IX Brazilian Symposium on GeoInformatics, Campos do Jordão, Brazil, November 25-28, 2007, INPE, p. 121-132.

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

Juan C. Vivar[1], **Marco A. R. Ferreira**[2]

[1]Departamento de Métodos Estatísticos
Universidade Federal do Rio de Janeiro (UFRJ) – RJ – Brazil

[2]University of Missouri
Columbia – U.S.A.

jcvivar@dme.ufrj.br, ferreiram@missouri.edu

Abstract. *We present a method to perform model selection based on predictive density in a class of spatio-temporal dynamic generalized linear models for areal data. These models assume a latent random field process that evolves through time with random field convolutions; the convolving fields follow proper Gaussian Markov random field processes. Parameter and latent process estimation based on Markov Chain Monte Carlo and the forward information filter backward sampler, respectively, is showed. Finally, an application using several specifications of the general model on homicide data in the State of Espírito Santo is presented showing the results of model selection.*

1. Introduction

The last decade has seen an upsurge of research on spatio-temporal modelling. Massive amounts of spatio-temporal data have become available and the increasing power of computers has made possible the analysis of these datasets with progressively realistic models. Several spatio-temporal models have been proposed in the literature for specific applications for point-referenced data. Examples include meteorology [Ghil et al. 1981], ozone analysis [Guttorp et al. 1994], prediction of snow water [Huang and Cressie 1996], calibration of radar rainfall data [Brown et al. 2001], and analysis of pollutant levels [Huerta et al. 2004]. For areal data have been developed in the disease mapping literature [Bernardinelli et al. 1995, Waller et al. 1997, Knorr-Held 2000, Knorr-Held and Richardson 2003, Schmid and Held 2004].

In [Vivar and Ferreira 2007] we proposed a linear Gaussian spatio-temporal models for areal data that uses *proper Markov random fields*. These models can be cast within a state-space formulation [West and Harrison 1997]. More specifically, we considered a latent random field process that evolves through time with random field convolutions; the convolving fields follow proper Gaussian Markov random field (PGMRF) processes. At each time, the latent random field process is linearly related to observations through an observation equation with errors that also follow a PGMRF.

The spatio-temporal model is

$$\mathbf{y}_t = \mathbf{F}_t'\beta_t + \epsilon_t, \qquad \epsilon_t \sim PGMRF(\mathbf{0}_S, \mathbf{V}_t^{-1}), \tag{1}$$

$$\beta_t - \mu_\beta = \mathbf{G}_t(\beta_{t-1} - \mu_\beta) + \omega_t, \qquad \omega_t \sim PGMRF(\mathbf{0}_S, \mathbf{W}_t^{-1}), \tag{2}$$

where $\mathbf{0}_S$ is the S-dimensional null vector and the errors $\epsilon_1, \ldots, \epsilon_T$ and the system innovations $\omega_1, \ldots, \omega_T$ are independent. The matrix $\mathbf{F}_t$ connects the latent random field process to the field observations, the matrix $\mathbf{G}_t$ describes the spatio-temporal evolution of the process, ω_t is the state innovation field, and the covariance matrices $\mathbf{V}_t$ and $\mathbf{W}_t$ describe the covariance structure of the observational and system errors, respectively. The mean level field μ_β describes the temporal stationary expected behavior of the latent process. It is only defined if the process is temporally stationary, otherwise, it is omitted from the model.

Following the notation of [Ferreira and De Oliveira 2007], $\mathbf{Z} \sim PGMRF(\mu_z, \mathbf{P})$ means that the variable $\mathbf{Z}$ follows a PGMRF process with mean vector μ_z and precision matrix $\mathbf{P}$, that is, the density function of $\mathbf{Z}$ is proportional to $\exp\left(-\frac{1}{2}(\mathbf{z}-\mu_z)'\mathbf{P}(\mathbf{z}-\mu_z)\right)$, where $\mathbf{P} = \tau(\mathbf{I}_S + \phi\mathbf{M})$ where $\mathbf{M}$ is called the neighborhood matrix, τ is a scale parameter, $\mathbf{I}_S$ is the $S \times S$ identity matrix and $\phi \geq 0$ controls the degree of spatial correlation. When $\phi = 0$, the regions are independent and the spatial dependence increases when ϕ increases. See [Vivar and Ferreira 2007, Vivar 2004] for more details, special cases and properties of this class of models.

Because this is a class of linear models it has a good performance with Gaussian or approximately Gaussian data. A natural extension is a model that deals with observations that are distributed in the exponential family. In the next Section we present a specific model for count data. Section 3 is developed Bayesian inference for the parameters. Section 4 describes several specifications of the general model and the method to perform model selection. An application using homicide data in the State of Espírito Santo is shown in Section 5.

2. Spatio-temporal model for count data

Our data consists on the annual number of homicides per county in the State of Espírito Santo, Brazil, from 1979 to 1998. During the 80s and 90s new counties were created in the State of Espírito Santo by fusion or division of older counties. For compatibility purposes, we use here the political map of 1979 in a total of 52 counties.

For each year t and county s, $t = 1, \ldots, T$, $s = 1, \ldots, S$, let n_{ts} denote the population size and y_{ts} the observed number of homicides. As it is typical for count data such as these, we assume that y_{ts} follows a Poisson distribution. More specifically, we assume that $y_{ts}|\lambda_{ts} \sim Po(n_{ts}\lambda_{ts})$, where λ_{ts} is the underlying risk at time t in county s. The Poisson distribution belongs to the exponential family distributions with canonical parameter $\eta_{ts} = \log(\mu)$. Specifically, we identify the following components

- Mean $\mu_{ts} = \exp(\eta_{ts})$ and variance $\Sigma_{ts} = \exp(\eta_{ts})$.
- Linear predictor: $\theta_{ts} = \log(\lambda_{ts})$.
- Link function: $g(\mu_{ts}) = \log(\mu_{ts}/n_{ts}) = \theta_{ts}$.
- Response function: $f(\theta_{ts}) = n_{ts}\exp(\theta_{ts}) = \mu_{ts}$.
- The function $\gamma(\theta_{ts})$ that transform directly the canonical parameter into the linear predictor:

$$\begin{aligned} \eta_{ts} &= \log(\lambda_{ts} n_{ts}) \\ &= \log(\lambda_{ts}) + \log(n_{ts}) \\ &= \theta_{ts} + \log(n_{ts}) \\ &= \gamma(\theta_{ts}) \end{aligned}$$

Then, our general spatio-temporal model for count data results as follows:

$$\begin{aligned} p(y_{ts}|\eta_{ts}) &\propto \exp\{y_{ts}\eta_{ts} - \exp(\eta_{ts}) + \log(y_{ts}!)\}, \\ \eta_{ts} &= \theta_{ts} + \log(n_{ts}), \\ \theta_t &= \mathbf{F}_t'\beta_t, \qquad \theta_t = (\theta_{t1}, \ldots, \theta_{tS})' \\ \beta_t &= \mathbf{G}_t\beta_{t-1} + \omega_t, \qquad \omega_t \sim PGMRF(\mathbf{0}, \mathbf{W}_t^{-1}), \end{aligned} \tag{3}$$

Our main interest is to make inference for $\beta_{(1:T)} = (\beta_1', \ldots, \beta_T')'$.

3. Bayesian inference

When the matrices $\mathbf{F}_t, \mathbf{G}_t$ and $\mathbf{W}_t$ are completely known, the extended Kalman filter can be used to perform inference about the latent process β_t. But in practice these matrices are known only up to the parameter vector ψ and numerical integration methods are required for the Bayesian statistical analysis. Here we favor Markov chain Monte Carlo (MCMC) methods [Gamerman and Lopes 2006, Robert and Casella 1999] that are quite powerful and applicable to general highly structured models [Green et al. 2003] such as our spatio-temporal models for areal data.

It is critically important to design Markov chains with good properties such as fast convergence and small autocorrelation between realizations. With that objective in mind, the Markov chain has to be tailored to the specific spatio-temporal model at hand and will depend on how $\mathbf{F}_t, \mathbf{G}_t$ and $\mathbf{W}_t$ depend on ψ. Nevertheless, the Markov chain may be partitioned in two blocks: simulation of ψ and simulation of $(\beta_0, \ldots, \beta_T)$. The simulation of β is model specific and is briefly discussed in the application of Section 5. For the simulation of the latent process, we use the forward information filter backward sampler (FIFBS) that combines the forward filter backward sampler [Carter and Kohn 1994, Frühwirth-Schnatter 1994] with the information filter and thus benefits from the sparsity of $\mathbf{V}_t^{-1}$ and $\mathbf{W}_t^{-1}$ to accelerate computations. For details of FIFBS, see [Vivar and Ferreira 2007].

4. Model selection

4.1. Proposed models

This subsection presents several special cases of our spatio-temporal model (3). Looking at the data, it is clear that this is not a stationary process, the we need non-stationary models. The first candidate is a model that smooth the data (I). Another alternative is a contamination model (II) since some clusters of counties can be detected through the period under study. Other models considered are related to the notorious increasing mean level of many counties through time. They are the second-order temporal trend model and some variants (III - VIII).

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

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

Models III e 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 $\phi_2 = 0$.

Model V: Second-order model with velocity equation including contamination and $\phi_2 = 0$

- $\mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{G}_{1t} \\ \mathbf{0}_S & \mathbf{G}_{2t} \end{pmatrix}$,

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

- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0}_S \\ \mathbf{0}_S & \mathbf{W}_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i\mathbf{M}), i = 1, 2.$

Model VI: Second-order model with level equation including contamination

- $\mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{G}_{1t} \\ \mathbf{0}_S & \mathbf{G}_{2t} \end{pmatrix}$,

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

- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0}_S \\ \mathbf{0}_S & \mathbf{W}_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i\mathbf{M}), i = 1, 2.$

Model VII: Second-order model including contamination on both equations and $\phi_1 = 0$

- $\mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{G}_{1t} \\ \mathbf{0}_S & \mathbf{G}_{2t} \end{pmatrix}$, $\mathbf{G}_{it} = \frac{1}{1+\kappa_i h}\mathbf{H}_i \longrightarrow \{\mathbf{H}_i\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa_i, & k \in N_l, \\ 0, & o.c. \end{cases}, i = 1, 2,$
- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0}_S \\ \mathbf{0}_S & \mathbf{W}_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i \mathbf{M}), i = 1, 2.$

Model VIII: Second-order model with same acceleration for all sites

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

4.2. Predictive density

With the ability to fit more complex models comes the necessity to compare those models. There are many criteria to select models in the literature. For example, criteria based on a predictive distribution include those of [Geisser and Eddy 1979, San Martini and Spezzaferri 1984] and [Gelfand et al. 1992] and the references therein. We prefer the predictive density as our criterion of model selection because it naturally penalizes complex models, differently from other ones, like the deviance information criterion (DIC), [Spiegelhalter et al. 2002] that favors overfitting and tends to select very complex models.

Bayesian model selection is usually performed by comparing the posterior probabilities of the competing models. When the competing models have equal prior probabilities, their posterior probabilities are proportional to the respective predictive densities. These densities will depend on the prior distribution of the parameter vector ψ. In order to overcome this difficulty, we use here a training sample approach [Frühwirth-Schnatter 1995] of the first p time observations; this results in calibrated priors for the parameters of each model. Then, Monte Carlo integration is used to compute the predictive distribution under each model for the remaining $T - p$ time observations.

Suppose that there are Q competing spatio-temporal models $M_1, \ldots, M_Q$. The qth model has observational density $p_q(\mathbf{y}_t|\eta_t(\beta_t))$ and evolution density $p_q(\beta_t|\beta_{t-1}, \psi)$. Note that the definitions of β_t and ψ may be (and in general will be) different under each model, but this distinction is omitted in order to keep the notation simple.

Let $p_q(\beta_{1:t-1}, \psi|\mathbf{D}_{t-1})$ denote the joint posterior distribution of ψ and $\beta_1, \ldots, \beta_{t-1}$ under model q up to time $t-1$. Then, the predictive distribution of $\mathbf{y}_t$ under model q given the information up to time $t-1$ will be

$$
\begin{aligned}
p_q(\mathbf{y}_t|\mathbf{D}_{t-1}) &= \int p_q(\mathbf{y}_t|\eta_t(\beta_t))p_q(\beta_t|\beta_{t-1},\psi)p_q(\beta_{1:t-1},\psi|\mathbf{D}_{t-1})d\beta_{1:t-1}d\beta_t d\psi \\
&= \int p_q(\mathbf{y}_t|\beta_{t-1},\psi)p_q(\beta_{1:t-1},\psi|\mathbf{D}_{t-1})d\beta_{1:t-1}d\psi \qquad (4)
\end{aligned}
$$

since

$$
p_q(\mathbf{y}_t|\beta_{t-1},\psi) = \int p_q(\mathbf{y}_t|\eta_t(\beta_t))p_q(\beta_t|\beta_{t-1},\psi)d\beta_t.
$$

The simulation scheme outlined in Section 3 can be used to simulate a sample $(\beta_{t-1}^{(1)},\psi^{(1)}),\ldots,(\beta_{t-1}^{(L)},\psi^{(L)})$ from the joint posterior distribution $p_q(\beta_{1:t-1},\psi|\mathbf{D}_{t-1})$. Then, a Rao-Blackwellized estimate of the predictive density of $\mathbf{y}_t$ given the information up to time $t-1$ is

$$
\hat{p}_q(\mathbf{y}_t|\mathbf{D}_{t-1}) = \frac{1}{L}\sum_{l=1}^{L} p_q(\mathbf{y}_t|\beta_{t-1}^{(l)},\psi^{(l)}) \qquad (5)
$$

Thus, we adjust an MCMC scheme for each time point t and make the prediction for the subsequent time $t+1$. Using the fact that the joint predictive density of $\mathbf{y}_{t^*},\mathbf{y}_{t^*+1},\ldots,\mathbf{y}_T$ can be written as $p_q(\mathbf{y}_{t^*},\mathbf{y}_{t^*+1},\ldots,\mathbf{y}_T|\mathbf{D}_{t^*-1}) = \prod_{t=t^*+1}^{T} p_q(\mathbf{y}_t|\mathbf{D}_{t-1})$, an estimate of the joint predictive density under model q is

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

where t^* is such that $p_q(\psi|D_{t^*})$ is proper for all $q = 1,\ldots,Q$.

As a result, one selects the model with the highest predictive density. Thus, the selected model will not only be the model with highest posterior probability but will also be the model with the best predictive performance. When one model has posterior probability close to one, that model is clearly the winner. But very often several models will have similar posterior probabilities. In that case, those several models should be reported and predictions should be computed by model averaging [Clyde and George 2004].

5. Application

Within the Bayesian paradigm, the models are complete with the specification of prior distributions for β_0, τ_i, ϕ_i and $\kappa_i, i = 1,2$; the usual assumption of independent priors is used here. The prior for β_0 is a multivariate normal with certain mean vector and diagonal precision matrix with elements close to 0, corresponding to vague information. The prior for τ_i is a gamma distribution $Ga(4,4)$ leading to a gamma full conditional distribution. The prior for ϕ_i proportional to 1 if $0 < \phi_i < 1$ and proportional to ϕ_i^{-5} if $\phi_i \geq 1$. The prior for κ_i is a uniform distribution in the interval $(0,1)$. Usually is not an easy task the estimation of ϕ_i and τ_i, hence their priors are somehow semi-informative.

For each model, the MCMC scheme discussed in Section 3 was used with the simulation of the latent process by FIFBS. Moreover, the updating for τ_i was performed with independent Gibbs steps and the updating for ϕ_i and κ_i was performed with Metropolis steps for $\log\phi_i$ and for κ_i. This MCMC scheme was implemented using Ox [Doornik 2002]. For each model, 90000 iterations were run and the first 10000 iterations

were discarded as burn-in. Then, we saved every 20th iteration yielding a final sample of size 4000 for each parameter.

For model selection, the first ten time points were used as training sample; that is, $t^* = 10$. Table 1 shows the logarithm of the predictive density for the several models and the mean squared error of the one-step ahead prediction (as a simpler comparison method, since it doesn't take account of all uncertainty). From the table, the best model is the contamination model (Model II). In order to understand the difference in performance between the models, Figure 1 shows the one-step-ahead predictive densities for all the models. Model II is the winner in almost all the time points.

Table 1. 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

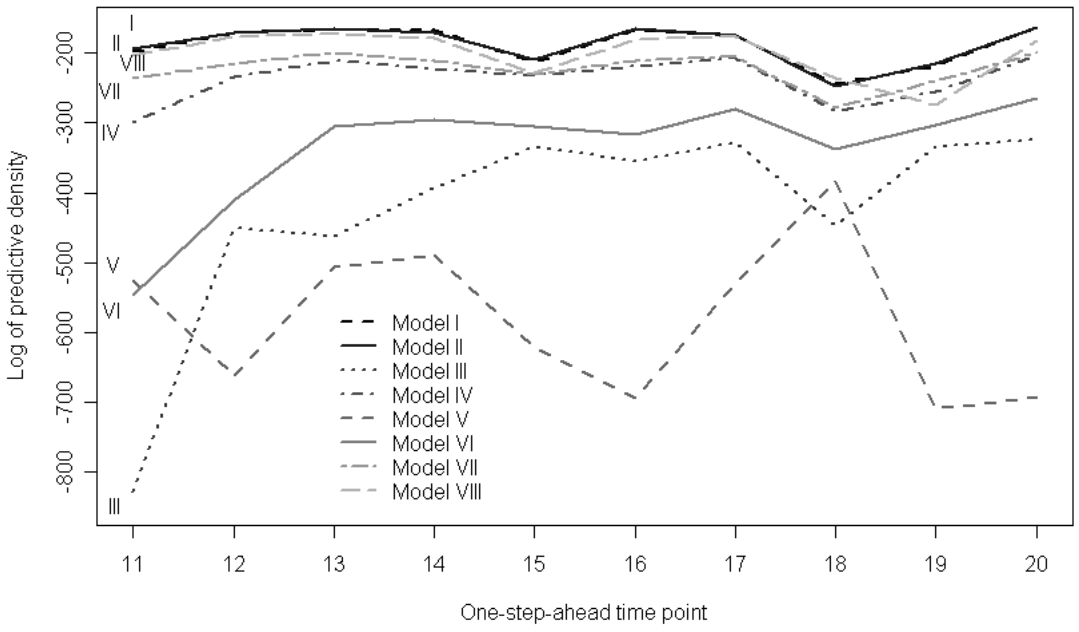

Figure 1. One-step ahead logarithm of the predictive density.

Table 2 shows point estimates of the parameters of the contamination model using all time observations and the respective quantiles in order to elaborate confidence intervals

of 95%. These results show that the innovations have a strong spatial correlation and a moderate precision indicating that the innovations have low magnitude. The value of the contamination index κ indicates very low but (statistically) significant interaction between sites in subsequent times.

Table 2. Posterior results using all data for the best model selected: Contamination model (II).

		Quantiles		
Parameter	Mean (s.d.)	2.5%	50%	97.5%
τ	1.518 (0.441)	0.795	1.459	2.514
ϕ	6.469 (2.376)	3.081	6.108	7.778
κ	0.0003 (0.0002)	0.00002	0.0003	0.0008

Figure 2 shows the posterior means of both the latent level (first row of each panel) and the innovation process (second row of each panel) for the years 1981, 1984, 1987, 1990, 1993 and 1996. The maps on the first rows of each panel represent the posterior mean of the risk per 100 thousand inhabitants in a scale from white (low risk) to black (high risk). In the beginning, the risk level was almost uniform for all counties, including metropolitan area (east-center region). Moreover, throughout the years the violence increases at the center of the State and some northern counties. The maps on the second rows of each panel represent the posterior mean of the innovations in a scale from white (high reduction in the risk) to black (high increase in the risk). These innovation maps are very informative, as they represent estimates of year specific spatially structured effects. For example, after accounting for the contamination effect, in 1993 there was a year-specific increase of the risk level in the center and southern regions and a decrease of the risk level in the northern part of Espírito Santo State.

6. Conclusions

We have presented a method to perform model selection of spatio-temporal models for areal observations in the exponential family. This method is based on predictive densities and in this paper we present a specific general spatio-temporal model for count data. Several different specifications of this model are presented and applied to the annual number of homicides in the State of Espírito Santo in the 1979-1998 period.

The proposed Bayesian analysis using MCMC with embedded FIFBS allows for full account of the uncertainty and the predictive-density-based model selection pointed to the contamination model as the best model among the proposed ones. A possible explanation is that neighbor counties may have similar security policies causing an increasing or decreasing violence process. This behavior was reflected on the maps representing the innovations, showing estimates of spatially structured effects.Further research will include different models for the homicides maybe considering some covariates; and model selection for other kind of data, like binomial data.

References

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 14:2433–2443.

Brown, P. E., Diggle, P. J., Lord, M. E., and Young, P. C. (2001). Space-time calibration of radar-rainfall data. *Applied Statistics*, 50:221–242.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81:541–553.

Clyde, M. and George, E. I. (2004). Model uncertainty. *Statistical Science*, 19:81–94.

Doornik, J. A. (2002). *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press and Oxford, London, 3rd. edition. URL: http://www.nuff.ox.ac.uk/Users/Doornik.

Ferreira, M. A. R. and De Oliveira, V. (2007). Bayesian reference analysis for Gaussian Markov Random Fields. *Journal of Multivariate Analysis*, 98:789–812.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series*, 15(2):183–202.

Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and bayes factors for linear gaussian state-space models. *Journal of the Royal Sttistics Society, Series B*, 57:237–246.

Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, 2nd edition.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74:153–160.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 4*. London: Oxford University Press.

Ghil, M., Cohn, S., Tavantzis, J., Bube, K., and Isaacson, E. (1981). Applications of estimation theory to numerical weather prediction. In Bengtsson, L., Ghil, M., and Källén, E., editors, *Dynamic Meteorology: Data Assimilation Methods*, pages 139–224. New York: Springer-Verlag.

Green, P. J., Hjort, N. L., and Richardson, S. (2003). *Highly Structured Stochastic Systems*. Oxford University Press.

Guttorp, P., Meiring, W., and Sampson, P. D. (1994). A space-time analysis of ground-level ozone data. *Environmetrics*, 5:241–254.

Huang, H. C. and Cressie, N. A. C. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis*, 22:159–175.

Huerta, G., Sansó, B., and Stroud, J. R. (2004). A spatio-temporal model for Mexico City ozone levels. *Journal of the Royal Statistical Society, Ser. C*, 53:231–248.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19:2555–2567.

Knorr-Held, L. and Richardson, S. (2003). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Applied Statistics*, 52:169–183.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag.

San Martini, A. and Spezzaferri, F. (1984). A predictive model selection criterion. *Journal of the Royal Statistical Society, Series B*, 46:296–303.

Schmid, V. and Held, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics*, 60:1034–1042.

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64:583–640.

Vivar, J. C. (2004). Uma nova classe de modelos espaço-temporais para dados de área. Master's thesis, Departamento de Métodos Estatísticos, IM – UFRJ, Rio de Janeiro, Brasil.

Vivar, J. C. and Ferreira, M. A. R. (2007). Spatio-temporal models for gaussian areal data. *(Submitted to Journal of Computational and Graphical Statistics)*.

Waller, L. A., Carlin, B. P., Xia, H., and E., G. A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92:607–617.

West, M. and Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, New York, 2nd. edition.

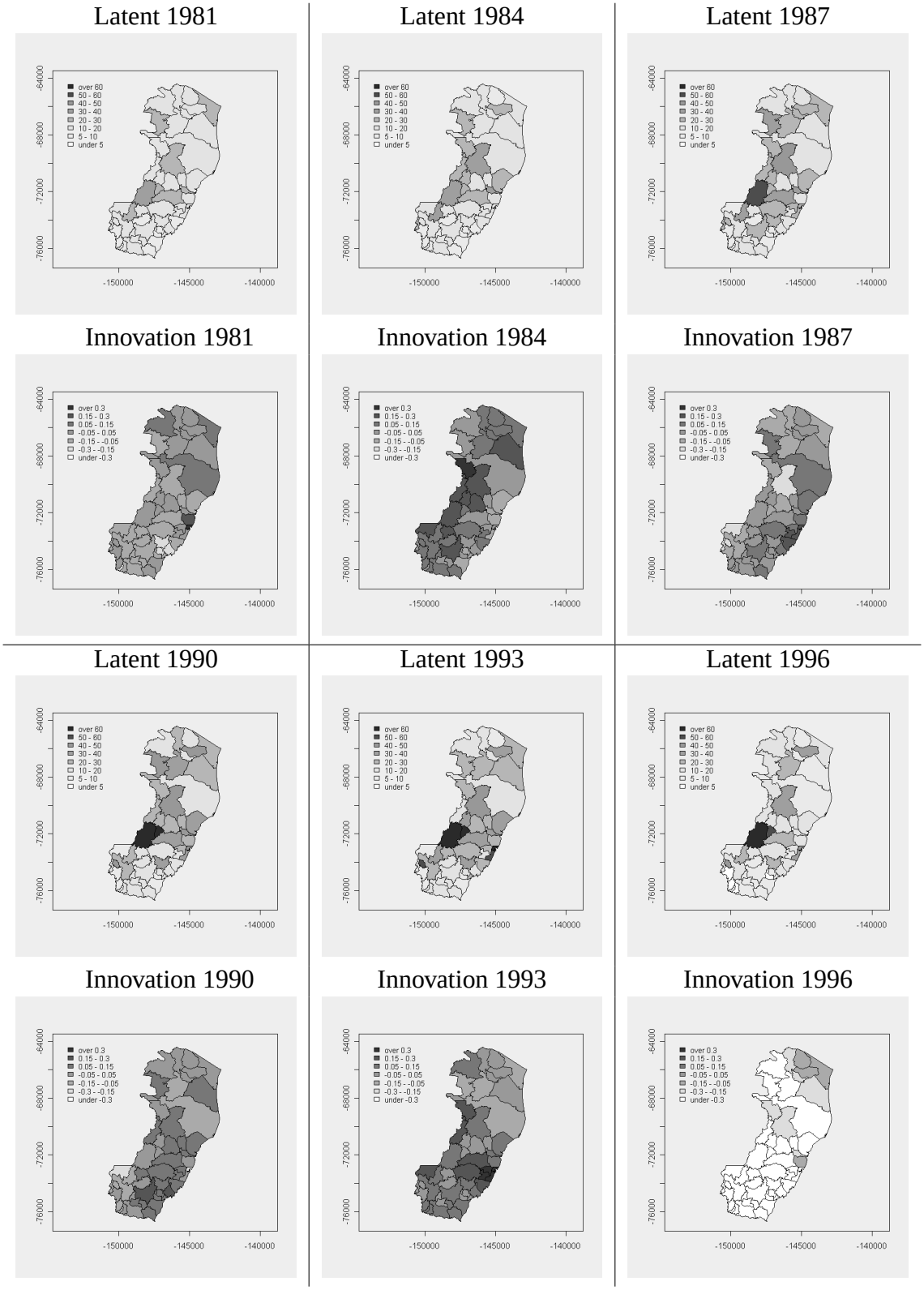

Figure 2. Homicides data - Spatio-temporal contamination model. Posterior means of latent and innovations fields for years 1981, 1984, 1987, 1990, 1993 and 1996.