Model selection for a class of spatio-temporal models for areal data

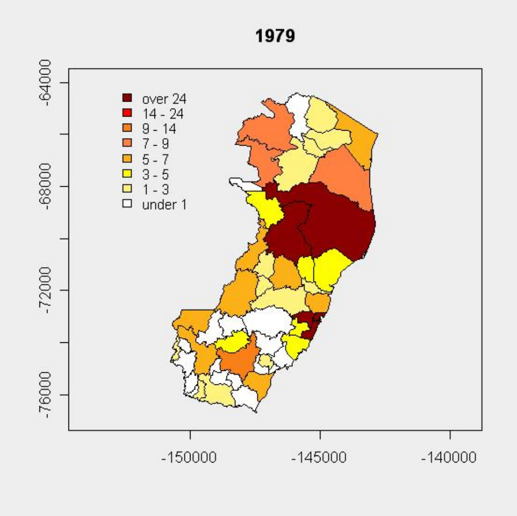
Juan C. Vivar
DME - UFRJ
jcvivar@dme.ufrj.br

and

Marco A. R. Ferreira
University of Missouri - Columbia
ferreiram@missouri.edu

GEOINFO 2007
IX Brazilian Symposium on GeoInformatics
November $27^{th}, 2007$

Contents



Homicide data - State of Espírito Santo

Objective

Main objective

Perform model selection based on predictive density in a class of spatio-temporal generalized dynamic models for areal data.

Our general model for count data

$$
\begin{aligned}
y_{ts}|\lambda_{ts} &\sim Po(n_{ts}\lambda_{ts}), \qquad (1)\\
p(y_{ts}|\eta_{ts}) &\propto \exp\{y_{ts}\eta_{ts} - \exp(\eta_{ts}) + \log(y_{ts}!)\},\\
\eta_{ts} &= \theta_{ts} + \log(n_{ts}),\\
\boldsymbol{\theta}_t &= \mathbf{F}'_t\boldsymbol{\beta}_t, \qquad \boldsymbol{\theta}_t = (\theta_{t1}, \ldots, \theta_{tS})'\\
\boldsymbol{\beta}_t &= \mathbf{G}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim PGMRF(\mathbf{0}, \mathbf{W}_t^{-1}),
\end{aligned}
$$

- For each year t and county s, y_{ts} represents the observed number of homicides; $t = 1, \ldots, 20$ and $s = 1, \ldots, 52$, n_{ts} is the population size
- $\boldsymbol{\beta}_{(1:T)}$ is the latent process of interest, $\boldsymbol{\beta}_{(1:T)} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T)$
- $\mathbf{W}_t$ describes the covariance structure of the errors

PGMRF = proper Gaussian Markov random field

$\mathbf{Z} \sim PGMRF(\boldsymbol{\mu}, \mathbf{P})$ means that

$$p(\mathbf{z}) \propto \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\mathbf{P}(\mathbf{z}-\boldsymbol{\mu})\right)$$

with $\mathbf{P} = \tau(\mathbf{I}_S + \phi\mathbf{M})$, where $\mathbf{I}_S$ is a $S \times S$ identity matrix, $\mathbf{M}$ is the neighborhood matrix, τ is a scale parameter and $\phi \geq 0$ controls the degree of spatial correlation

Bayesian inference

Markov chain Monte Carlo (MCMC) methods to obtain samples from the posterior distributions of the unknown quantities.

Markov chain partitioned in two blocks:

✓ Simulation of the parameter vector ψ (ψ is specific for each model):

- Gibbs sampling
- Metropolis-Hastings algorithm

✓ Simulation of the latent process $\boldsymbol{\beta}_{(1:T)}$:

- Using an efficient scheme named FEIFBS (*forward extended information filter backward sampler)*

Some specific models

Different specifications of $\mathbf{F}_t$ and $\mathbf{G}_t$ in (1) lead to the following models:

Model I - First-order temporal trend

- $\mathbf{F}'_t = \mathbf{I}_S$ and $\mathbf{G}_t = \mathbf{I}_S$,
- $\mathbf{W}_t^{-1} = \tau(\mathbf{I}_S + \phi\mathbf{M})$.

Model II - Contamination

- $\mathbf{F}'_t = \mathbf{I}_S$,
- $\mathbf{G}_t = \frac{1}{1+\kappa h}\mathbf{H} \longrightarrow \{\mathbf{H}\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa, & k \in N_l, \\ 0, & o.c. \end{cases}$
- $\mathbf{W}_t^{-1} = \tau(\mathbf{I}_S + \phi\mathbf{M})$.

Some specific models

Different specifications of $\mathbf{F}_t$ and $\mathbf{G}_t$ in (1) lead to the following models:

Model I - First-order temporal trend

- $\mathbf{F}'_t = \mathbf{I}_S$ and $\mathbf{G}_t = \mathbf{I}_S$,
- $\mathbf{W}_t^{-1} = \tau(\mathbf{I}_S + \phi\mathbf{M})$.

Model II - Contamination

- $\mathbf{F}'_t = \mathbf{I}_S$,
- $\mathbf{G}_t = \frac{1}{1+\kappa h}\mathbf{H} \longrightarrow \{\mathbf{H}\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa, & k \in N_l, \\ 0, & o.c. \end{cases}$
- $\mathbf{W}_t^{-1} = \tau(\mathbf{I}_S + \phi\mathbf{M})$.

Some specific models

Models III and IV - Second-order temporal trend

- $\mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{G}_{1t} \\ \mathbf{0}_S & \mathbf{G}_{2t} \end{pmatrix}$, $\mathbf{G}_{it} = \mathbf{I}_S, i = 1, 2.$
- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0}_S \\ \mathbf{0}_S & \mathbf{W}_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i \mathbf{M}), i = 1, 2.$

Model III considers $\phi_2 = 0$.

Some specific models

Models V, VI and VII - Second-order including contamination

- Model V: In velocity equation and $\phi_2 = 0$

$$\mathbf{G}_{1t} = \mathbf{I}_S \text{ and } \mathbf{G}_{2t} = \frac{1}{1+\kappa_2 h}\mathbf{H} \longrightarrow \{\mathbf{H}\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa_2, & k \in N_l, \\ 0, & o.c. \end{cases}$$

- Model VI: In he level equation

$$\mathbf{G}_{2t} = \mathbf{I}_S \text{ and } \mathbf{G}_{1t} = \frac{1}{1+\kappa_1 h}\mathbf{H} \longrightarrow \{\mathbf{H}\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa_1, & k \in N_l, \\ 0, & o.c. \end{cases}$$

- Model VII: In both equations and $\phi_1 = 0$

$$\mathbf{G}_{it} = \frac{1}{1+\kappa_i h}\mathbf{H}_i \longrightarrow \{\mathbf{H}_i\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa_i, & k \in N_l, \\ 0, & o.c. \end{cases}, i = 1, 2.$$

Some specific models

Model VIII - Second-order with same acceleration for all regions

- $\mathbf{F}_t' = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{1} \\ \mathbf{0} & G_{2t} \end{pmatrix}$, $\mathbf{G}_{1t} = \mathbf{I}_S, G_{2t} = 1.$
- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0} \\ \mathbf{0} & W_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{1t}^{-1} = \tau_1(\mathbf{I}_S + \phi_1 \mathbf{M}),\ W_{2t}^{-1} = \tau_2$

Predictive distribution

The predictive distribution $p_q(\mathbf{y}_t|\mathbf{D}_{t-1})$ under model q is:

$$p_q(\mathbf{y}_t|\mathbf{D}_{t-1}) \;=\; \int p_q(\mathbf{y}_t|\boldsymbol{\beta}_{t-1},\psi)p_q(\boldsymbol{\beta}_{1:t-1},\psi|\mathbf{D}_{t-1})d\boldsymbol{\beta}_{1:t-1}d\psi$$

where $\mathbf{D}_t = (\mathbf{y}_1, \ldots, \mathbf{y}_t)$, and can be approximated by

$$\hat{p}_q(\mathbf{y}_t|\mathbf{D}_{t-1}) = \frac{1}{L}\sum_{l=1}^{L} p_q(\mathbf{y}_t|\beta_{t-1}^{(l)},\psi^{(l)}) \qquad (2)$$

Our criterion

An estimate of the joint predictive density under model q is

$$\hat{p}_q(\mathbf{y}_{t^*}, \mathbf{y}_{t^*+1}, \ldots, \mathbf{y}_T | \mathbf{D}_{t^*-1}) = \prod_{t=t^*+1}^{T} \hat{p}_q(\mathbf{y}_t | \mathbf{D}_{t-1}), \qquad (3)$$

t^* is the training sample such that $p_q(\psi | \mathbf{D}_{t^*})$ is proper for all $q = 1, \ldots, Q$.

Application

We fit each model within the Bayesian paradigm

Parameter	Prior	Simulation
$\boldsymbol{\beta}_0$	$N(\mathbf{m}_0, \mathbf{C}_0^{-1})$	FEIFBS
τ_i	$Ga(4,4)$	Gibbs sampler
ϕ_i	$\propto \begin{cases} 1, & 0 < \phi_i < 1 \\ \phi_i^{-5}, & \phi_i \geq 1 \end{cases}$	Metropolis-Hastings
κ_i	$U(0,1)$	Metropolis-Hastings

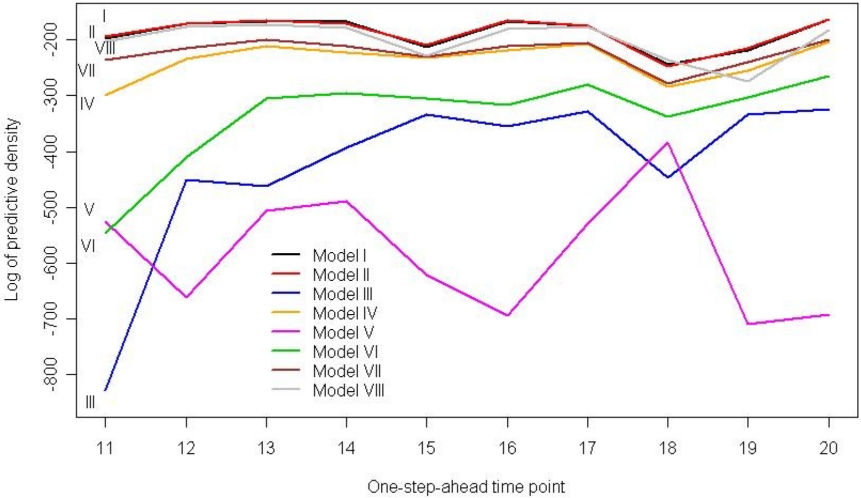
using a training sample of $t^* = 10$.

Results

Table: Logarithm of the predictive density and mean squared error of the prediction for all the considered models.

Model	Log p.d.	MSE
I	-1881.40	1766.06
II	**-1873.36**	**1749.60**
III	-4254.69	13164.13
IV	-2364.12	3033.33
V	-5815.18	13857.93
VI	-3365.69	8094.99
VII	-2227.64	2337.85
VIII	-2008.27	2228.14

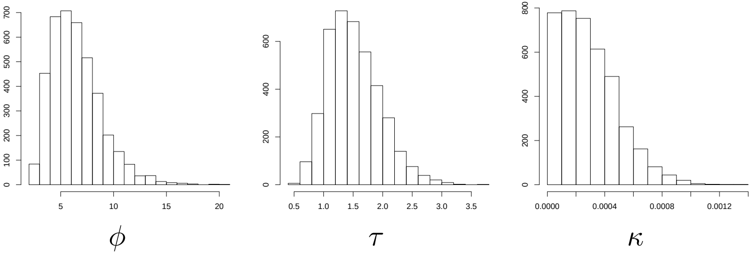
One-step ahead logarithm of the predictive density

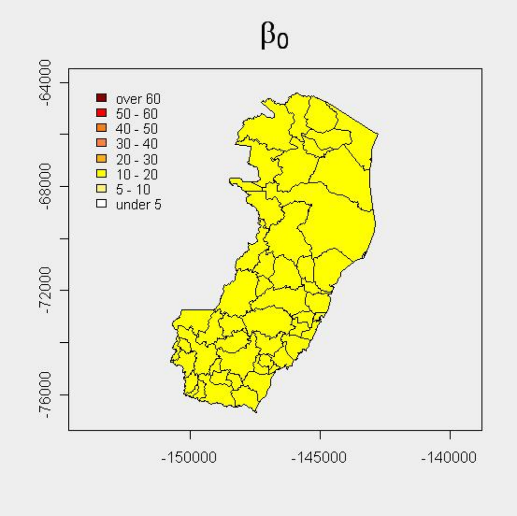


Results for contamination model

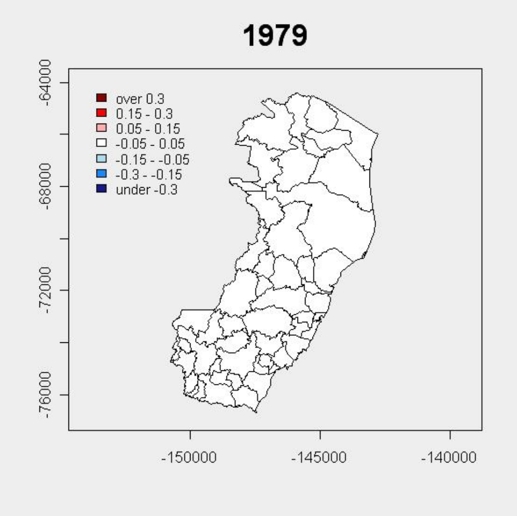
Table: Posterior results ($\hat{R}$ - statistic of Gelman & Rubin).

Parameter	Mean	s.d.	2.5%	50%	97.5%	$\hat{R}$
ϕ	6.47	2.38	3.08	6.11	10.2	1.0
τ	1.52	0.44	0.08	1.46	2.51	1.0
κ	0.0003	0.0002	0.00001	0.0003	0.0008	1.0





Posterior means of estimated latent process



Posterior means of innovations - Spatially structured effects

Conclusions

- We presented a method to perform model selection based on predictive density.
- Flexible class of models permitted several specifications of spatio-temporal models for the data: the annual number of homicides in the State of Espírito Santo in the 1979-1998 period.
- We described an efficient MCMC scheme with embedded FEIFBS to perform Bayesian inference for the parameters and the latent process.
- The selection method indicated the contamination model as the best model among the proposed ones.