ENVIRONMETRICS
Environmetrics 2005; **16**: 291–304
Published online 2 February 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/env.704

Spatio-temporal models for mapping the incidence of malaria in Pará

Aline A. Nobre[1,*,†], Alexandra M. Schmidt[1] and Hedibert F. Lopes[2]

[1]*Departamento de Métodos Estatísticos—UFRJ, Brazil*
[2]*Graduate School of Business—University of Chicago, U.S.A.*

SUMMARY

Our main aims in this article are: (i) to model the means by which rainfall affects malaria incidence in the state of Pará, one of Brazil's largest states; and (ii) to check for similarities along the counties in the state. We use state of the art spatial–temporal models which can, we believe, anticipate various kinds of interactions and relations that might be present in the data.

We use the traditional Poisson–normal model where, at any given time, the incidences of malaria for any two counties are conditionally independent and Poisson distributed with log-mean explained by rainfall and random effects terms. Our methodological contribution is in allowing some of the random effects variances to evolve with time according to a dynamic model. Additionally, the change of support problem caused by combining malaria counts (per county) with rainfall (per station) is partially solved by interpolating the whole state through a Gaussian process.

Posterior inference and model comparison are computationally assessed via Markov chain Monte Carlo (MCMC) methods and deviance information criteria (DIC), respectively. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: Bayesian kriging; change of support; conditional autoregressive models; relative risk; spatio-temporal interaction

1. INTRODUCTION

Malaria is the main endemic infectious disease in the Brazilian Amazon region, having about 600 million cases per year worldwide. The disease is transmitted by mosquitoes from the *Anopheles sp* genus, which is developed by accumulated water either on the ground or on rivers and lakes. Pará is the state of Brazil with the highest number of cases—168 688 cases in 2001.

There are natural and social factors affecting the dynamics of the transmission of infection. Temperature and rainfall are considered major natural risk factors that affect the cycle's life and the breeding of the mosquitoes. Under certain climatological conditions, mainly humidity and temperature, the eggs take an average of 10 days to hatch. On the one hand, reasonable amounts of rainfall

*Correspondence to: A. A. Nobre, Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, Rio de Janeiro, RJ, Brazil. CEP 21945-970.
†E-mail: anobre@dme.ufrj.br.

Received 11 November 2003
Accepted 18 August 2004

Copyright © 2005 John Wiley & Sons, Ltd.

create additional breeding sites for mosquitoes, increasing their population. However, huge amounts of rain clean the ground and kill the eggs. On the other hand, if there is drought, the mosquitoes disappear from the ground, the insects' breeding becomes more difficult and, as a consequence, the index of disease transmission decreases. Despite the fact that mosquitoes fertilize only once in their lifespan, they continue to produce eggs on a regular basis every 2 to 4 days (Dias, 2003). The exodus of population from outside the endemic region represents a major social risk factor, that puts at risk the susceptible locals, who are biologically more vulnerable to infections. Another important social factor is the lack of preventive public health policies.

Our main aims are to look for possible relationships between rainfall and malaria and to learn about spatial and temporal similarities amongst counties in the state. However, in order to use rainfall to explain malaria we tackle the existing change of support problem by interpolating rainfall along the state through a Gaussian process (Cressie, 1993).

This paper is organized as follows. Section 2 reviews current methodology for disease mapping, and this is specialized for the problem at hand. Section 3 presents the inference procedure based on the Bayesian paradigm. Results are presented in Section 4, where we also discuss the change of support problem and issues about model adequacy. Finally, Section 5 summarizes our findings and states the directions of our future research agenda.

2. SPACE AND SPACE–TIME MODELS FOR DISEASE MAPPING

Traditionally, disease mapping has been used to describe the geographic distribution of diseases and to identify areas of high risk, both useful, for instance, in resource allocation policies. Moreover, these methods can be used in the investigation of epidemiologic and environmental hypotheses. As a result, a growth in public and scientific interest is apparent.

We start by doing a brief review on spatial models for disease mapping. Then we discuss some models for spatio-temporal data and present our proposal.

2.1. Spatial models

We initially assume that the region of interest has been divided into n contiguous subregions, and that y_i represents the number of cases of the disease for each area $i = 1, \ldots, n$. In the classical literature, it is assumed that y_i follows a Poisson distribution with mean $e_i r_i$, where e_i is the expected counting of the disease and r_i represents the relative risk, i.e.

$$y_i \sim \text{Poisson}(e_i r_i) \tag{1}$$

The expected counts are a known quantity based on common risk factors, and play an important role in standardizing the information from areas of different characteristics. Therefore, the unknown quantities of interest are the relative risks, r_i.

It is already well documented that disease maps based on the maximum likelihood estimate (MLE) of r_i, denoted here by $\hat{r}_i = y_i/e_i$, the crude rates, are misleading when, for instance, the disease is rare or the areas are small or both (Mollié, 1996). Besides, ML estimates do not accommodate possible spatial dependence. Another aspect to consider is the homogeneity within each area, where individual risks are expected to be equal within the area. Therefore, it is natural to include an extra Poisson variation in the model so that the estimated relative risks take into account this effect.

Copyright © 2005 John Wiley & Sons, Ltd.

The literature on modeling relative risks is very rich. A classical reference is Besag *et al.* (1991), which assumes that, for area i, the log relative risks are modeled through

$$\log(r_i) = \alpha + \beta x_i + b_i + u_i \tag{2}$$

where α is a common intercept for the entire region, x_i is a vector of covariates and u_i and b_i are random effect terms, such that $u_i \sim \mathrm{N}(0, \sigma_u^2)$, $i = 1, \ldots, n$, representing some unstructured noise term, and

$$b_i \mid b_j, j \neq i \sim \mathrm{N}\left(\frac{\sum_{j\in\delta i} W_{ij} b_j}{\sum_{j\in\delta i} W_{ij}}, \frac{\sigma_b^2}{\sum_{j\in\delta i} W_{ij}}\right), \quad i = 1, \ldots, n \tag{3}$$

where δi is the set containing areas adjacent to i, and W_{ij} is the weight that neighboring area j has on i. The simplest, and most commonly used, neighboring structure is based on adjacent areas, where neighbor regions share a common boundary. This will be the case considered throughout this paper. This implies that $W_{ij} = 1$ if i and j share boundaries ($i \sim j$) and 0 otherwise. The prior distribution in (3) is known in the literature as a conditional autoregressive (CAR) prior, and will be denoted here by CAR(σ_b^2). Additional references using CAR prior distributions for disease mapping are Bernardinelli *et al.* (1995), Best *et al.* (1999), Kelsal and Wakefield (2002) and references therein. Best *et al.* (1999) investigate the choice of neighborhood structure and the specification of hyperprior distributions. On the other hand, Kelsal and Wakefield (2002) model the relative risks through a Gaussian random field where the correlation structure is derived from an underlying continuous risk surface. They claim that this leads to more realistic correlation structures between neighboring areas.

2.2. *Spatio-temporal models*

Now, besides having a region divided into n contiguous subregions, we assume that observations arise at successive, discrete instants of time, i.e. we now observe Y_{it} for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, and assume that

$$\begin{aligned} y_{it} &\sim \mathrm{Poisson}(e_{it} r_{it}) \\ \log(r_{it}) &= \alpha_t + \beta_t x_{it} + b_{it} \end{aligned} \tag{4}$$

Most of the studies on disease mapping do not deal with the spatial and temporal components jointly. The presence of longitudinal information and spatially referenced data for disease incidence encourages the study and development of new models to deal with this important issue. Waller *et al.* (1997) consider a model like the one in Besag *et al.* (1991), Equation (2), where both error terms are time-dependent. They consider the random effect as the sum of heterogeneity and spatial patterns over time. However, in their approach, the coefficient of the covariate and the variances of the error terms are fixed for all t. They analyzed lung cancer mortality data in 88 counties in the state of Ohio over a period of 21 years. Knorr-Held and Besag (1998) adopt the model in (4), with α_t following a random walk prior and $b_{it} = b_i$ for all $t = 1, \ldots, T$. They also investigated the same lung cancer data as Waller *et al.* (1997), and mentioned that they use a fixed spatial random effect because they believe that all relevant information on temporal covariates was available. Assunção *et al.* (2001) adopt an area-specific second-degree polynomial trend model for the diffusion and prediction of Leishmaniasis

Copyright © 2005 John Wiley & Sons, Ltd.

through time. The spatio-temporal interaction is modeled through the use of a CAR prior distribution on the parameters of these polynomials. Knorr-Held and Richardson (2003) propose a hierarchical model for space–time surveillance data on meningococcal disease incidence which is able to capture hyperendemic periods through the use of temporal indicator variables.

2.3. *Our spatio-temporal model*

We propose a spatio-temporal generalization of the priors in model (4) that, among other things, allows β_t and b_{it} to evolve smoothly over time:[§]

$$\begin{aligned} \log(r_{it}) &= \beta_t x_{it} + b_{it} \\ \beta_t &= \beta_{t-1} + w_t, \quad w_t \sim \mathrm{N}(0, \tau_\beta^2) \end{aligned}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. As can be easily recognized, this structure resembles a dynamic generalized linear model (West *et al.*, 1985). We model the random effects terms, b_{it}, as either independent and identically normally distributed with mean zero and variance σ_t^2, $b_{it} \sim \mathrm{N}(0, \sigma_t^2)$, or as a $\mathrm{CAR}(\sigma_t^2)$. Additionally, σ_t^2s can be either independent and identically distributed inverse gamma with (fixed) parameters a and c, $\sigma_t^2 \sim \mathrm{IG}(a, c)$, for $t = 1, \ldots, T$, or $\log(\sigma_t^2) = \log(\sigma_{t-1}^2) + \varepsilon_t$, with $\varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$. Therefore, we entertain the following four prior distributions:

$$\mathrm{M}_1\text{: } b_{it} \sim \mathrm{N}(0, \sigma_t^2) \quad \text{and} \quad \sigma_t^2 \sim \mathrm{IG}(a, c)$$

$$\mathrm{M}_2\text{: } \mathrm{CAR}(\sigma_t^2) \quad \text{and} \quad \sigma_t^2 \sim \mathrm{IG}(a, c)$$

$$\mathrm{M}_3\text{: } b_{it} \sim \mathrm{N}(0, \sigma_t^2) \quad \text{and} \quad \log(\sigma_t^2) = \log(\sigma_{t-1}^2) + \varepsilon_t$$

$$\mathrm{M}_4\text{: } \mathrm{CAR}(\sigma_t^2) \quad \text{and} \quad \log(\sigma_t^2) = \log(\sigma_{t-1}^2) + \varepsilon_t$$

It is clear that models M_1 and M_3 are very similar. The random effects are allowed to vary independently a priori with variances that also vary independently in M_1 but evolve smoothly through time in M_3. Likewise, models M_2 and M_4 are also similar, whereas now, the random effects are assumed to follow a spatially structured prior and the variability of these effects is allowed to change with time.

Unlike previous work, our modeling structure allows the data to speak up whenever any structure in the evolution of the variability of the effects is apparent.

3. POSTERIOR INFERENCE

It is obvious, in our context, that y_{it} and x_{it} are misaligned, since x_{it} would represent rainfall in county i, while the only available information is rainfall for several points (monitoring stations) in the region

[§]Due to the absence of confounding factors, here we compute the e_{it}s simply as $e_{it} = p_{it} p_t^*$, where $p_t^* = \sum_i y_{it} / \sum_i p_{it}$ and p_{it} is the population in city i in the month t.

Copyright © 2005 John Wiley & Sons, Ltd.

under study (the state of Pará). Nonetheless, we assume, in this section, that the covariate x_{it} is known for all areas i and times t. We discuss this problem further in Section 4. That being said, let $\boldsymbol{\theta}$ be the parameter vector containing all the unknown quantities of interest. More precisely, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\sigma}^2, \tau_\beta^2, \sigma_\varepsilon^2)$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)$, $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_T)$, $\boldsymbol{b}_t = (b_{1t}, \ldots, b_{nt})'$ for $t = 1, \ldots, T$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_T^2)$.

For illustration purposes consider model M_4 as defined above. From Bayes' theorem, the posterior distribution of $\boldsymbol{\theta}$ is proportional to the prior times the likelihood function; that is,

$$
\begin{aligned}
p(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{x}) \propto & \prod_{t=1}^{T} \prod_{i=1}^{n} \exp(\beta_t x_{it} + b_{it})^{y_{it}} \exp\{-e_{it} \exp(\beta_t x_{it} + b_{it})\} \\
& \times \prod_{t=1}^{T} \left[p\left(\beta_t \mid \beta_{t-1}, \tau_\beta^2\right) p\left(\boldsymbol{b}_t \mid \sigma_t^2\right) p\left(\sigma_t^2 \mid \sigma_{t-1}^2, \sigma_\varepsilon^2\right) \right] \\
& \times p(\beta_0) p\left(\tau_\beta^2\right) p\left(\sigma_\varepsilon^2\right)
\end{aligned}
$$

where $\boldsymbol{y} = \{y_{it} : i = 1, \ldots, n; t = 1, \ldots, T\}$ and $\boldsymbol{x} = \{x_{it} : i = 1, \ldots, n; t = 1, \ldots, T\}$. We are now left to assign prior distributions to the parameters of the model at time $t = 0$. For β_0 we assign a zero mean normal prior with large (fixed) variance. For the variance parameters we use inverse gamma prior distributions with parameters fixed at some suitable values.

Clearly, the posterior distribution is analytically intractable. However, several Markov chain Monte Carlo algorithms have already been proposed to sample from $p(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{x})$. More specifically, we run a Gibbs sampling (Gelfand and Smith, 1990) algorithm based on the posterior full conditional distributions of each of the parameters. Except for the scale parameters, $\tau_\alpha^2, \tau_\beta^2, \boldsymbol{\sigma}^2$ and σ_ε^2, whose full conditional distributions are inverse gamma distributions, all the others have unknown posterior full conditional distributions. However, they are log concave and one could use, for example, the adaptive rejection method (Gilks and Wild, 1992) to obtain samples from them. We use the software WinBugs version 1.4[¶] (Spiegelhalter *et al.*, 2002) to perform all posterior inference for models M_1–M_4. An example of the code used to fit model M_1 is shown in the Appendix.

4. CASE STUDY: MAPPING MALARIA IN PARÁ

The available information on malaria comprises monthly recorded counts from January 1996 until December 1998 for 69 counties in the state of Pará, with 34 considered missing, but included in the model in order to facilitate neighboring allocation. We refer to Nobre (2003) for further information about how we ended up using only 69 of Pará's 140 counties. For the same number of months/years, rainfall measurements (in millimeters) were collected in 78 monitoring stations throughout the state (see Figure 1). The county of Anajás is highlighted in the map by its name as it is the county with the highest number of cases of malaria along the years.

Our analysis has essentially two steps: (i) x_{it} is estimated for each county i and time t by using a Gaussian spatial interpolator; and (ii) conditioned on x_{it}, the four models described in the end of Section 2 are entertained.

[¶]WinBugs is freely available from the web at http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.

Copyright © 2005 John Wiley & Sons, Ltd.

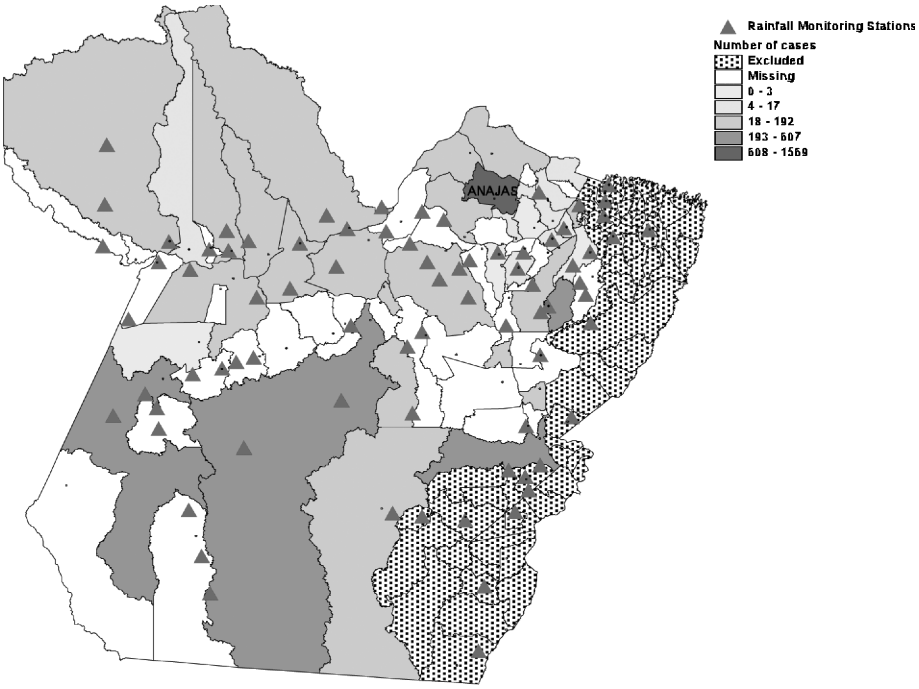

Figure 1. State of Pará with the observed number of cases of Malaria in January of 1997, the region under study and the rainfall monitoring stations

4.1. Step 1: modeling the rainfall

As previously outlined we have a change of support problem, as rainfall is observed at fixed locations s (point referenced data), and the counts of malaria are observed at each county (areal data). Following (4), we therefore need to obtain a predicted amount of rainfall, x_{it}, for each area (block) of the region under study.

Gelfand *et al.* (2001) propose an approach for prediction from points to points, points to blocks, blocks to points and blocks to blocks. They give an example of their approach for a static spatial case using a dataset of point-level ozone measurements in the Atlanta metropolitan area. Zhu *et al.* (2003) go further and investigate the relationship between ozone and pediatric asthma ER visits. They also have the change of support problem which is dealt with by applying the method described in Gelfand *et al.* (2001). They model ozone and asthma data jointly, obtaining more realistic estimates as the uncertainty in estimating the block averages of ozone at each zip code is clearly considered. They model the log-ozone by a stationary spatial Gaussian process and also extend for spatio-temporal modeling.

We consider here the modeling of the amount of rainfall observed at each of the 78 sites from October 1995 to October 1998 (see Figure 1). Let $\{X(s) : s \in G\}$ with $G \in \mathbb{R}^p$ be a spatial random field. Here we assume that $p = 2$ such that s describes the latitude and longitude of a point within the region of interest. Considering that each county is a block and following Gelfand *et al.*

Copyright © 2005 John Wiley & Sons, Ltd.

(2001), for block data we assume that the observations arise as block averages; i.e. for a block $B \subset G$,

$$X(B) = |B|^{-1} \int_B X(s)\,\mathrm{d}s \tag{5}$$

where $|B|$ represents the area of B. Let $(X(s_1), \ldots, X(s_q))'$ represent the amount of rainfall observed at q fixed sites. According to the model in (1), for each time t we need to predict $[X_t(B_1), \ldots, X_t(B_n)] \equiv [X_{1t}, \ldots, X_{nt}]$, the amount of rainfall for each of the n counties considered in this study. This means we need to solve the stochastic integral just described.

Following Stidd (1973) and Rodriguez-Iturbe and Mejia (1974), we consider that the cubic root of rainfall follows a Gaussian process with mean given by $\alpha_x + \beta_1 \text{latitude}(s) + \beta_2 \text{longitude}(s)$ and covariance matrix with elements given by the product of a common variance, σ_x^2, by an exponential correlation function, i.e. $\text{cov}(x(s), x(s')) = \sigma_x^2 \exp(-\phi||s - s'||)$, with $||.||$ denoting Euclidean distance. Under the Bayesian framework the model specification is complete after assigning the prior distributions for the parameters $\alpha_x, \beta_1, \beta_2, \sigma_x^2$ and ϕ. We assume that they are all independent a priori. For the intercept and the coefficients of latitude and longitude we use a zero mean, normal prior with some large fixed variance. An inverse gamma prior was assigned for σ_x^2 with parameters $a = 0.004$ and $b = 0.02$. Finally, a gamma prior was assigned to ϕ with mean $= 0.5$ and variance $\simeq 0.1$. Posterior inference is facilitated through MCMC algorithms and, again, we use WinBugs to perform all needed computations.

The predictions for each block (county) B_i were obtained by approximating the integral in (5) by a sum. This was achieved by overlaying a 231-point grid over the state of Pará. The interpolation for all these *unmonitored* points results from the properties of the multivariate normal distribution (Gelfand *et al.*, 2001). The predicted amount of rainfall for each county i was estimated as the mean of the posterior predictive distribution of the process at points of the grid that have fallen within county i. This procedure was independently repeated at each time point, $t = 1, \ldots, T$.

4.2. Step 2: Model selection/comparison

Considering the dynamics of malaria described in Section 1, an important issue is whether rainfall has an instantaneous or delayed impact on the incidence of malaria. Nobre (2003) has run the generalized spatial linear model, (1) and (2), for several lags of rainfall and found out that a three month lag fits the data quite significantly. In the great majority of counties the rainfall effect, measured through β, was statistically relevant.

As a preliminary analysis, Nobre (2003) used models M_1 and M_2 for each time $t = 1, \ldots, 36$, independently. In her analysis, model M_1 always exhibited a better fit in terms of the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002).

Due to the large number of parameters in the models here considered, we used an empirical Bayes type of approach (O'Hagan, 1994) when dealing with the missing data at those 34 counties that had no data on malaria. These data were input into the spatio-temporal model fitting in the following way. Let y_{jt}^m, $j = 1, \ldots, 34$, be the missing information for each one of those 34 counties at time t. In these separated analyses mentioned above, the missing data were considered as another parameter of the model and a posterior distribution then obtained. Therefore, for $t = 1, \ldots, 36$, we substituted in our data set all the (34×36) missing data by the posterior median of each of their respective distributions. This procedure overlooks model uncertainty and may result in over-optimistically small posterior variances. A fully Bayesian algorithm is under investigation.

Copyright © 2005 John Wiley & Sons, Ltd.

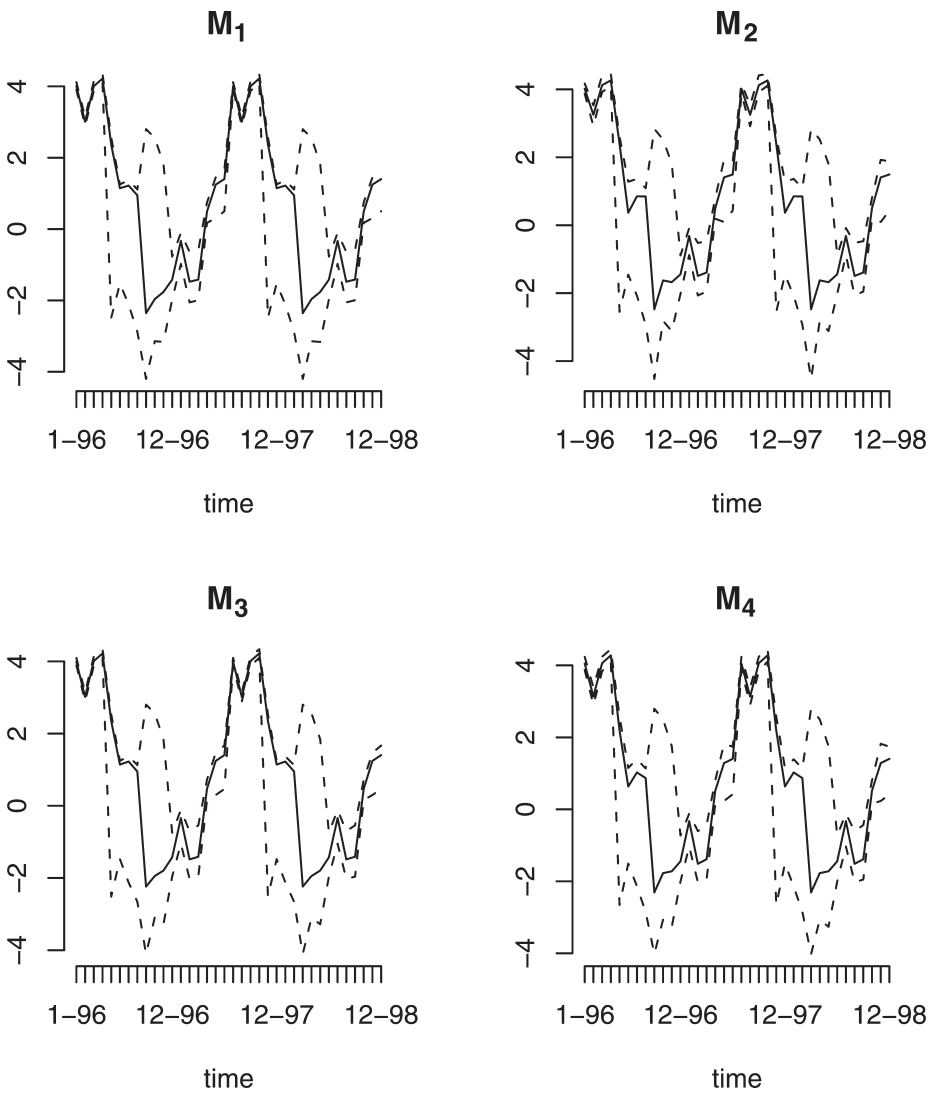

Figure 2. Posterior median (solid line) and 95% credible intervals (dashed lines) of the random effects b_{it} for the county of Anajás, under models M_1–M_4

For each model proposed, we considered a burn-in period of 10 000 iterations and retained every 20th set of parameter values to obtain approximately posterior samples of size 1000. The prior distributions for $1/\tau_\beta$, $1/\sigma_\varepsilon^2$ and $1/\sigma_t^2$ for $t = 1, \ldots, 36$ are gamma distributions with mean and variance both equal to 1. Figure 2 presents the evolution through time of the posterior median of the random effects for the county of Anajás, the one with the highest counts of Malaria over many years. We clearly observe a seasonal pattern of random effects under the four models. Also they tend to present a very similar behavior, showing a smaller adjustment to the log-relative risks during the months of the summer and higher adjustments in the winter months. Figure 3 shows the posterior

Copyright © 2005 John Wiley & Sons, Ltd.

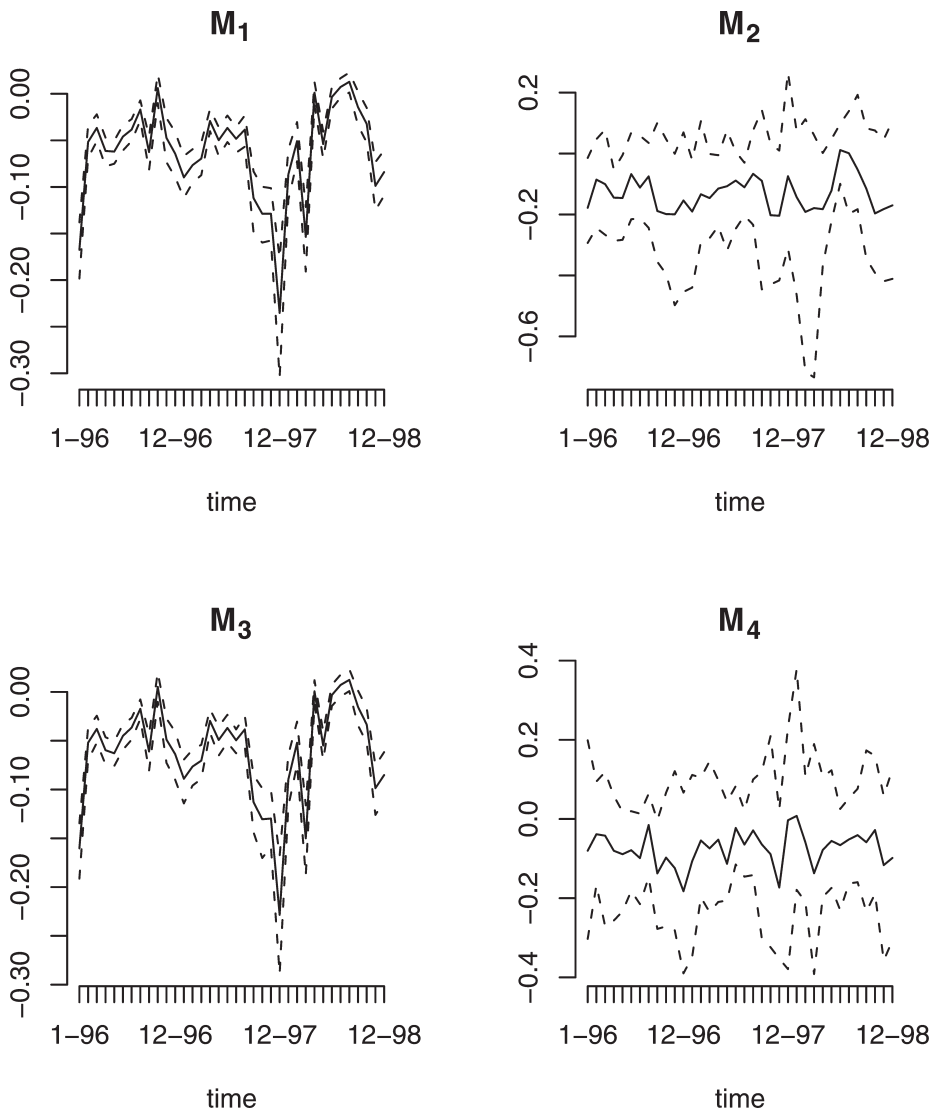

Figure 3. Posterior median (solid line) and 95% credible interval (dashed line) for the rainfall coefficient β_t under models M_1–M_4

median and associated 95% credibility interval for the rainfall coefficient estimated under each of the four models. It is noticeable that rainfall is significant for models M_1 and M_3 and not so significant for models M_2 and M_4; also, for all the models it has a negative effect on the log-relative risk. Still for models M_1 and M_3 we see that at around December 1997 the coefficient drops quite abruptally, and this is probably due to the fact that the amount of rainfall predicted in October 1997 is the highest, showing perhaps that too much rain tends to 'wash' the ground and kill the eggs, reducing the effect of rainfall on $\log(r_{it})$. In Figure 4 we present the evolution of the posterior mean of the random effects'

Copyright © 2005 John Wiley & Sons, Ltd.

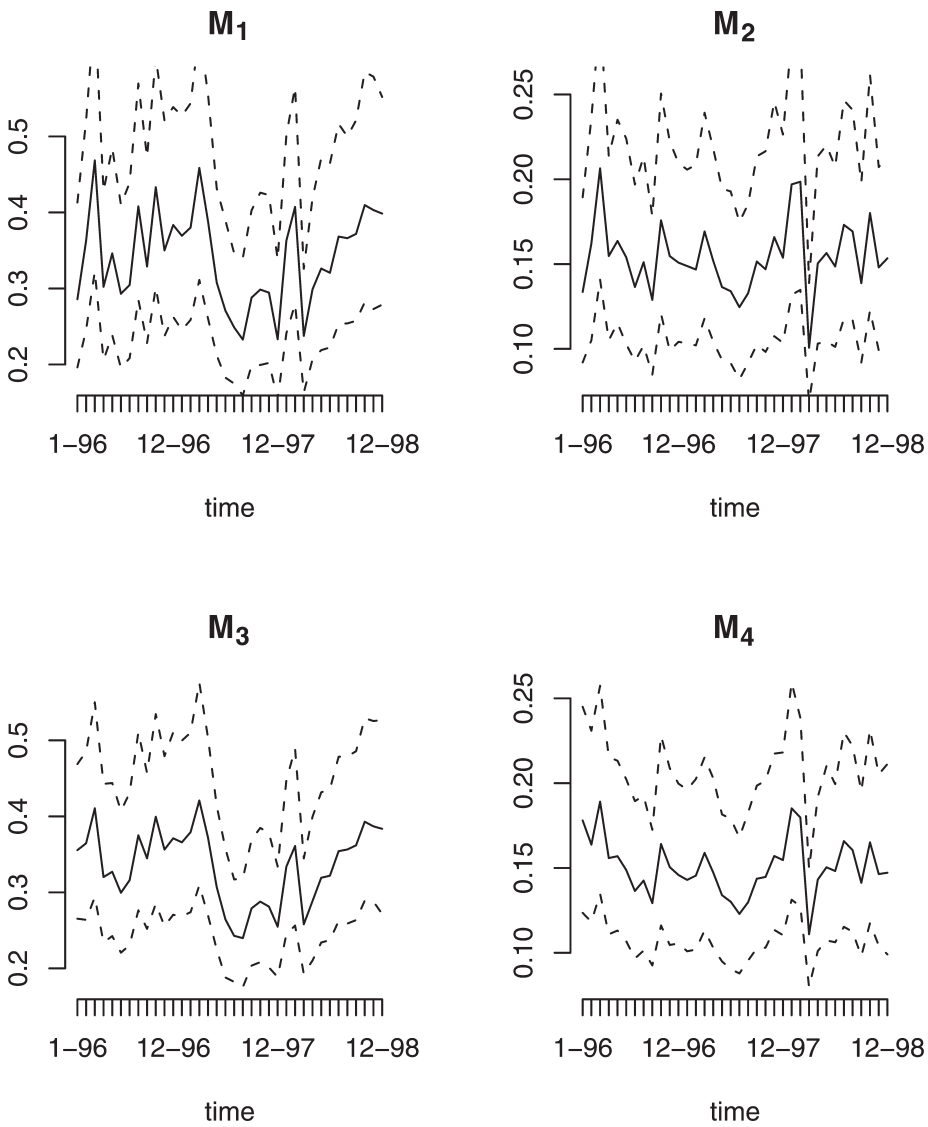

Figure 4. Posterior median (solid line) and associated 95% credible interval (dashed line) for the variance of the random effects under models M_1–M_4

variances and their associated 95% credible intervals. Although models M_1 and M_3 use different prior structures for these variances, we observe that their posterior means tend to behave quite similarly through time, and both differently from models M_2 and M_4. It seems that they are influenced by the prior structure of the random effects, with that of the CAR prior presenting much smaller values than the unstructured ones. In Figure 5 we see the posterior median of the log-relative risks for each of the counties during 1997. We clearly observe a smooth change of the log relative risks throughout the months, confirming the need of a spatio-temporal model. The county of Anajás always presents a high

Copyright © 2005 John Wiley & Sons, Ltd.

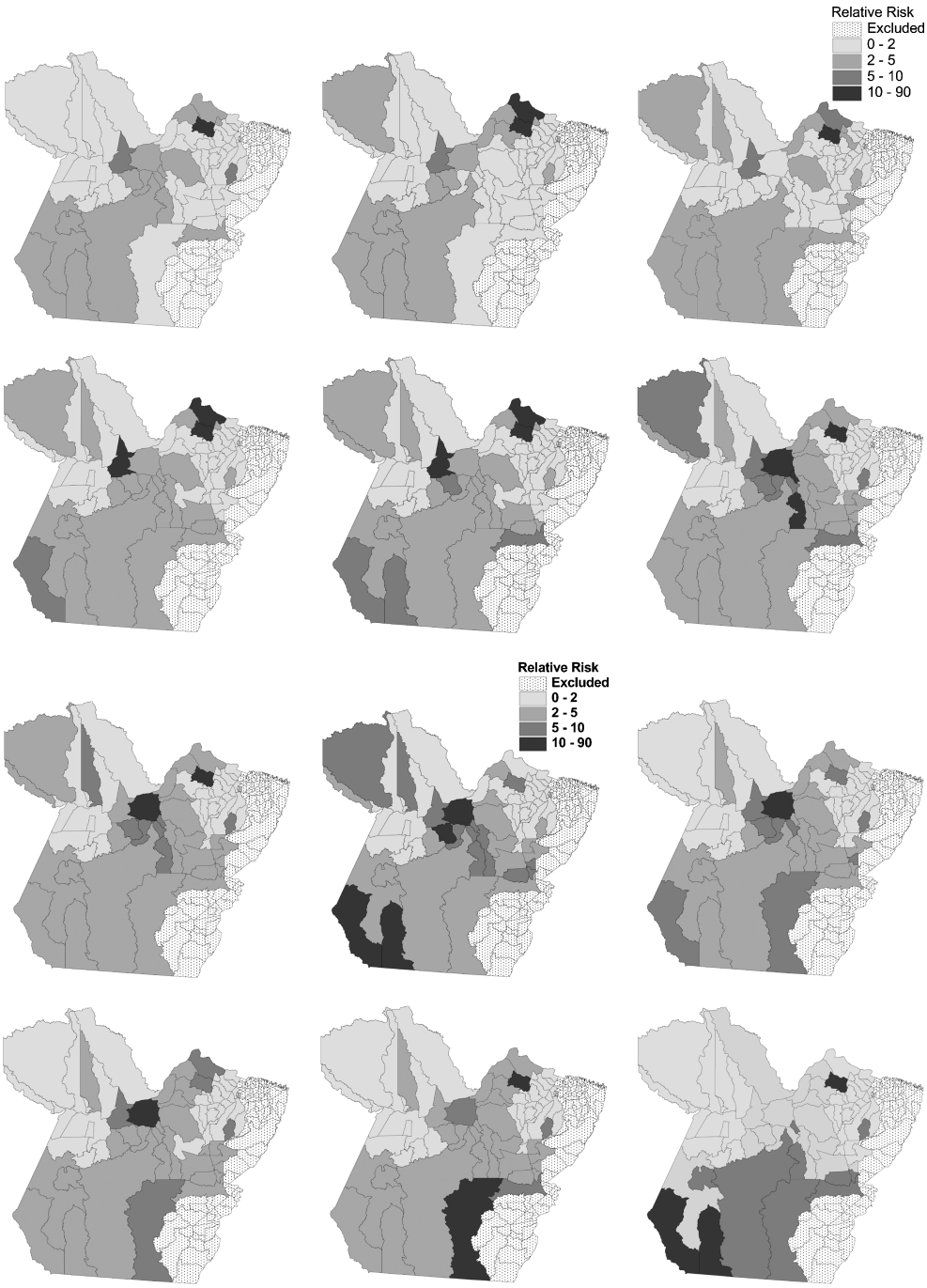

Figure 5. Evolution of the posterior median of the log-relative risks during the year of 1997 under model M_3

Copyright © 2005 John Wiley & Sons, Ltd.

Table 1. Deviance summaries computed for each of the models considered, where $\overline{D}$ is posterior mean deviance, p_D is effective number of parameters and DIC is deviance information criterion

Model	$\overline{D}$	p_D	DIC
M_1	15 569	2374	17 943
M_2	15 650	2319	17 969
M_3	15 574	2369	17 933
M_4	15 646	2319	17 965

risk, maybe due to the fact that it is located in an area of many rivers and close to the Marajó island. We also notice that the north part of the region tends to present a smaller value of the risks in the spring, (September and October). On the other hand, the south part of the region shows an increase of the relative risks in these months.

In order to compare models M_1–M_4 we made use of the DIC suggested by Spiegelhalter *et al.* (2002) which is based on the deviance posterior distribution. Table 1 presents the results for these models. According to the DIC, the best model, amongst those considered, is M_3, the one with unstructured random effects and random walk prior for the log of their variances.

5. CONCLUSIONS AND FUTURE WORK

We have proposed here a spatio-temporal model to analyze the cases of malaria in the state of Pará, Brazil. In our approach we clearly allow for a prior spatio-temporal interaction amongst the random effects. However, when we observe the posterior mean of the random effects and of their hyperparameters, the behavior of the posterior did not seem to differ much. Recall that they are counting for both natural and social effects which might have entered in the model if the information was available. Implementation was easily attained through the use of WinBugs, which is an important issue as these models might be used by epidemiologists in Brazil.

An important issue is to specify with which lag the rainfall should enter the model. Maybe we should even consider more than one lag, but how many and which? This is a difficult question to answer. In the dynamic linear models (DLM) literature (West and Harrison, 1997), if it is believed that a covariate has a persistent effect over the response, it is common practice to consider a transfer function model (West *et al.*, 1985). Alves *et al.* (2004) propose such a model to study the effect of dioxide carbon on the number of children (under 5 years old) with asthma in the city of São Paulo, Brazil. As we have data available for each county i and time t, we can consider fitting their model, but we would be adding a complexity into it as we would again consider a random effect for each area and time. This approach can be challenging as we may not have sufficient information in the time scale to identify all the parameters. Also there might be some unidentifiability between the transfer function component and the random effects. This is the subject of our current research on this project.

ACKNOWLEDGEMENTS

This work was developed as part of A. Nobre's M.Sc. dissertation under the supervision of H. F. Lopes and A. M. Schmidt. Her work was supported by CAPES and CNPq. The authors are thankful to Dr. Ulisses Confalonieri at the Oswaldo Cruz Foundation, PMAGS-ENSP/FIOCRUZ, for making the data set available and for insightful discussions. We also thank Bruno Sansó for useful comments and suggestions on earlier stages of this work.

Copyright © 2005 John Wiley & Sons, Ltd.

REFERENCES

Alves MB, Gamerman D, Ferreira MA. 2004. Mortalidade infantil e qualidade do ar: um estudo dos efeitos defasados da poluição atmosférica sobre doenças respiratórias em São Paulo. Technical Report. Instituto de Matemática, Universidade Federal do Rio de Janeiro.

Assunção RM, Reis IA, Oliveira CDL. 2001. Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine* **20**: 2319–2335.

Bernardinelli L, Clayton D, Montomoli C. 1995. Bayesian estimates of disease maps: how important are priors? *Statistics in Medicine* **14**: 2411–2431.

Besag J, York J, Mollié A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**: 1–59.

Best NG, Arnold RA, Thomas A, Waller LA, Conlon EM. 1999. Bayesian models for spatially correlated disease and exposure data. In *Bayesian Statistics 6*, Bernardo JM, Dawid JO, Berger AP, Smith AFM (eds). Oxford University Press: Oxford, UK; 131–156.

Cressie NAC. 1993. *Statistical for Spatial Data*. New York: Wiley.

Dias RC. 2003. Uma contribuição ao estudo da malária no estado de Roraima e sua associação com a precipitação pluviométrica no período de 1985 a 1996. PhD thesis. Escola Nacional de Saúde Pública, ENSP/FIOCRUZ, Brasil (In Portuguese).

Gelfand AE, Smith AFM. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409.

Gelfand AE, Zhu L, Carlin BP. 2001. On the change of support problem for spatio-temporal data. *Biostatistics* **2**: 31–45.

Gilks WR, Wild P. 1992. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**(2): 337–348.

Kelsal J, Wakefield J. 2002. Modeling spatial variation in disease risk: a geostatistical approach. *Journal of the American Statistical Association* **97**: 692–701.

Knorr-Held L, Besag J. 1998. Modelling risk from a disease in time and space. *Statistics in Medicine* **17**: 2045–2060.

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

Mollié A. 1996. Bayesian mapping of disease. In *Markov Chain Monte Carlo in Practice*, Gilks WR, Richardson S, Spiegelhalter DJ (eds). Chapman & Hall: London, UK; 359–379.

Nobre AA. 2003. A relação entre a malária e a chuva no estado do Pará: Uma análise espaço-temporal. Master's dissertation. Universidade Federal do Rio de Janeiro, Brasil (In Portuguese).

O'Hagan A. 1994. *Kendall's Advanced Theory of Statistics, Bayesian Inference*, Vol. 2B. John Wiley & Sons, Inc.

Rodriguez-Iturbe I, Mejia J. 1974. On the transformation of point rainfall to areal rainfall. *Water Resources Research* **10**: 729–735.

Spiegelhalter DJ, Best NG, Carlin BP, Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* **64**: 583–639.

Spiegelhalter DJ, Thomas A, Best NG. 2002. WinBUGS version 1.4 user manual. Research Report. Cambridge: Medical Research Council Biostatistics.

Stidd CK. 1973. Estimating the precipitation climate. *Water Resources Research* **9**: 1235–1241.

Waller LA, Carlin BP, Xia H, Gelfand AE. 1997. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**: 607–617.

West M, Harrison J. 1997. *Bayesian Forecasting and Dynamic Models*. New York: Springer—Second Edition.

West M, Harrison J, Migon H. 1985. Dynamic generalised linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association* **80**: 73–97.

Zhu L, Carlin BP, Gelfand AE. 2003. Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics* **14**: 537–557.

Copyright © 2005 John Wiley & Sons, Ltd.

APPENDIX A: WINBUGS CODE FOR MODEL M_1

```
model{
  for (i in 1 : 68) {
    b1[i] ~ dnorm(0, tau.b[1])
    ⋮
    b36[i] ~ dnorm(0, tau.b[36])
                  }
    b1[69] <- -sum(b1[1:68])
    ⋮
    b36[69] <- -sum(b36[1:68])
  for (i in 1 : 69) {
    for (t in 1:36){
      Y[i,t] ~ dpois(mu[i,t])
                  }
      log(mu[i,1]) <- log(E[i,1]) + beta[1]*chuva[i,1] + b1[i]
      ⋮
      log(mu[i,36]) <- log(E[i,36]) + beta[36]*chuva[i,36] + b36[i]
      RR[i,1] <- exp(beta[1]*chuva[i,1] + b1[i])
      ⋮
      RR[i,36] <- exp(beta[36]*chuva[i,36] + b36[i])
                    }
  beta[1] ~ dnorm(0,tau.beta)
  for (t in 2:36) {
    beta[t]~ dnorm(beta[t-1],tau.beta)
                }
  for (t in 1:36) {
    tau.b[t] ~ dgamma (1, 1)
                }
  tau.beta ~ dgamma (1, 1)
    }
```

Copyright © 2005 John Wiley & Sons, Ltd.