Gaussian Multiscale Spatio-temporal Models for Areal Data

Marco A. R. Ferreira (University of Missouri)

Scott H. Holan (University of Missouri)

Adelmo I. Bertolde (UFES)

Outline

Motivation

Multiscale factorization

The multiscale spatio-temporal model

Bayesian analysis

Application: Agricultural Production in Espírito Santo

Unknown multiscale structure

Conclusions

Outline

Motivation

Multiscale factorization

The multiscale spatio-temporal model

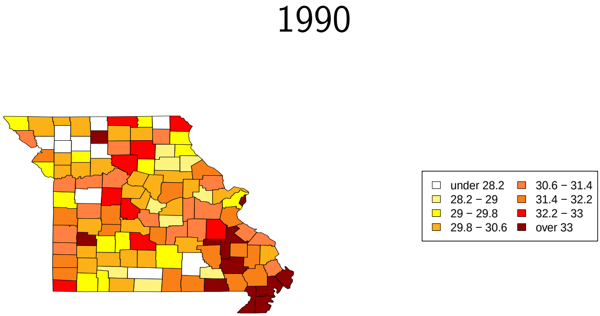
Bayesian analysis

Application: Agricultural Production in Espírito Santo

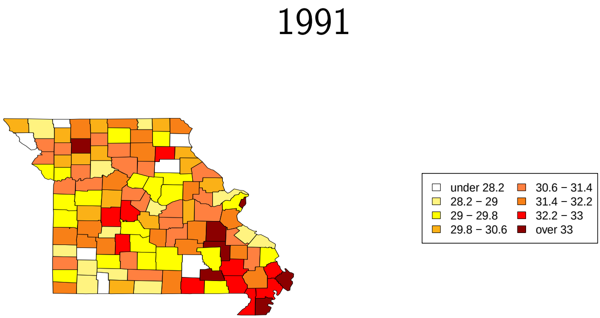
Unknown multiscale structure

Conclusions

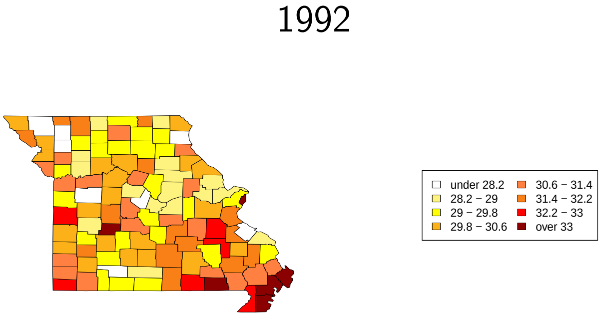
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



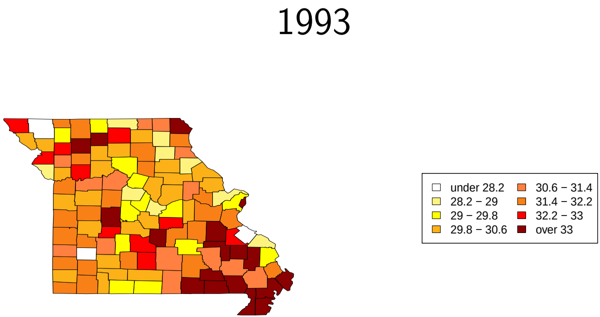
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



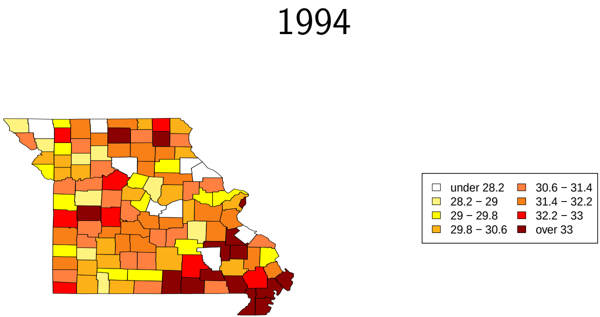
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



Missouri: square root of standardized mortality ratio per $100,000$ inhabitants

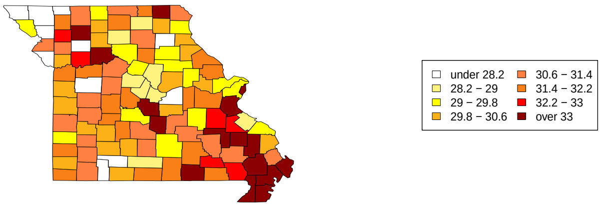


Missouri: square root of standardized mortality ratio per 100, 000 inhabitants

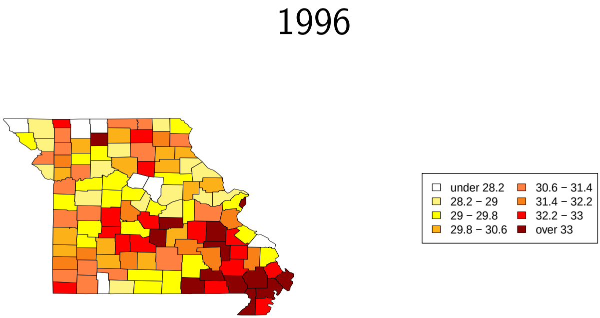


Missouri: square root of standardized mortality ratio per 100,000 inhabitants

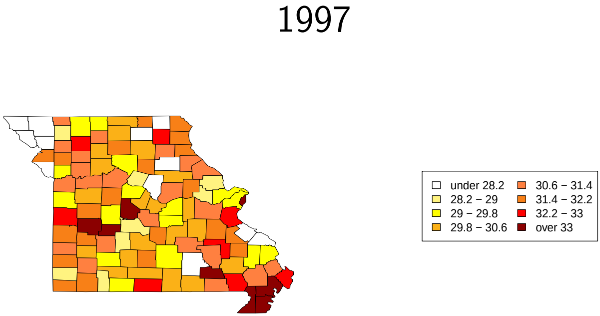
1995



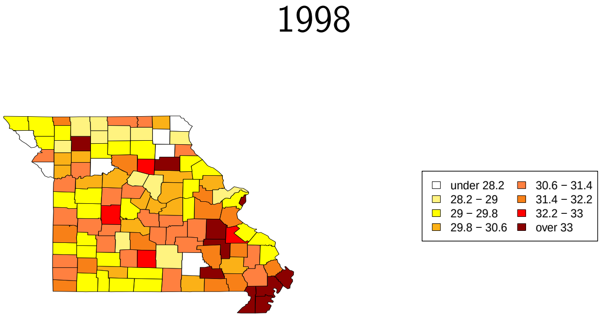
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



Missouri: square root of standardized mortality ratio per 100, 000 inhabitants

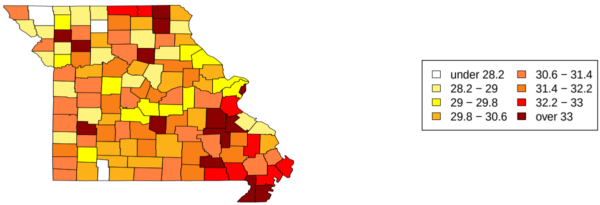


Missouri: square root of standardized mortality ratio per $100,000$ inhabitants

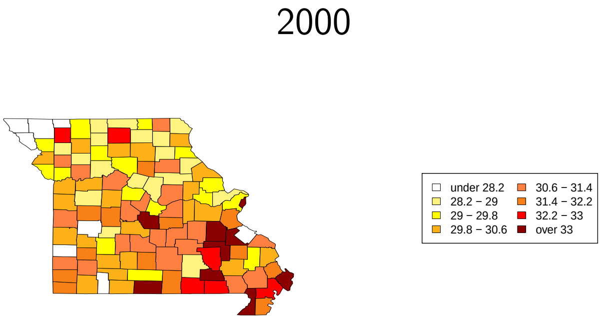


Missouri: square root of standardized mortality ratio per 100, 000 inhabitants

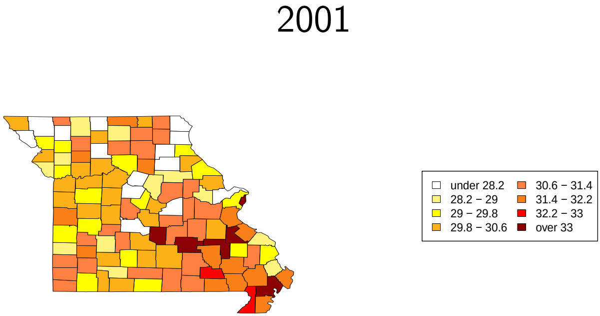
1999



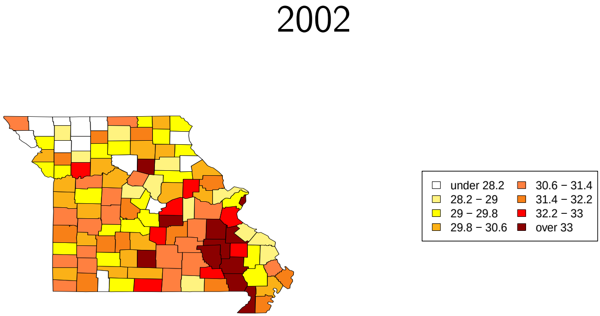
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



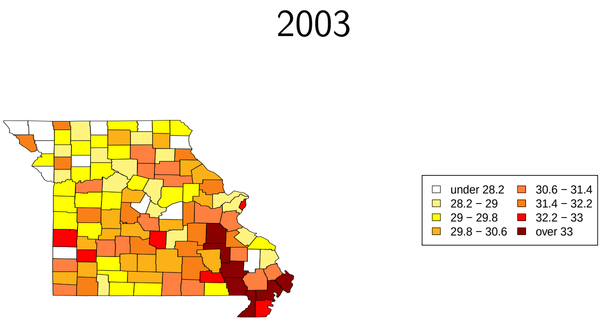
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



Missouri: square root of standardized mortality ratio per 100,000 inhabitants

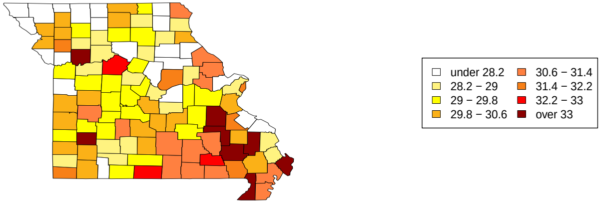


Missouri: square root of standardized mortality ratio per 100, 000 inhabitants

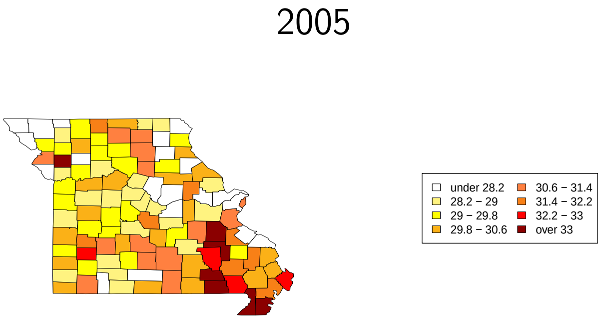


Missouri: square root of standardized mortality ratio per $100,000$ inhabitants

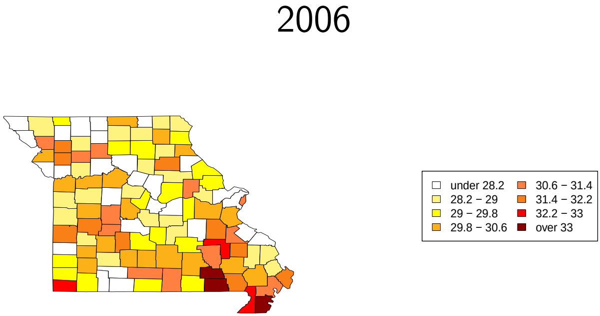
2004



Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



Some background

- Many processes of interest are naturally spatio-temporal.
- Frequently, quantities related to these processes are available as areal data.
- These processes may often be considered at several different levels of spatial resolution.
- Related work on dynamic spatio-temporal multiscale modeling: Berliner, Wikle and Milliff (1999), Johannesson, Cressie and Huang (2007).

Data Structure

Here, the region of interest is divided in geographic subregions or blocks, and the data may be averages or sums over these subregions.

Moreover, we assume the existence of a hierarchical multiscale structure. For example, each state in Brazil is divided into counties, microregions and macroregions.

Geopolitical organization

(a) (b) (c)

Figure: Geopolitical organization of Espírito Santo State by (a) counties, (b) microregions, and (c) macroregions.

Outline

Motivation

Multiscale factorization

The multiscale spatio-temporal model

Bayesian analysis

Application: Agricultural Production in Espírito Santo

Unknown multiscale structure

Conclusions

Multiscale factorization

At each time point we decompose the data into empirical multiscale coefficients using the spatial multiscale modeling framework of Kolaczyk and Huang (2001). See also Chapter 9 of Ferreira and Lee (2007).

Interest lies in agricultural production observed at the county level, which we assume is the L^{th} level of resolution (i.e. the finest level of resolution), on a partition of a domain $S \subset \mathbb{R}^2$.

For the j^{th} county, let y_{Lj}, $\mu_{Lj} = E(y_{Lj})$, and $\sigma^2_{Lj} = V(y_{Lj})$ respectively denote agricultural production, its latent expected value and variance.

Let D_{lj} be the set of descendants of subregion (l, j).
The aggregated measurements at the l^{th} level of resolution are recursively defined by

$$y_{lj} = \sum_{(l+1,j') \in D_{lj}} p_{l+1,j'} y_{l+1,j'}.$$

Analogously, the aggregated mean process is defined by

$$\mu_{lj} = \sum_{(l+1,j') \in D_{lj}} p_{l+1,j'} \mu_{l+1,j'}.$$

Assuming conditional independence,

$$\sigma^2_{lj} = \sum_{(l+1,j') \in D_{lj}} p^2_{l+1,j'} \sigma^2_{l+1,j'}.$$

Define $\boldsymbol{\sigma}_l^2 = (\sigma_{l1}^2, \ldots, \sigma_{ln_l}^2)'$, $\boldsymbol{\rho}_l = \mathbf{p}_l \odot \boldsymbol{\sigma}_l^2$, and $\boldsymbol{\Sigma}_l = \mathrm{diag}(\boldsymbol{\sigma}_l^2)$, where $\odot$ denotes the Hadamard product. Then

$$\mathbf{y}_{D_{lj}} \Big| y_{lj}, \boldsymbol{\mu}_L, \boldsymbol{\sigma}_L^2 \sim N(\boldsymbol{\nu}_{lj} y_{lj} + \boldsymbol{\theta}_{lj}, \boldsymbol{\Omega}_{lj}),$$

with

$$\begin{aligned} \boldsymbol{\nu}_{lj} &= \boldsymbol{\rho}_{D_{lj}} / \sigma_{lj}^2, \\ \boldsymbol{\theta}_{lj} &= \boldsymbol{\mu}_{D_{lj}} - \boldsymbol{\nu}_{lj} \mu_{lj}, \end{aligned}$$

and

$$\boldsymbol{\Omega}_{lj} = \boldsymbol{\Sigma}_{D_{lj}} - \sigma_{lj}^{-2} \boldsymbol{\rho}_{D_{lj}} \boldsymbol{\rho}'_{D_{lj}}.$$

Consider

$$\boldsymbol{\theta}_{lj}^{e} = \mathbf{y}_{D_{lj}} - \boldsymbol{\nu}_{lj} y_{lj},$$

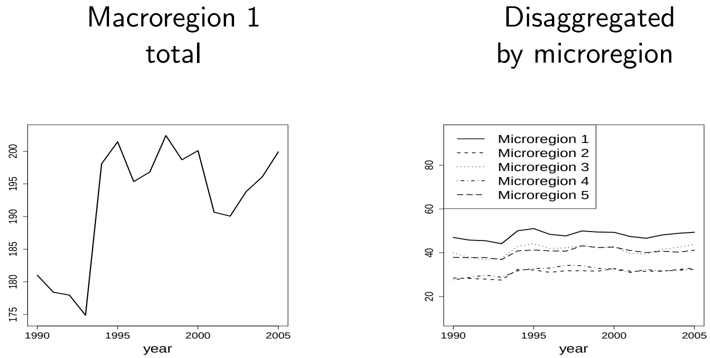
which is an empirical estimate of $\boldsymbol{\theta}_{lj}$.

Then (Kolaczyk and Huang, 2001)

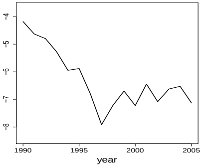
$$\boldsymbol{\theta}_{lj}^{e} | y_{lj}, \boldsymbol{\mu}_L, \boldsymbol{\sigma}_L^2 \sim N(\boldsymbol{\theta}_{lj}, \boldsymbol{\Omega}_{lj}),$$

which is a singular Gaussian distribution (Anderson, 1984).

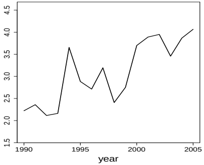
Exploratory Multiscale Data Analysis



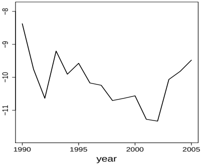
Empirical multiscale coefficient for Macroregion 1



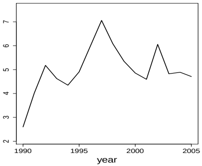
θ^e_{t111}



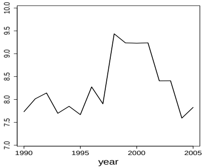
θ^e_{t112}



θ^e_{t113}



θ^e_{t114}



θ^e_{t115}

Outline

Motivation

Multiscale factorization

The multiscale spatio-temporal model

Bayesian analysis

Application: Agricultural Production in Espírito Santo

Unknown multiscale structure

Conclusions

The multiscale spatio-temporal model

Observation equation:

$$\mathbf{y}_{tL} = \boldsymbol{\mu}_{tL} + \mathbf{v}_{tL}, \quad \mathbf{v}_{tL} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_L)$$

where

$$\boldsymbol{\Sigma}_L = diag(\sigma^2_{L1}, \ldots, \sigma^2_{Ln_L}).$$

Multiscale decomposition of the observation equation:

$$y_{t1k} \mid \mu_{t1k} \sim N(\mu_{t1k}, \sigma^2_{1k})$$

$$\boldsymbol{\theta}^e_{tlj} \mid \boldsymbol{\theta}_{tlj} \sim N(\boldsymbol{\theta}_{tlj}, \boldsymbol{\Omega}_{lj})$$

System equations:

$$\mu_{t1k} = \mu_{t-1,1k} + w_{t1k}, \quad w_{t1k} \sim N(0, \xi_k \sigma^2_{1k})$$

$$\boldsymbol{\theta}_{tlj} = \boldsymbol{\theta}_{t-1,lj} + \boldsymbol{\omega}_{tlj}, \quad \boldsymbol{\omega}_{tlj} \sim N(\mathbf{0}, \psi_{lj} \boldsymbol{\Omega}_{lj})$$

Priors

$$\mu_{01k}|D_0 \sim N(m_{01k}, c_{01k}\sigma^2_{1k}),$$

$$\boldsymbol{\theta}_{0lj}|D_0 \sim N(\mathbf{m}_{0lj}, C_{0lj}\boldsymbol{\Omega}_{lj}),$$

$$\xi_k \sim IG(0.5\tau_k, 0.5\kappa_k),$$

$$\psi_{lj} \sim IG(0.5\varrho_{lj}, 0.5\varsigma_{lj}),$$

Outline

Motivation

Multiscale factorization

The multiscale spatio-temporal model

Bayesian analysis

Application: Agricultural Production in Espírito Santo

Unknown multiscale structure

Conclusions

Empirical Bayes estimation of $\boldsymbol{\nu}_{lj}$ and $\boldsymbol{\Omega}_{lj}$

$\boldsymbol{\nu}_{lj}$: vector of relative volatilities of the descendants of (l, j),
$\boldsymbol{\Omega}_{lj}$: singular covariance matrix of the empirical multiscale coefficient of subregion (l, j)

In order to obtain an initial estimate of σ^2_{Lj}, we perform a univariate time series analysis for each county using first-order dynamic linear models (West and Harrison, 1997). These analyses yield estimates $\tilde{\sigma}^2_{Lj}$.

Let $\tilde{\boldsymbol{\rho}}_l = \mathbf{p}_l \odot \tilde{\boldsymbol{\sigma}}^2_l$. We estimate $\boldsymbol{\nu}_{lj}$ and $\boldsymbol{\Omega}_{lj}$ by

$$\begin{aligned} \tilde{\boldsymbol{\nu}}_{lj} &= \tilde{\boldsymbol{\rho}}_{D_{lj}} / \tilde{\sigma}^2_{lj}, \\ \tilde{\boldsymbol{\Omega}}_{lj} &= \tilde{\boldsymbol{\Sigma}}_{D_{lj}} - \tilde{\sigma}^{-2}_{lj} \tilde{\boldsymbol{\rho}}_{D_{lj}} \tilde{\boldsymbol{\rho}}'_{D_{lj}}. \end{aligned}$$

Posterior exploration

Let

$$
\begin{aligned}
\boldsymbol{\theta}_{\bullet lj} &= (\boldsymbol{\theta}'_{0lj}, \ldots, \boldsymbol{\theta}'_{Tlj})', \\
\boldsymbol{\theta}_{t\bullet j} &= (\boldsymbol{\theta}'_{t1j}, \ldots, \boldsymbol{\theta}'_{tLj})', \\
\boldsymbol{\theta}_{\bullet\bullet\bullet} &= (\boldsymbol{\theta}'_{\bullet 11}, \ldots, \boldsymbol{\theta}'_{\bullet 1n_1}, \boldsymbol{\theta}'_{\bullet 21}, \ldots, \boldsymbol{\theta}'_{\bullet 2n_2}, \ldots, \boldsymbol{\theta}'_{\bullet L1}, \ldots, \boldsymbol{\theta}'_{\bullet Ln_L})',
\end{aligned}
$$

with analogous definitions for the other quantities in the model.

It can be shown that, given $\boldsymbol{\sigma}^2_{\bullet}$, $\xi_{\bullet}$, and $\psi_{\bullet\bullet}$, the vectors $\boldsymbol{\mu}_{\bullet 11}, \ldots, \boldsymbol{\mu}_{\bullet 1n_1}$, $\boldsymbol{\theta}_{\bullet 11}, \ldots, \boldsymbol{\theta}_{\bullet 1n_1}, \ldots, \boldsymbol{\theta}_{\bullet L1}, \ldots, \boldsymbol{\theta}_{\bullet Ln_L}$, are conditionally independent *a posteriori*.

Gibbs sampler

- $\boldsymbol{\mu}_{\bullet 1k}$: Forward Filter Backward Sampler (FFBS) (Carter and Kohn, 1994; Fruhwirth-Schnatter,1994).
- $\xi_k|\mu_{\bullet 1k}, \sigma^2_{1k}, D_T \sim IG\left(0.5\tau^*_k, 0.5\kappa^*_k\right)$, where $\tau^*_k = \tau_k + T$ and $\kappa^*_k = \kappa_k + \sigma^{-2}_{1k} \sum_{t=1}^T (\mu_{t1k} - \mu_{t-1,1k})^2$.
- $\psi_{lj}|\boldsymbol{\theta}_{\bullet lj}, D_T \sim IG(0.5\varrho^*_{lj}, 0.5\varsigma^*_{lj})$, where $\varrho^*_{lj} = \varrho_{lj} + T(m_{lj} - 1)$ and $\varsigma^*_{lj} = \varsigma_{lj} + \sum_{t=1}^T (\boldsymbol{\theta}_{tlj} - \boldsymbol{\theta}_{t-1,lj})' \boldsymbol{\Omega}^-_{lj} (\boldsymbol{\theta}_{tlj} - \boldsymbol{\theta}_{t-1,lj})$, where $\boldsymbol{\Omega}^-_{lj}$ is a generalized inverse of $\boldsymbol{\Omega}_{lj}$.
- $\boldsymbol{\theta}_{\bullet lj}$: Singular FFBS.

Singular FFBS

1. Use the singular forward filter to obtain the mean and covariance matrix of $f(\boldsymbol{\theta}_{1lj}|\sigma^2, \psi_{lj}, D_1)$, ..., $f(\boldsymbol{\theta}_{Tlj}|\sigma^2, \psi_{lj}, D_T)$:
   - posterior at $t-1$: $\boldsymbol{\theta}_{t-1,lj}|D_{t-1} \sim N\left(\mathbf{m}_{t-1,lj}, C_{t-1,lj}\boldsymbol{\Omega}_{lj}\right)$;
   - prior at t: $\boldsymbol{\theta}_{tlj}|D_{t-1} \sim N\left(\mathbf{a}_{tlj}, R_{tlj}\boldsymbol{\Omega}_{lj}\right)$, where $\mathbf{a}_{tlj} = \mathbf{m}_{t-1,lj}$ and $R_{tlj} = C_{t-1,lj} + \psi_{lj}$;
   - posterior at t: $\boldsymbol{\theta}_{tlj}|D_t \sim N\left(\mathbf{m}_{tlj}, C_{tlj}\boldsymbol{\Omega}_{lj}\right)$, where $C_{tlj} = (1 + R_{tlj}^{-1})^{-1}$ and $\mathbf{m}_{tlj} = C_{tlj}\left(\boldsymbol{\theta}^e_{tlj} + R_{tlj}^{-1}\mathbf{a}_{tlj}\right)$.
2. Simulate $\boldsymbol{\theta}_{Tlj}$ from $\boldsymbol{\theta}_{Tlj}|\sigma^2, \psi_{lj}, D_T \sim N(\mathbf{m}_{Tlj}, C_{Tlj}\boldsymbol{\Omega}_{lj})$.
3. Recursively simulate $\boldsymbol{\theta}_{tlj}$, $t = T-1, \ldots, 0$, from

$$\boldsymbol{\theta}_{tlj}|\boldsymbol{\theta}_{t+1,lj}, \ldots, \boldsymbol{\theta}_{Tlj}, D_T \;\equiv\; \boldsymbol{\theta}_{tlj}|\boldsymbol{\theta}_{t+1,lj}, D_t \;\sim\; N(\mathbf{h}_{tlj}, H_{tlj}\boldsymbol{\Omega}_{lj}),$$

where $H_{tlj} = \left(C_{tlj}^{-1} + \psi_{lj}^{-1}\right)^{-1}$ and

$$\mathbf{h}_{tlj} = H_{tlj}\left(C_{tlj}^{-1}\mathbf{m}_{tlj} + \psi_{lj}^{-1}\boldsymbol{\theta}_{t+1,lj}\right).$$

Reconstruction of the latent mean process

One of the main interests of any multiscale analysis is the estimation of the latent mean process at each scale of resolution.

From the g^{th} draw from the posterior distribution, we can recursively compute the corresponding latent mean process at each level of resolution using the equation

$$\boldsymbol{\mu}_{t,D_{lj}}^{(g)} = \boldsymbol{\theta}_{tlj}^{(g)} + \boldsymbol{\nu}_{tlj}\mu_{tlj}^{(g)},$$

proceeding from the coarsest to the finest resolution level.

With these draws, we can then compute the posterior mean, standard deviation and credible intervals for the latent mean process.

Outline

Motivation

Multiscale factorization

The multiscale spatio-temporal model

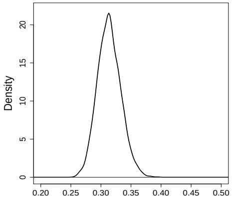
Bayesian analysis

Application: Agricultural Production in Espírito Santo

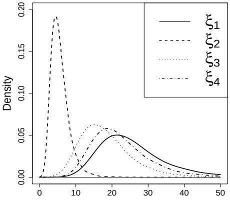
Unknown multiscale structure

Conclusions

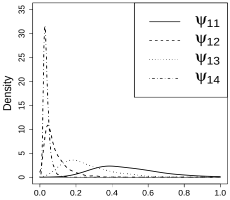
Posterior densities



σ^2

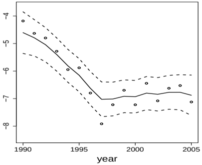


$\xi_1, \ldots, \xi_4$

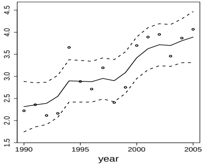


$\psi_{11}, \ldots, \psi_{14}$

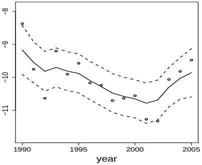
Multiscale coefficient for Macroregion 1



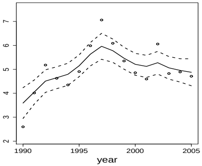
θ_{t111}



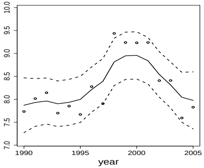
θ_{t112}



θ_{t113}



θ_{t114}



θ_{t115}

Outline

Motivation

Multiscale factorization

The multiscale spatio-temporal model

Bayesian analysis

Application: Agricultural Production in Espírito Santo

Unknown multiscale structure

Conclusions

Discrepancy measurement model

Consider two subregions at the same level of resolution with observations y_{t1} and y_{t2} at time t, observational variances σ_1^2 and σ_2^2, and aggregation weights p_1 and p_2. We define the Discrepancy Measurement Model (DMM) by the dynamic linear model

$$
\begin{aligned}
\begin{pmatrix} y_{t1} \\ y_{t2} \end{pmatrix} &= \begin{pmatrix} p_1\sigma_1^2/(p_1^2\sigma_1^2 + p_2^2\sigma_2^2) & p_2 \\ p_2\sigma_2^2/(p_1^2\sigma_1^2 + p_2^2\sigma_2^2) & -p_1 \end{pmatrix} \begin{pmatrix} \mu_t \\ \Delta_t \end{pmatrix} + \mathbf{v}_t, \\
\begin{pmatrix} \mu_t \\ \Delta_t \end{pmatrix} &= \begin{pmatrix} \mu_{t-1} \\ \Delta_{t-1} \end{pmatrix} + \mathbf{w}_t,
\end{aligned}
$$

where $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V})$, $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W})$, $\mathbf{V} = \mathrm{diag}(\sigma_1^2, \sigma_2^2)$, and $\mathbf{W} = \mathrm{diag}(W_\mu, W_\Delta)$.

Relative discrepancy

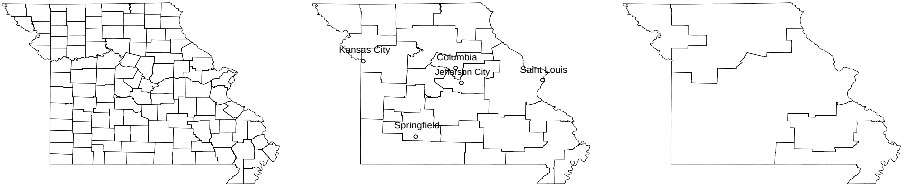
Consider the DMM model and two subregions at the same resolution level. We define the *relative discrepancy* between these two subregions as W_{Δ}/W_{μ}.

Multiscale clustering algorithm

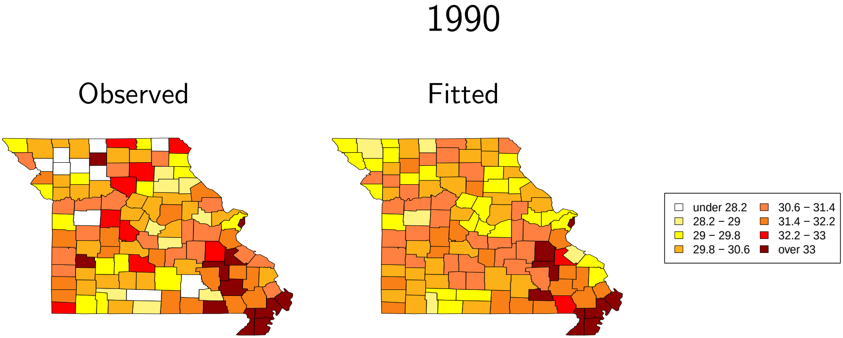
1. For each county, create a link to one of its neighbors which minimizes the estimated relative discrepancy.
2. Create intermediate-level clusters of counties. Counties are in the same cluster if and only if there is a path connecting them.
3. Obtain the data at the intermediate-level through aggregation.
4. Apply steps 1 and 2 to the intermediate-level clusters in order to obtain the coarse-level clusters.

We illustrate the algorithm with a dataset on mortality in Missouri.

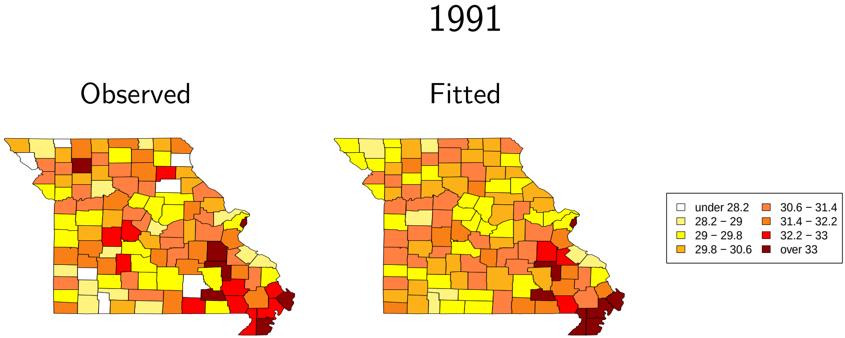
Estimated multiscale structure for the State of Missouri



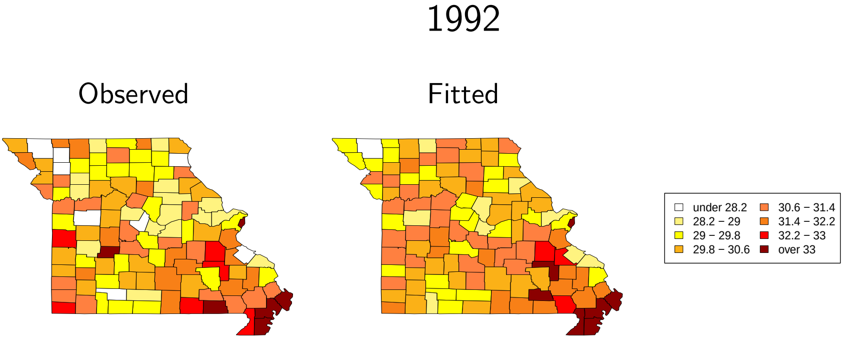
Missouri: square root of standardized mortality ratio per 100,000 inhabitants



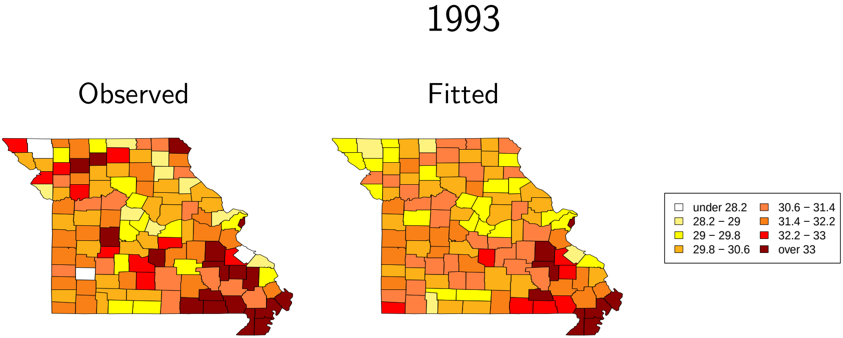
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



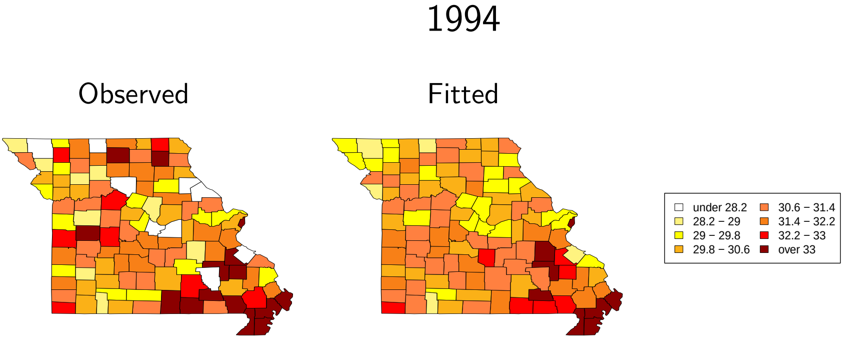
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



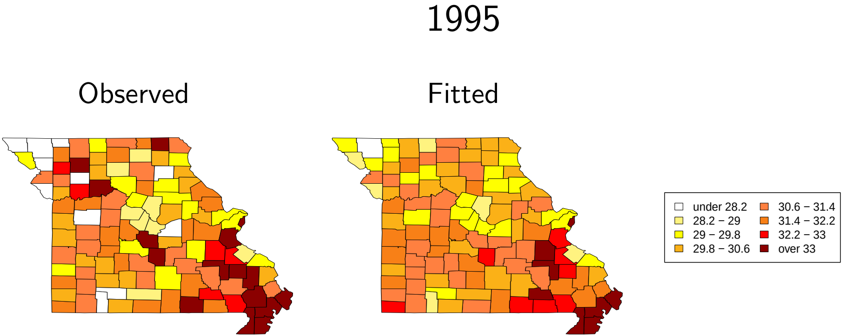
Missouri: square root of standardized mortality ratio per $100,000$ inhabitants



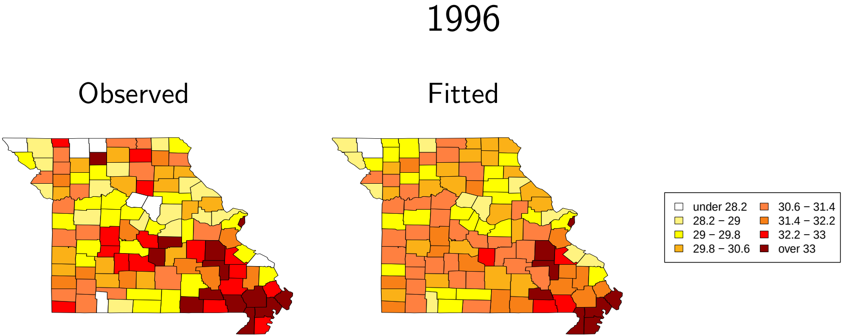
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



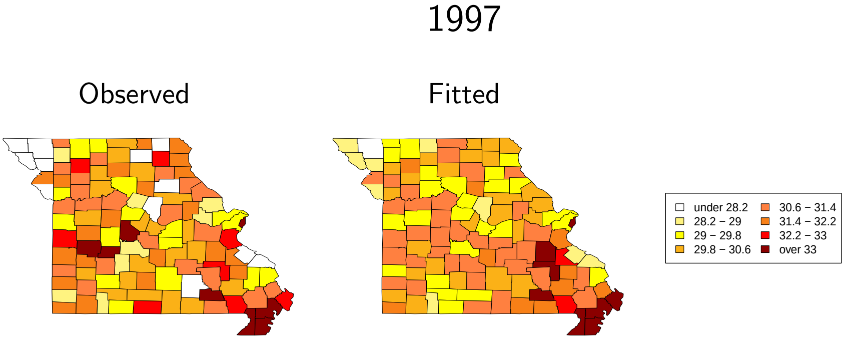
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



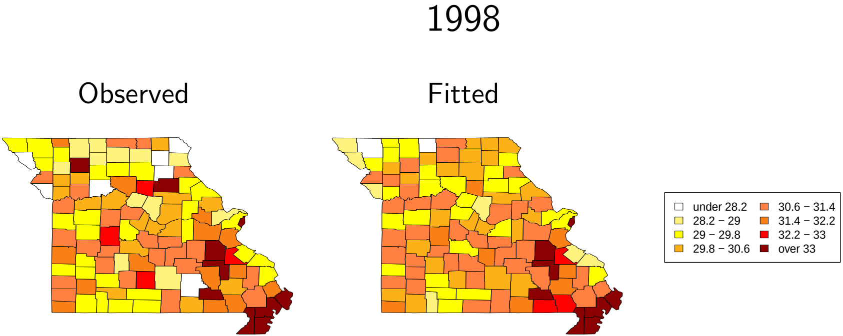
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



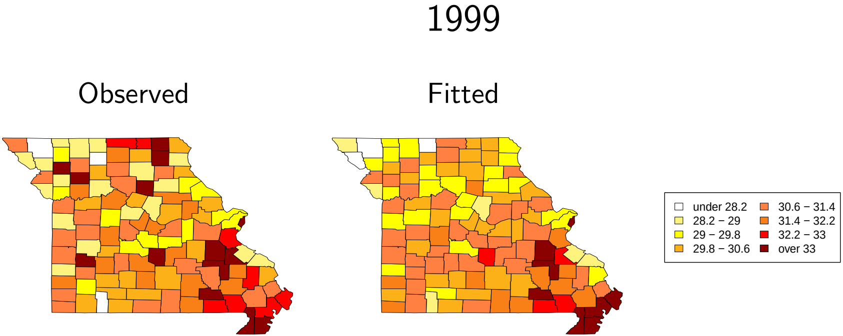
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



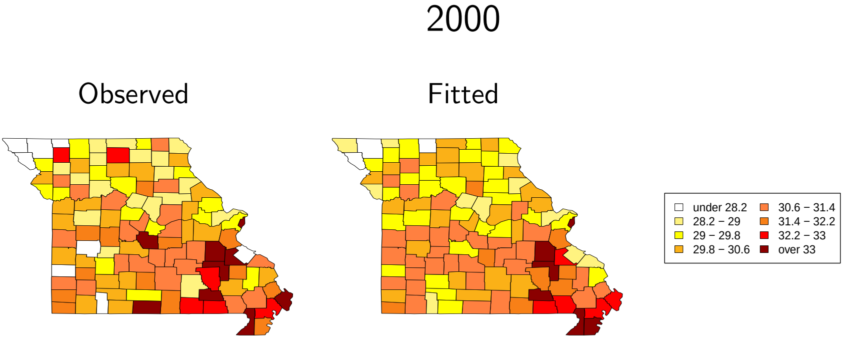
Missouri: square root of standardized mortality ratio per 100,000 inhabitants



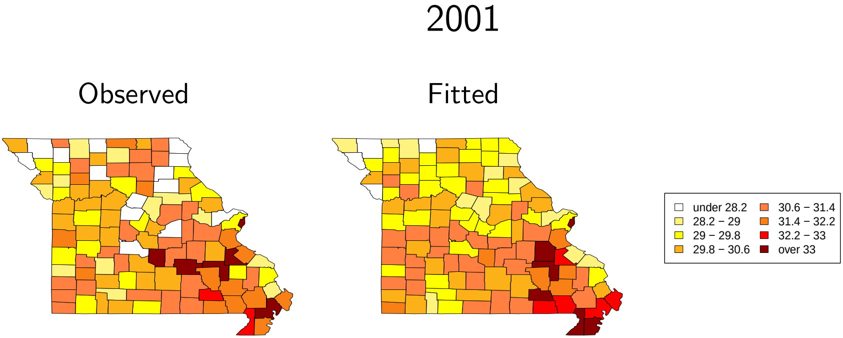
Missouri: square root of standardized mortality ratio per 100,000 inhabitants



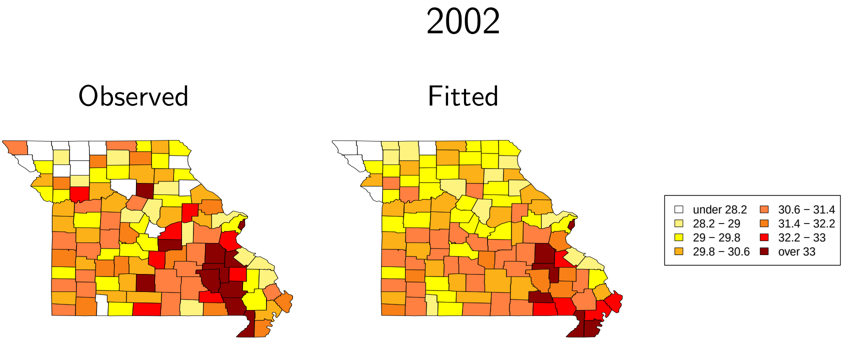
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



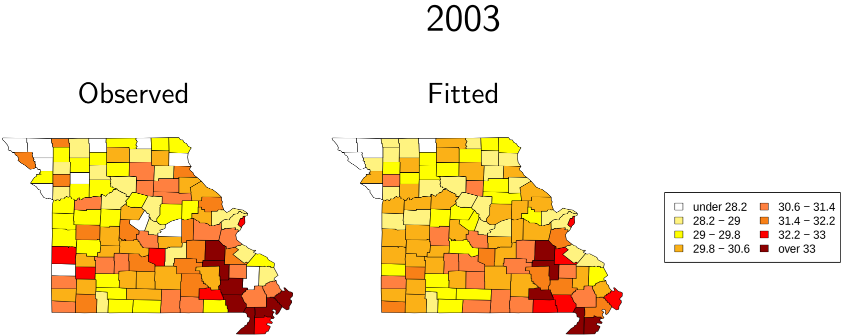
Missouri: square root of standardized mortality ratio per 100,000 inhabitants



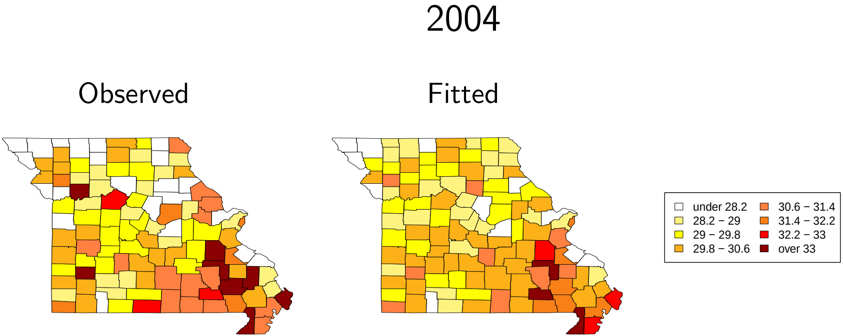
Missouri: square root of standardized mortality ratio per 100, 000 inhabitants



Missouri: square root of standardized mortality ratio per $100,000$ inhabitants



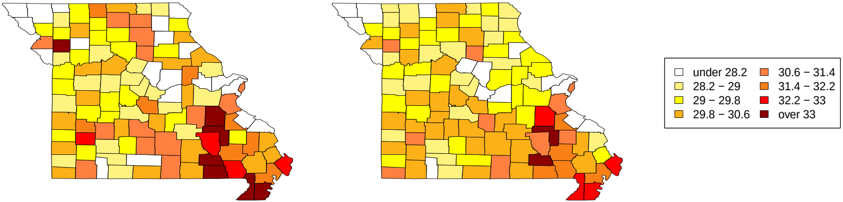
Missouri: square root of standardized mortality ratio per 100,000 inhabitants



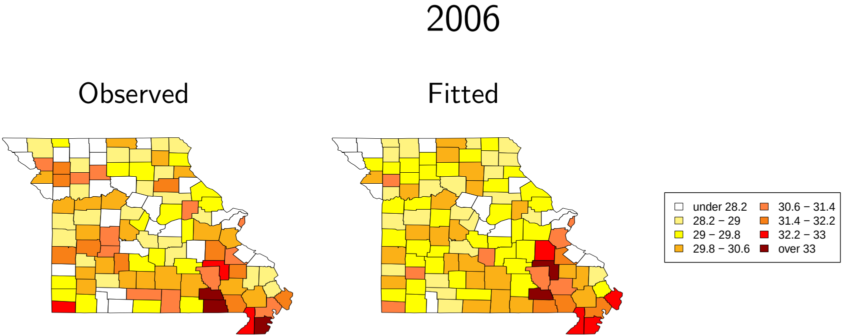
Missouri: square root of standardized mortality ratio per $100,000$ inhabitants

2005

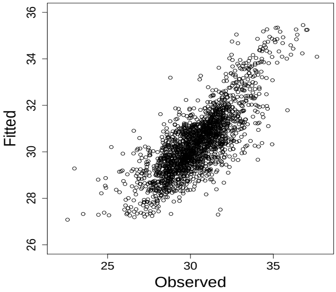
Observed Fitted

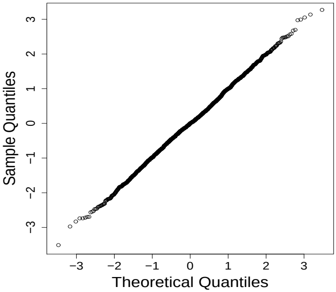


Missouri: square root of standardized mortality ratio per $100,000$ inhabitants



State of Missouri application. Residuals check.





Outline

Motivation

Multiscale factorization

The multiscale spatio-temporal model

Bayesian analysis

Application: Agricultural Production in Espírito Santo

Unknown multiscale structure

Conclusions

Conclusions

- New multiscale spatio-temporal model for areal data.
- Estimated multiscale coefficients shed light on similarities and differences between regions within each scale of resolution.
- Modeling strategy naturally respects nonsmooth transitions between geographic subregions.
- Our divide-and-conquer modeling strategy leads to computational procedures that are scalable and fast.