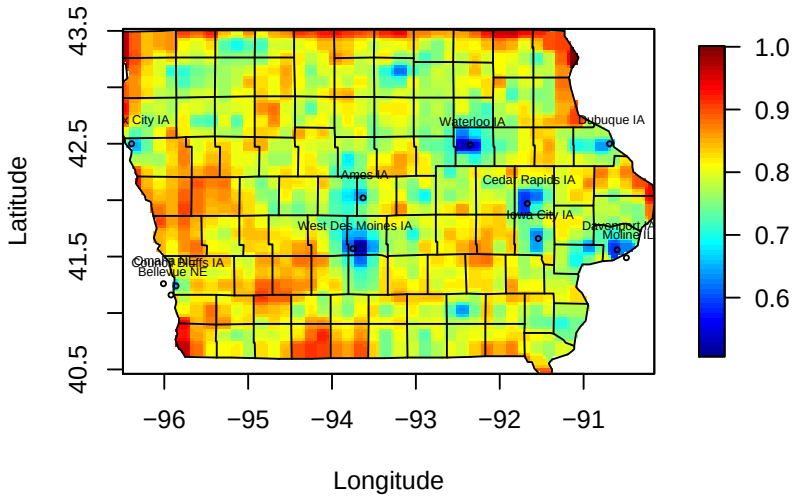

Estimated geometric mean
of basement long−term radon levels
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
7
6
5
4
3
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Omaha Bluffs IA
Bellevue NE

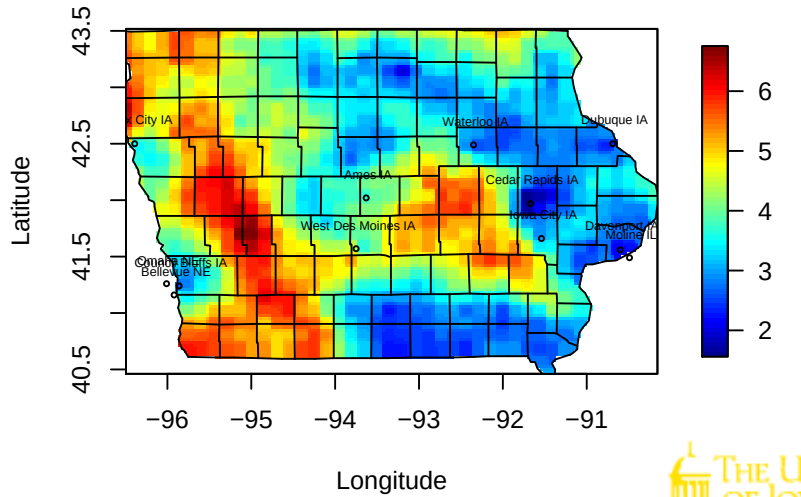

Widths of pointwise
95% credible intervals
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
1.4
1.2
1.0
0.8
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Omaha Bluffs IA
Bellevue NE

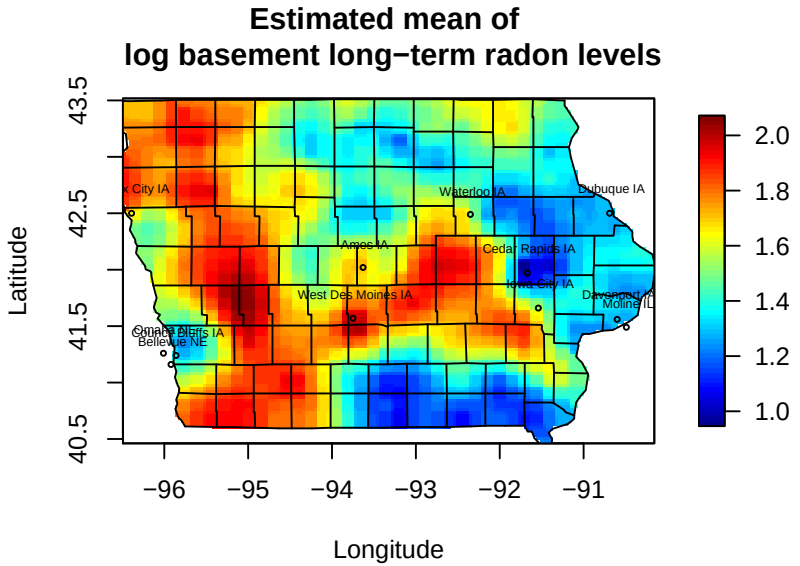

Widths of pointwise
95% credible intervals
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
10
8
6
4
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Omaha Bluffs IA
Bellevue NE

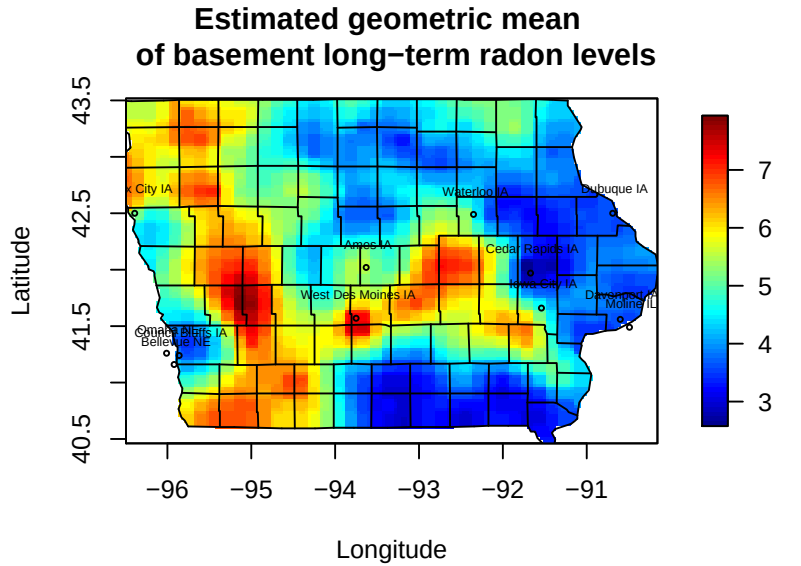

Estimated mean of
log basement long−term radon levels
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
2.0
1.8
1.6
1.4
1.2
1.0
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Omaha Bluffs IA
Bellevue NE

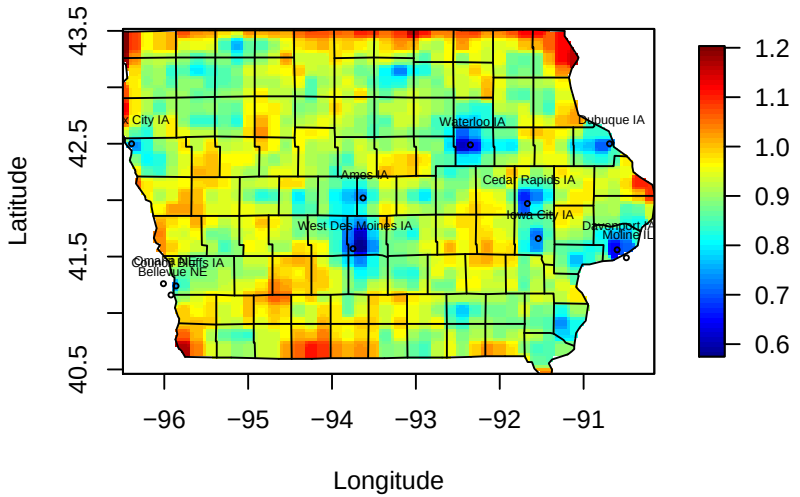

Estimated geometric mean
of basement long−term radon levels
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
7
6
5
4
3
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Omaha Bluffs IA
Bellevue NE

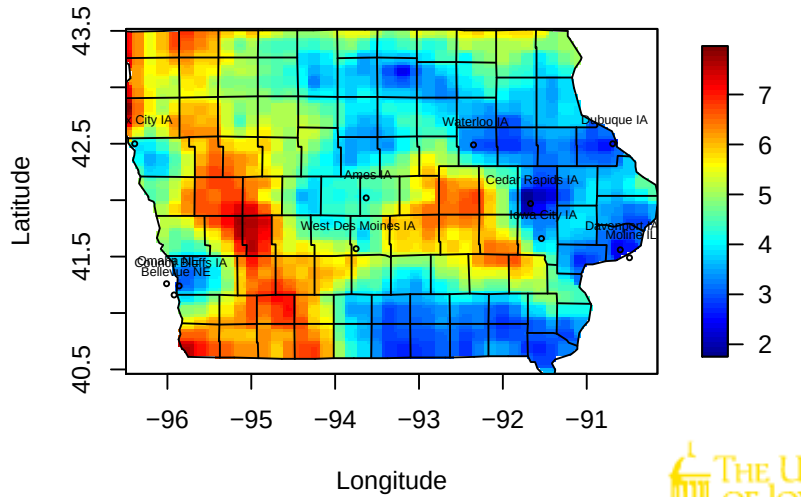

Widths of pointwise
95% credible intervals
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
1.0
0.9
0.8
0.7
0.6
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Omaha Bluffs IA
Bellevue NE


Widths of pointwise
95% credible intervals
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
6
5
4
3
2
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Omaha Bluffs IA
Bellevue NE


Estimated mean of
log basement long−term radon levels
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
2.0
1.8
1.6
1.4
1.2
1.0
City IA
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Davenport
Moline IL
Council Bluffs IA
Omaha
Bellevue NE


Estimated geometric mean
of basement long−term radon levels
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
7
6
5
4
3
City IA
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Davenport
Moline IL
Council Bluffs IA
Omaha
Bellevue NE


Widths of pointwise
95% credible intervals
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
1.2
1.1
1.0
0.9
0.8
0.7
0.6
City IA
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Davenport
Moline IL
Council Bluffs IA
Omaha
Bellevue NE


Widths of pointwise
95% credible intervals
Latitude
Longitude
43.5
42.5
41.5
40.5
−96
−95
−94
−93
−92
−91
7
6
5
4
3
2
City IA
Waterloo IA
Dubuque IA
Ames IA
Cedar Rapids IA
Iowa City IA
West Des Moines IA
Davenport
Moline IL
Council Bluffs IA
Omaha
Bellevue NE

Outline

Conclusions

- Combining data from disparate sources increases precision of prediction.
- RAMPS algorithm works for complex Bayesian geostatistical models.
- Ames may have higher average radon levels than Iowa City.

Model formulation

- again reparameterize in terms of κ, σ^2_{tot}
  - $\sigma^2_{tot} = \sigma^2_z + \sigma^2_{re} + \sigma^2_{e,IRLCS} + \sigma^2_{e,SRRS}$
  - $\kappa_1 = \frac{\sigma^2_z}{\sigma^2_{tot}}$

Likelihood for prediction

$$\mathbf{y} \sim N\left(X\beta + K\mathbf{z},\ \sigma^2_{tot}\ diag\left(\frac{\kappa}{\mathbf{N}}\right)\right)$$

Second and third stages

- second stage

$$\mathbf{z} \sim N\left(\mathbf{0},\ \sigma^2_{tot}\ diag(sqrt(\kappa))\ \Omega(\phi)\ diag(sqrt(\kappa))\right)$$

- third stage: prior distributions on $\kappa,\ \sigma^2_{tot},\ \phi$
  - $phi \sim U(l, r)$
  - κ and σ^2_{tot} get their own slides

Third stage: prior distributions

Suppose that independent inverse gamma priors with parameters α_j and β_j have been placed on each variance σ_j^2. (We parameterize the inverse gamma density such that if $Y \sim IG(a, b)$ then $E(Y) = \frac{b}{a-1}$.) Then the joint prior density induced on **kappa** and σ_{tot}^2 is

$$p(s_1, s_2, \ldots, s_{K-1}, \sigma_{tot}^2) = \prod_{j=1}^{K} \left[\frac{\beta_j}{\Gamma(\alpha_j)} \frac{1}{s_j^{\alpha_j+1}} \right] \frac{1}{(\sigma_{tot}^2)^{\sum_{j=1}^{K} \alpha_j + 1}} exp\left(-\frac{1}{\sigma_{tot}^2} \sum_{j=1}^{K} \frac{\beta_j}{s_j} \right),$$

where $s_K \equiv 1 - \sum_{j=1}^{K-1} s_j$.

This joint density may be factored into a marginal density for **S** times a conditional density for σ^2_{tot} given **S**:

$$p(s_1, s_2, \ldots\ldots, s_{K-1}) = \prod_{j=1}^{K} \left[\frac{\beta_j}{\Gamma(\alpha_j)} \frac{1}{s_j^{\alpha_j+1}} \right] \frac{\Gamma\left(\sum_{j=1}^{K} \alpha_j\right)}{\left(\sum_{j=1}^{K} \frac{\beta_j}{s_j}\right)^{\sum_{j=1}^{K} \alpha_j}}$$

$p(\sigma^2_{tot}|\mathbf{s})$ is inverse gamma with parameters $\sum_{j=1}^{K} \alpha_j$ and $\sum_{j=1}^{K} \frac{\beta_j}{s_j}$.

Reformulation of entire means structure as in Hodges, 1998

$$\left(\begin{array}{c} \mathbf{y} \\ \hline \mathbf{0} \end{array}\right) = \left(\begin{array}{c|c} \mathbf{X} & \mathbf{K} \\ \hline \mathbf{0} & -I \end{array}\right) \left(\begin{array}{c} \beta \\ \hline \mathbf{z} \end{array}\right) + \left(\begin{array}{c} \psi \\ \delta \end{array}\right)$$

or

$$\mathbf{Y} = \mathbf{Q}\,\theta + \mathbf{E}$$

where **Y** consists of known values, **X** is a known design matrix, θ is a vector of unknown parameters, and ψ and δ are error vectors.

Covariance matrix of **E**

$$Cov\begin{pmatrix} \psi \\ \delta \end{pmatrix} = \sigma^2_{tot}\left(\begin{array}{c|c} diag(\frac{\kappa}{\mathbf{n}}) & \mathbf{0} \\ \hline \mathbf{0} & diag(sqrt(\kappa))\ \Omega(\phi)\ diag(sqrt(\kappa)) \end{array}\right)$$

or

$$\sigma^2_{tot}\Omega(\phi,\kappa)$$