Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Small area population prediction via hierarchical models

by Debora F. Souza, Fernando A.S. Moura and Helio S. Migon

December 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2009
Vol. 35, No. 2, pp. 203-214
Statistics Canada, Catalogue No. 12-001-X

Small area population prediction via hierarchical models

Debora F. Souza, Fernando A.S. Moura and Helio S. Migon [1]

Abstract

This paper proposes an approach for small area prediction based on data obtained from periodic surveys and censuses. We apply our approach to obtain population predictions for the municipalities not sampled in the Brazilian annual Household Survey (PNAD), as well as to increase the precision of the design-based estimates obtained for the sampled municipalities. In addition to the data provided by the PNAD, we use census demographic data from 1991 and 2000, as well as a complete population count conducted in 1996. Hierarchically non-structured and spatially structured growth models that gain strength from all the sampled municipalities are proposed and compared.

Key Words: Markov Chain Monte Carlo (MCMC); Population projection; Spatial models.

1. Introduction

Like many other countries, the demand for detailed and updated small area statistics has been steadily growing in Brazil. This increasing demand is motivated by the need to have a more precise picture of subregions and has been driven by issues of distribution, equity and disparity. For instance, there may exist subregions or subgroups that are not keeping up with the overall average in certain respects. Therefore, there is a need to identify such regions and to have statistical information at that geographical level before taking any possible remedial action. Besides these national requirements, local authorities are faced with the need of having reliable estimates, such as demographic characteristics, for analysis, planning and administration purposes.

In Brazil, one important example of the demand for reliable estimates is related to how constitutionally mandated federal revenue sharing is apportioned annually to the various municipalities (Brazil is a federated republic made up of states and the Federal District. The states are divided into municipalities, which share characteristics of cities and counties - they can contain more than one urban area, but they have a single mayor and municipal council). The predicted number of inhabitants in a municipality is used by the federal government as a criterion to distribute funding. Hence, there is a need to obtain reliable municipal population forecasts in order to fairly apply this criterion, regulated by federal law.

An important source of demographic data is the annual Household Survey (PNAD). However, this survey is not designed to produce estimates at the municipal level. In other words, apart from a few municipalities, the municipal sample sizes are not large enough to yield acceptable standard errors when the direct survey estimates are used. Furthermore, a considerable number of municipalities are not sampled at all.

The current approach to obtain municipal population estimates is based on making prediction for a larger area at first, and then using some auxiliary information to allocate the total predicted population to the municipalities. In turn, prediction for a larger area is done by assuming that birth, mortality and migration rates are the same for all municipalities. The major drawback of this approach is that it relies on the assumed model evolution. It does not take into account all uncertainties and does not provide, in general, error measures of the estimates.

The small area estimation problem has received attention in the statistical literature due to the growing demand for detailed statistical information from the public and private sectors. An excellent and updated account of methods and applications of small area estimation can be found in Rao (2003). The main source of small area data is provided by periodic surveys whose sample sizes are not large enough to provide reliable estimates for the areas. One way of tackling this problem is to gain strength from all areas and through other sources of related data. As stated in Pfeffermann (2002), the sources of data suitable for this task can be classified into two categories: data obtained from other similar areas with respect to the characteristic of interest and past data obtained for the characteristic of interest and auxiliary information. In our demographic context, the main source of related data is provided by the 1991 and 2000 censuses and a complete count of the population carried out in 1996.

The aim of this work is to obtain estimates of the municipal populations based on survey data provided by the PNAD and census data. A non-structured hierarchical model is proposed and its fitness and predictive power are

1. Debora F. Souza, Department of Methods and Quality, IBGE, Rio de Janeiro, 20031-170. E-mail: debora.souza@ibge.gov.br; Fernando A.S. Moura, Universidade do Brasil-UFRJ, Rio de Janeiro, Brazil. E-mail: fmoura@im.ufrj.br; Helio S. Migon, Universidade do Brasil-UFRJ, Rio de Janeiro, Brazil. E-mail: migon@im.ufrj.br.

evaluated. We also consider a spatially structured hierarchical model, in the spirit of Moura and Migon (2002), since the population per area and its growth pattern might be related to the development of its neighboring areas. For the sake of simplicity, from now on we respectively call the non-structured hierarchical and spatially structured hierarchical models as the Hierarchical model and Spatial model.

In Section 2 the main data sources used in this work are described. In Section 3, the proposed models and a model selection criteria are presented. Applications with real and a simulated data are presented in Section 4. Finally, Section 5 contains a brief summary with an outline for future research.

2. Data set

The input data for the models introduced in Section 3 are taken from the annual Household Surveys (PNADs) from 1992 to 1999, the 1991 and 2000 census data and a complete enumeration of the population carried out in 1996. In order to evaluate the proposed approach, the municipalities of São Paulo State are considered as the areas of interest.

In this section we present a brief description of these data sources, reporting their main advantages and limitations. The population direct estimates of sampled municipalities were obtained from the PNAD. As explained in Section 3, these estimates are regarded as the input data for making inference about our target parameters. The two censuses and the 1996 population count are also utilized in our application.

The Brazilian Demographic Census is the main source of information about the population. It is carried out every ten years, usually in the beginning of the decade. Although the objective is to count all the population, some enumeration errors are found. The magnitude of the errors is evaluated through a post enumeration survey carried out soon after the completion of the census.

The annual Household Survey (PNAD) is designed to produce basic information about the socioeconomic situation of the country. The investigation unit is the household, for which yearly information about the number of dwellers, their gender, education level, employment, *etc.* is collected. The survey is not carried out in a census year, and was also not conducted in 1994 for administrative reasons. The sample is selected by a three-stage cluster sampling design. The primary and secondary units are respectively the municipality and enumeration areas (with 250 households on average). The municipalities are stratified according to their population sizes as obtained from the last census. In the first stage, all municipalities belonging to the metropolitan regions and the state capitals (which in Brazil are normally the largest cities in the respective states) are sampled. The municipalities whose populations are greater than some cutoff value are also included in the sample with probability one. The ones left are stratified and two of them are sampled from each stratum with probability proportional to their population sizes.

The enumeration areas are sampled with probability proportional to the number of households residing in the area in the last census. Finally, in the last stage the households are sampled systematically with equal probability from a list, which is updated at the beginning of the survey. The municipalities and enumeration districts are kept the same in all the surveys carried out in the same decade, while households are sampled every year.

Since each area is sampled with probability proportional to its respective number of households, it could be argued that the sampling mechanism is informative with respect to the population of the area. However, since the response variable actually used in this work is the area density, it is reasonable to assume that the sample selection mechanism is not relevant. Thus, this issue is not exploited in this work. A good reference about how to make small area inference under informative sampling is Pfeffermann and Sverchkov (2007). We also recommend Pfeffermann, Moura and Silva (2006) for readers interested in how to employ a Bayesian approach to hierarchically modeling under informative sampling.

3. Model specification

3.1 Exponential growth model

Let y_t be sample values of a distribution belonging to an exponential family with expected value given by $\pi_t = E(y_t \mid \theta_t)$ where θ_t is a vector of unknown parameters.

An important and wide class of exponential growth models parameterized by $(\alpha, \beta, \gamma, \phi)$ is defined as:

$$\pi_t = [\alpha + \beta \exp(\gamma t)]^{1/\phi}. \tag{1}$$

Some special well-known cases in the literature are:

(1) Logistic: with $\phi = -1$, $\pi_t^{-1} = \alpha + \beta \exp(\gamma t)$;

(2) Gompertz: with $\phi = 0$, defining (1) as $\log(\pi_t) = \alpha + \beta \exp(\gamma t)$;

(3) Modified exponential: with $\phi = 1$, $\pi_t = \alpha + \beta \exp(\gamma t)$.

The main advantage of using model (1) is the possibility of keeping the observations y_t in the original scale, changing only the trajectory of π_t, making interpretation easy. Furthermore, the time intervals do not need to be of the same length, allowing the data to come from different reference sources (see Section 4 for further details).

When $\psi = \exp(\gamma) < 1$, the process is non-explosive, implying that π_t converges to $\alpha^{1/\phi}$ when $t \to \infty$, with the convention that for $\phi = 0$, this quantity is equal to $\log(\alpha)$. When $\psi > 1$, the curves are concave for $\phi \geq 0$ and $\beta > 0$, leading to an explosive process. This class of models is called the generalized exponential growth model. Migon and Gamerman (1993) show how the exponential growth model can be viewed as a particular case of a general dynamic model.

3.2 Hierarchical growth models

In this paper our main parameters of interest π_{it} are nonlinear exponential growth functions with some parameters that are hierarchically or spatially structured. Spatially structured models provide alternative ways for connecting similar neighboring areas. We further assume that the sampling variance σ_{it}^2 follows a model that depends on the sample size in the respective municipality. In this work, hierarchical and spatial models are fitted and compared.

We assume that the population sizes are available for all the m municipalities of São Paulo State for the census years of 1991 and 2000, as well as the complete population count in 1996. From now on, we simply refer to them as the census data. In order to improve the hypothesis of exchangeability of the parameters describing the mean of the process, our response variables are set as the sampled municipal density estimates instead of the municipal population estimates. See also the end of Section 2 for further reasons for using the densities.

For each period, estimates of these quantities are available only for $k < m$ first-stage units municipalities of the PNAD sample. In order to estimate the municipal density, we simply divide the total population estimate by the respective municipal area.

Let y_{it} be the population density obtained from the census data or estimated by the PNAD at time t, $t = 1, \ldots n$ for the i^{th} municipality, $i = 1, \ldots, m$. Our aim is to make inferences about the true population density π_{it} for the population of all municipalities, including those that are not sampled. In the next section, true municipal population densities π_{it} are modeled via a stochastic nonlinear hierarchical growth function. We assume that the random quantities y_{it} are normally distributed with mean π_{it} and variance σ_{it}^2.

We use a Bayesian approach in this work. Therefore, predictions are described by probability distributions, giving the opportunity for users to analyze the uncertainties involved in the decision process. This fact is one of the advantages, among many others, of using this kind of approach.

Only in the census years are the y_{it} obtained for all the municipalities of São Paulo State. Although the census attempts to obtain complete enumeration of the whole population, coverage errors can occur. The following model is assumed therefore for the census data and the data obtained from the PNAD, with exception that the variances σ_{it}^2 are set to be smaller for the census data (see Section 3.4 and also the final remarks in Section 5):

$$\begin{aligned} y_{it} &= \pi_{it} + \varepsilon_{it},\ \varepsilon_{it} \sim N(0,\ \sigma_{it}^2) \\ \pi_{it} &= \{\alpha_i + \beta \exp(\gamma_i t)\}^{1/\phi} \\ \alpha_i &= \alpha + \xi_{\alpha_i},\ \xi_{\alpha_i} \sim N(0,\ \sigma_{\xi_\alpha}^2) \\ \gamma_i &= \gamma + \xi_{\gamma_i},\ \xi_{\gamma_i} \sim N(0,\ \sigma_{\xi_\gamma}^2) \end{aligned} \tag{2}$$

where the prior distributions of α, β and γ are given by: $\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$, $\beta \sim N(\mu_\beta, \sigma_\beta^2)$, $\gamma \sim N(\mu_\gamma, \sigma_\gamma^2)$. It should be noted that information from all areas is obtained through the hierarchical structure of the parameters α_i, and γ_i. Another way of borrowing information between municipalities is to assume that α_i are spatially structured (see Section 3.3). Supposing that the mean π_{it} is non-explosive, the parameter $\alpha^{1/\phi}$ can be regarded as the value at which the mean municipal population stabilizes. The parameters β and γ affect the evolution of the density over time. The prior distributions of α, β and γ can be chosen by taking advantage of some prior demographic knowledge of the expected population evolution. In our application, we set $\phi = 1$, implying that for $t = 0$ the true value density in each municipality is given by $\alpha_i + \beta$. The hierarchical structure imposed on the parameters α_i, implies that the expected value of the true density for any municipality at $t = 0$ is $\alpha + \beta$. To assume that the growth parameters, γ_i, have a hierarchical structure means that the densities have different growth rates but share the same mean. A small simulation study (see Section 4.1) guides us to keep the β parameter fixed for all areas, without any loss of generality, since the levels are still different for different municipalities. In all models considered in our application, we assume that $\tau_\alpha^2 = \sigma_\alpha^{-2} \sim G(a_\alpha, b_\alpha)$, $\tau_\gamma^2 = \sigma_\gamma^{-2} \sim G(a_\gamma, b_\gamma)$. In order to assign vague priors, in Section 4.2 we set small values for the parameters related to these precision prior distributions.

The assumption that the mean function π_{it} is given by an exponential growth curve allows adjusting for increasing or decreasing population density. The sources of data used have different reference data and are not equally spaced in time. In this case, the use of an exponential growth curve yields an extra advantage, since we can simply make a scale of time in order to conform with the different data sources, as explained in the application section 4.

3.3 Spatial model

In the Hierarchical model presented in the previous section, the information from all areas is combined in order to predict the population of a particular area. However, it is reasonable to assume that two or more neighboring municipalities have more similar demographic densities than two other arbitrarily chosen ones. The regional structure is represented in the joint prior distribution of the random spatial effects. We consider that two areas are neighbors if they share a border.

In our proposed model, the demographic density in an area i at time t, π_{it}, is affected by its neighboring areas by adding random spatial effects δ_{α_i} to the parameters α_i, that is, $\alpha_i = \alpha + \delta_{\alpha_i}$, where α is a term representing the intercept. Therefore, α_i vary only with the spatial effect, representing a local effect, while the growth parameters γ_i 's are regarded as similar among all areas (overall effect).

The relationship between neighboring areas is defined in the prior distributions of δ_{α_i}. The prior joint distribution of $\delta_\alpha = (\delta_{\alpha_1}, ..., \delta_{\alpha_m})'$ given the hyperparameter σ^2_α, is defined as in Mollié (1996):

$$p(\delta_\alpha \mid \sigma^2_\alpha) \propto \frac{1}{\sigma_\alpha^{m/2}} \exp\left\{-\frac{1}{2\sigma^2_\alpha}\sum_{i=1}^{m}\sum_{k<i} w_{ik}(\delta_{\alpha_i} - \delta_{\alpha_k})^2\right\} \quad (3)$$

where w_{ik} are the weights associated with the regional structure. The weights were chosen such that $w_{ik} = 1$, if i and k are contiguous, and $w_{ik} = 0$, otherwise. The distribution of $\delta_\alpha \mid \sigma^2_\alpha$ is evidently improper, since we can add any constant to all of the δ_{α_i} and $p(\delta_\alpha \mid \sigma^2_\alpha)$ is not affected. Thus, we must impose a constraint to ensure that the model is identifiable. We set $\sum_{i=1}^m \delta_{\alpha_i} = 0$ and assign a uniform prior distribution on the whole real line to the intercept α. It is not difficult to see that this procedure leads to a proper $(m-1)$ dimensional likelihood density, see Besag and Kooperang (1995) for further details.

The prior conditional distribution of δ_{α_i}, given the effects δ_{α_k} of the remaining areas and the hyperparameter σ^2_α, is normal with mean and variance given by:

$$E[\delta_{\alpha_i} \mid \delta_{\alpha_k},\ k \in \partial i,\ \sigma^2_\alpha] = \overline{\delta}_{\alpha_i}$$

$$\text{Var}[\delta_{\alpha_i} \mid \delta_{\alpha_k},\ k \in \partial i,\ \sigma^2_\alpha] = \frac{\sigma^2_\alpha}{w_{i+}}$$

where $\overline{\delta}_{\alpha_i}$ denotes the arithmetic mean of the δ_{α_j} for $k \in \partial i$ (the contiguous areas of i), and $w_{i+} = \sum_{k=1}^m w_{ik}$ is the number of neighboring municipalities of i.

Figure 1 shows the demographic densities of São Paulo municipalities in 1991. These municipalities tend to be concentrated geographically according to density classes. This suggests that the spatial model can be usefully applied.

3.4 Modeling the sampling variances

Since we use data from two different sources, it makes sense to assume that the sampling variances vary over time. Furthermore, we can also consider that the variances change with the areas.

For the years in which the data are provided by the PNAD, we assume the following model for the sampling variances:

$$\log(\sigma^2_{it}) = \eta_0 + \eta_1 .(1/n_i) \quad (4)$$

with n_i representing the number of enumeration areas sampled in the i^{th} area. This model captures the expectation that the variance gets smaller as the sample size increases.

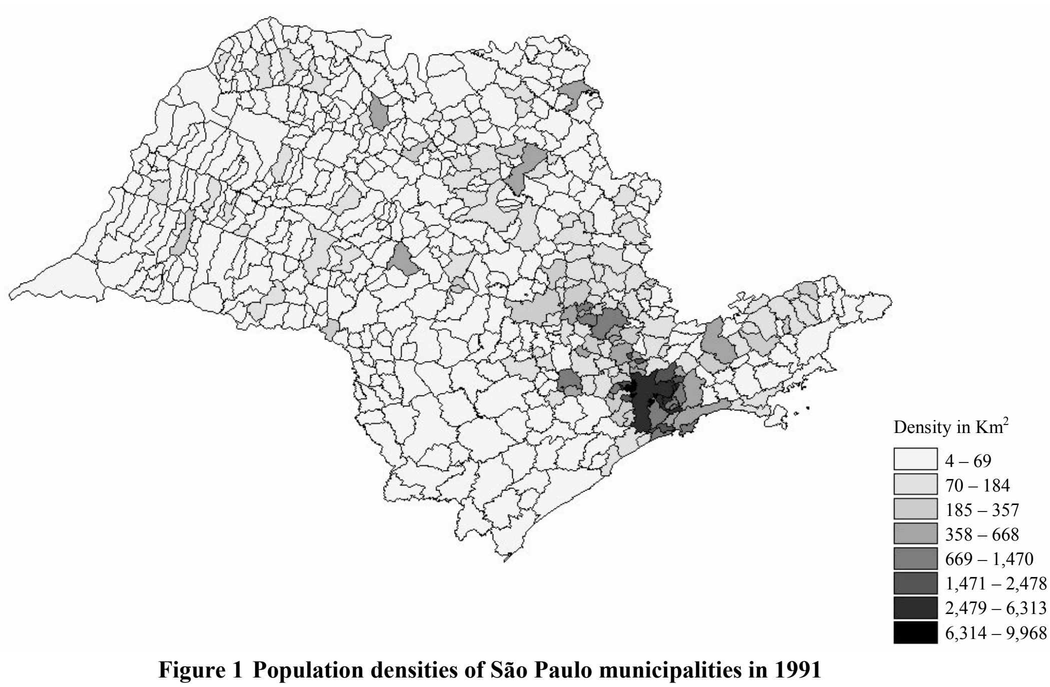

Figure 1 Population densities of São Paulo municipalities in 1991

For the years that the censuses were carried out, we assumed that σ_{it}^2 is known and $\log(\sigma_{it}^2) = \log(v_{it})$ where v_{it} is calculated in such a way that the census coverage error is 5% for all areas. This hypothesis implies that the true population in each area for census years lies in the interval given by the observed population in the census plus or minus 5% of this value. Therefore, for the census years we set the standard deviation as: $\sigma_{it} = 0.05 * (y_{it}/2)$. Assuming known variance in the census years is a way of giving more weight to census data, since one would expect a complete census to provide more reliable information than survey data. Independent normal distributions are assumed for the parameters η_0 and η_1: $\eta_k \sim N(\mu_{\eta_k}, \phi_{\eta_k})$; $k = 0,1$. In order to assign vague priors to the η 's, we set both prior means as zero and large values for the ϕ_η 's. See Section 4.2 for details.

3.5 Summary of the models

The prior distributions of the common parameters of the Spatial and Hierarchical models are the same as already described for the former. The distributions of the random spatial effects are specified in Section 3.3. The variance σ_{it}^2 in the Spatial model was stated as in the Hierarchical model. A summary of the models in Section 4 is presented in Table 1. For the sake of simplicity, the application was carried out by fixing $\phi = 1$ in both models.

3.6 Computational issues

The posterior distributions of the parameters for the models proposed cannot be obtained in closed forms. Therefore, it is necessary to use numerical approximation methods. One alternative, often used and easy to implement, is to generate samples of these distributions based on the Markov Chain Monte Carlo (MCMC) algorithm. Since the full conditional distributions of all the model parameters have closed form, except for the vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)$, we employed the Gibbs sampler algorithm with one acceptance/rejection algorithm step for sampling from the vector $\boldsymbol{\gamma}$. Let π_{it} be the population density in the i^{th} area at time t. The following steps summarize how to sample from the posterior distribution of π_{it}:

1. Generate $\alpha_i^{(l)}, \beta^{(l)}, \gamma_i^{(l)}, \alpha^{(l)}, \gamma^{(l)}, \tau_\alpha^{2(l)}, \tau_\gamma^{2(l)}, \eta_0^{(l)}$ and $\eta_1^{(l)}$ for $l = 1, \ldots, M$, where M is the number of MCMC samples generated from the full conditional distributions of all model parameters including the random effects;
2. Calculate $\pi_{it}^{(l)} = \alpha_i^{(l)} + \beta^{(l)}\exp(\gamma_i^{(l)}t)$;

Three informal checks for convergence, based on graphical techniques, were applied for assessing the convergence when fitting our proposed models. They consist of observing the histogram, the trace and the autocorrelation function for each of the sampled values calculated. The histogram analysis allows us to identify possible departures from convergence, such as the presence of multiple modes. The trace of the multiple chains simulated in parallel, each one with different starting points and overdispersed with respect to the target distribution, provides a rough indication of stationary behavior when the sequences of values tend to oscillate in the same region. The plot of the autocorrelation function allows identifying whether the sampling can be regarded as independent.

In addition to these informal checks, other more formal criteria were applied. The criteria introduced by Brooks and Gelman (1998) and implemented in WinBugs 1.4 (Spiegelhalter, Thomas, Best and Lunn 2004) permit diagnosing whether dispersion within chains is larger than dispersion between chains. Consider I parallel chain and a parameter of interest λ. Let λ_i^j be the j^{th} value of the i^{th} chain, for $i = 1, \ldots, K$ and $j = 1, \ldots, J$. Then the variances between chains $\hat{B}$ and within chains $\hat{W}$ are given by

$$\hat{B} = J(K-1)^{-1}\sum_{i=1}^{K}(\bar{\lambda}_i - \bar{\lambda})^2$$

and

$$\hat{W} = \{K(J-1)\}^{-1}\sum_{i=1}^{K}\sum_{j=1}^{J}(\lambda_i^j - \bar{\lambda}_i)^2$$

where $\bar{\lambda}_i$ and $\bar{\lambda}$ respectively are the average of observations of chain i, $i = 1, \ldots, K$ and the global average. Under convergence, all these KJ values are drawn from the posterior of λ and the variance of λ can be consistently estimated by $\hat{B}$, $\hat{W}$ and the weighted average $\hat{\sigma}_\lambda^2 = (1 - 1/J)\,\hat{W} + (1/J)\,\hat{B}$.

Table 1
Summary of the models employed

model	parameters	variance	prior distribution
Hierarchical	$\alpha_i = \alpha + \xi_{\alpha_i}$ β $\gamma_i = \gamma + \xi_{\gamma_i}$	$\log(\sigma_{it}^2) = \eta_0 + \eta_1(1/n_i)$, for survey data σ_{it}^2 is assumed to be known for census data	$\eta_0 \sim N(\mu_{\eta_0}, \phi_{\eta_0})$ $\eta_1 \sim N(\mu_{\eta_1}, \phi_{\eta_1})$
Spatial	$\alpha_i = \alpha + \delta_{\alpha_i}$ β $\gamma_i = \gamma + \xi_{\gamma_i}$	$\log(\sigma_{it}^2) = \eta_0 + \eta_1(1/n_i)$ in the survey σ_{it}^2 is assumed to be known for census data	$\delta_{\alpha_i} \mid \delta_{\alpha,-i}, \tau_\alpha^2 \sim N(\bar{\delta}_{\alpha_i}, \tau_\alpha^2/w_{i+})$ $\sum_{i=1}^{m}\delta_{\alpha_i} = 0$ $\eta_0 \sim N(\mu_{\eta_0}, \phi_{\eta_0})$ $\eta_1 \sim N(\mu_{\eta_1}, \phi_{\eta_1})$

If the chains have not yet converged, then initial values will still be influencing the trajectories and $\hat{\sigma}_\lambda^2$ will overestimate σ_λ^2 until stationarity be reached. On the other hand, before convergence, $\hat{W}$ will tend to underestimate σ_λ^2. Following these reasoning, Brooks and Gelman (1998) proposed an iterated graphical approach, which is implemented in WinBugs 1.4. It allows to check if: (i) the weighted posterior variance estimated $\hat{\sigma}_\lambda^2$ and the within-chain variance $\hat{W}$ stabilize as a function of J, and (ii) the variance reduction factor, $\hat{R} = \hat{\sigma}_\lambda^2/\hat{W}$, approaches 1.

4. Application

In this section we present two applications of our approach, the first one with a simulated data set and the second one with the real data set that motivated this work. The simulation study aims to check if the parameters of interest are being properly estimated, as well as to perform some sensitivity analysis with respect to the form of the prior distributions used for fitting the model.

4.1 Application to simulated data

We carried out a small simulation study fitting the Hierarchical and Spatial models presented in Section 3. The true model hyperparameters related to the growth curve were fixed as $\alpha = 40, \beta = 25, \gamma = 0.05$. Thus, we are considering a situation where the population size approximately doubles in 25 years. The parameters related to the sampling variance model were fixed as $\eta_0 = 6.5, \eta_1 = 0.5$. Finally, the precision parameters were respectively set as $\tau_\alpha^2 = 0.0001$ and $\tau_\gamma^2 = 400$. The precision τ_α^2 and τ_γ^2 were fixed to be in agreement with the scales of the quantities they respectively measure. The intercept presents more relative variation between areas than the growth parameter, which is expected in practical situations.

Since it is well recognized that the form of the priors has more impact on the component of variance parameters than the fixed parameters, we fitted the simulated data using two different vague priors for the parameters related to the variances: uniform for the standard deviation, which is one of the priors recommended by Gelman (2006) for linear hierarchical models, and gamma for the precision, commonly used as the default in some computational packages. In the first case, we assigned $\sigma_\alpha \sim U(0, 1{,}000)$ and $\sigma_\gamma \sim U(0, 100)$, where $\sigma_\alpha = 1/\tau_\alpha$ and $\sigma_\gamma = 1/\tau_\gamma$. In the second case, we considered $\tau_\alpha^2 \sim G(0.001, 0.001)$ and $\tau_\gamma^2 \sim G(0.001, 0.001)$. For the other parameters, we set $\alpha \sim U(-\infty, +\infty)$, for the Spatial Model (see Section 3.3 for further details) and $\alpha \sim N(0, 10^6)$ for the Hierarchical model. For the others parameters we set $\beta \sim N(0, 10^6)$, $\gamma \sim N(0, 10^2)$, $\eta_0 \sim N(0, 10^4)$ and $\eta_1 \sim N(0, 10^4)$ for both models. The effect of the number of small areas is also investigated. We simulated separate data from the Hierarchical and Spatial models with $m = 60$ and $m = 100$ areas in each case. For each combination of the number of areas and the model employed we generated 200 data sets. Therefore, a total of 800 sets of artificial data was simulated. The distribution of the sample sizes within the areas is the same for the simulated data sets with 60 and 100 areas. Table 2 presents the relative frequencies of the small areas sample sizes for the both simulated data sets. These sample sizes are very similar to the sample sizes in the real data that underlines this simulation study. The number of neighbors employed in the spatial model varies from 1 to 12 and each area has on average 5 neighbors. We considered a total period of $n = 9$ years.

Table 2
Relative frequencies of the small area samples sizes for both simulated data sets

Sample size	Relative frequency
2	0.05
5	0.20
8	0.25
10	0.25
12	0.20
15	0.05

In order to get rid of chain correlation, we generated 20,000 samples after discarding the first 10,000. There is no evidence for non-convergence of the Hierarchical and the Spatial model parameters. A careful analysis of some outputs obtained from the MCMC samples for some simulation sets suggests that convergence was achieved for all model parameters. We assessed the statistical properties of the population density (π_{it}) estimates by investigating the average of the absolute relative error of the estimates (ARE) and the mean square error (MSE), respectively given by:

$$\text{ARE}_{i,t} = \frac{1}{200}\sum_{l=1}^{200}\frac{|\hat{\pi}_{i,t}^{(l)} - \pi_{i,t}^{(l)}|}{\pi_{i,t}^{(l)}}$$

and

$$\text{MSE}_{i,t} = \frac{1}{200}\sum_{l=1}^{200}(\hat{\pi}_{i,t}^{(l)} - \pi_{i,t}^{(l)})^2,$$

$i = 1, \ldots, m,\ t = 1, \ldots, n.$ There is no much variation, as far as the ARE values are concerned. For the two models fitted and both small area sample sizes tried, the ARE values are around 1.5%.

Table 3 shows a summary of the MSE values obtained from the simulations carried out under the Spatial and Hierarchical models with 60 and 100 areas and respectively assigning gamma and uniform priors to the precision and to the standard deviation of the parameters related to the variance. It can be seen from Table 3 that the MSEs are not affected by the use of different vague priors. It is noteworthy that increasing the number of areas from 60 to 100 results in a small decrease of 6% in the median of the MSE for the Spatial model. However, for the case of the Hierarchical model, the decrease is about 13%.

Table 3
Summary of mean square error distribution for the spatial and hierarchical models

Model	Num. of areas	Gamma prior			Uniform prior		
		1st Qu.	Median	3rd Qu.	1st Qu.	Median	3rd Qu.
Spatial	60	0.398	1.741	3.574	0.394	1.737	3.595
	100	0.525	1.637	3.538	0.524	1.641	3.517
Hierarchical	60	0.542	2.218	6.262	0.646	2.223	6.278
	100	0.594	1.959	5.593	0.596	1.960	5.619

We also investigated the percentage coverage of nominal 95% credible intervals. The results are presented in Table 4. As far as this simulation study is concerned, the intervals for the parameters of interest have in general the correct coverage percentages for both models investigated and these results do not depend on wether we have 60 or 100 areas. However, with a small number of areas we could face convergence problems unless we tighten the priors for the hyperparameters. The simulation study reveals that the population prediction is not affected by the forms of the vague priors assigned to the variance of the intercept term.

Table 4
The coverage rates of nominal 95% credible intervals for the population densities

Model	Num. of Areas	Gamma prior coverage(%)	Uniform prior coverage(%)
Spatial	60	96	96
	100	96	96
Hierarchical	60	94	94
	100	95	95

We analyzed the model fit when data generated from a model were fitted by the correct and the wrong models. Figure 2 presents the mean square error for the following situations: (a) data generated from the Spatial model and fitted by the Spatial and Hierarchical models and (b) data generated from the Hierarchical model and fitted by the Spatial and Hierarchical models. Since the form of the priors assigned to the parameters related to the variance does not affect the inference, we set uniform priors for both models. The ARE measures are shown in Figure 3.

It can be seen from Figure 2 that when the data are generated from the simpler model (Hierarchical) the more complex estimation procedures (Spatial) do not suffer any appreciable worsening of efficiency. On the other hand when the data are generated from the more complex model (Spatial) the simpler estimator (Hierarchical) has some inferior properties. However, this result does not hold for the ARE measurements. Figure 3 shows that fitting the model not used for generating the data results in appreciable increase in the relative bias. As it might be expected, model fitting and diagnostics are crucial in order to get suitable prediction of the small area population.

4.2 Application to real data

The PNAD data sets from 1992 to 1999, (excluding 1994 and 1996) and the population census data of 1991, 1996 and 2001 were used in our application. Our areas of interest are all the municipalities in São Paulo State, a total of 572 areas, of which 111 areas were sampled by the PNAD survey. Figure 4 shows the areas sampled by the PNAD, classified by the sampling definition: areas belong to metropolitan regions and self-representing areas (sampled with probability equal to 1) and non-self-representing areas. It should be noted that the census and PNAD have different periods of reference. We set $t = 0$ for the 1991 census. Thus, the values of t for the data provided by the PNAD are equal to the number of years between the reference period of the 1991 census and the respective PNAD. For instance, a survey datum provided by the PNAD 18 months after the 1991 census corresponds to $t = 1.5$.

Figure 5 shows the estimated coefficient of variation of the direct estimator by areas'sample sizes. These estimates are based on PNAD data. It can be seen that these coefficients of variation vary considerably with the areas and tend to decrease as the sample size increases. The high values of these coefficients show the difficulty in using only the direct estimator to provide municipal estimates. Furthermore, we cannot make any prediction for nonsampled areas by using only the direct estimators.

4.3 Specification of the prior distributions

The mean of the normal prior distributions of the parameters α, β and γ, related to the population evolution, were assigned by first expanding the function $\alpha + \beta\exp(\gamma t)$ around zero in a Taylor series up to the second order and then equating the resulting expression to the values of the mean density in the 1991 and 2000 censuses and the 1996 population count. In the absence of prior information, we considered a reasonably large value (10^6) for the prior variances of α, β and γ. Thus, we set $\alpha \sim U(-\infty, +\infty)$ (see Section 3.3 for further details), for the Spatial Model and $\alpha \sim N(370, 10^6)$, for the Hierarchical model and $\beta \sim N(726, 10^6)$, $\gamma \sim N(0.04, 10^6)$ for both models. The reason for this adjustment is to obtain a reasonable value of the prior means, but one that is essentially vague. Regarding the precisions and η_0, η_1, we assigned relatively vague priors: $\tau_\alpha^2 \sim \text{Ga}(0.001, 0.001)$, $\tau_\gamma^2 \sim \text{Ga}(0.001, 0.001)$, $\eta_0 \sim N(0, 10^6)$ and $\eta_1 \sim N(0, 10^6)$.

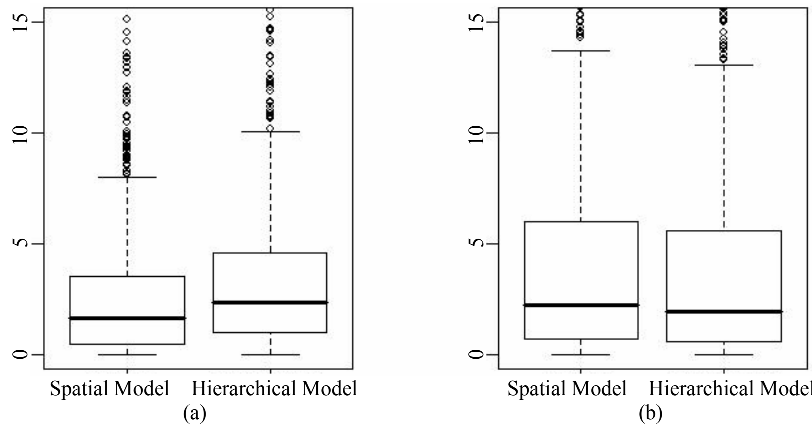

Figure 2 Box plots of mean square error (MSE) for the cases: (a) data generated from the Spatial model and respectively fitted by the Spatial and Hierarchical models and (b) data generated from the Hierarchical model and respectively fitted by the Spatial and Hierarchical models

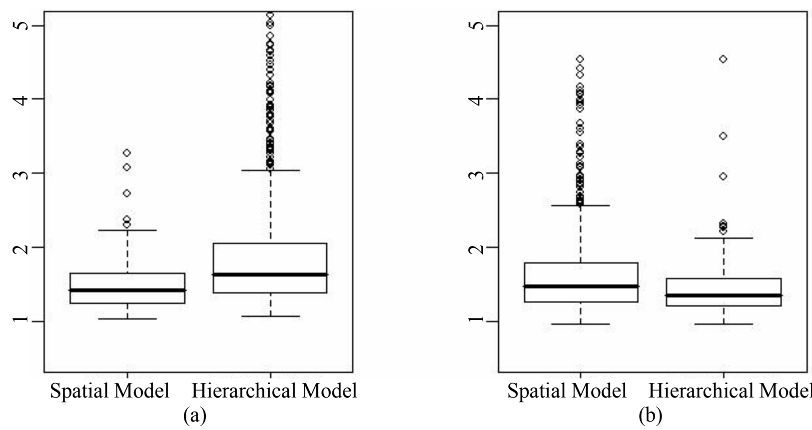

Figure 3 Box plots of absolute relative error (ARE) for the cases: (a) data generated from the Spatial model and respectively fitted by the Spatial and Hierarchical models and (b) data generated from the Hierarchical model and respectively fitted by the Spatial and Hierarchical models

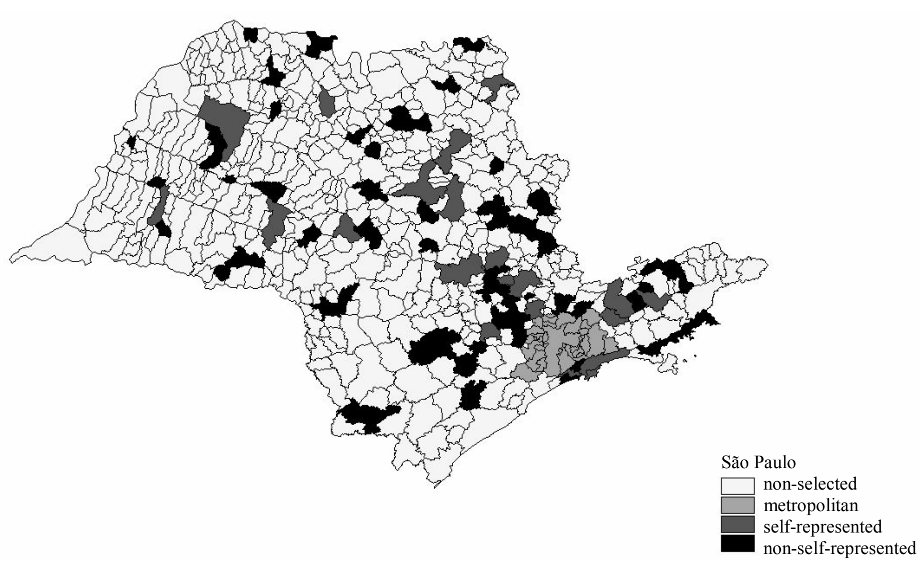

Figure 4 São Paulo municipalities sampled by the PNAD classified by the sampling definition

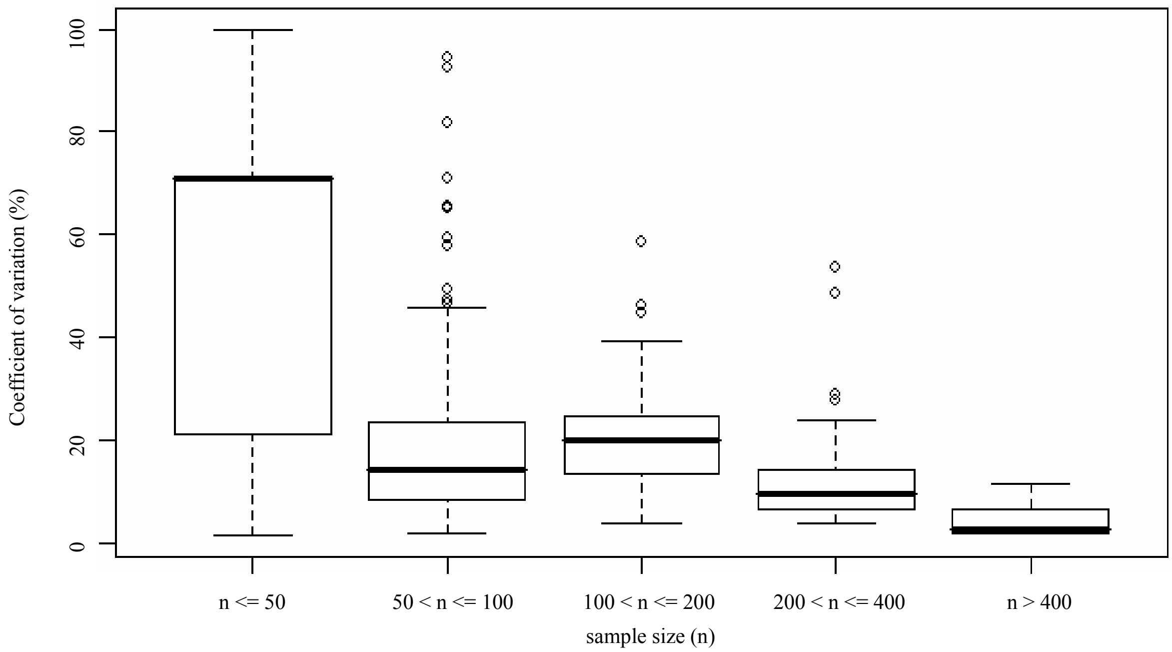

Figure 5 Boxplot of the coefficients of variation of the direct population estimates

4.4 Some results

We generated 20,000 samples after discarding the first 5,000. There is no evidence for non-convergence of the Hierarchical and the Spatial model parameters. A careful analysis of the MCMC outputs suggests that convergence was achieved for all model parameters. We summarize the results obtained by fitting the Hierarchical model (3) to the data provided by the PNAD survey. The posterior means of the model parameters were used as the point estimates. Table 5 presents these estimates together with the respective square root of the posterior variance. It can be seen from Table 5 that the estimate of η_1 is significantly positive, which agrees with what is expected by equation 4: the greater the sample size, the smaller σ^2_{it}.

Table 5
Summary of the model (2) parameter posterior distributions

parameter	posterior mean	posterior std
α	892.500	202.000
β	105.700	1.278
γ	0.072	0.008
η_0	10.620	0.133
η_1	3.185	0.484
τ^2_α	2.174E-7	2.961E-8
τ^2_γ	139.000	19.560

Figure 6 shows that the posterior means of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ that index the hierarchical model seem to be spatially distributed. The parameters of neighboring areas seem more alike than those of distant areas, which suggests applying the Spatial model.

4.5 Model selection

The Expected Prediction Deviance (EPD) (Gelfand and Ghosh 1998) measure was applied to help choose the most suitable model. The EPD measure is the sum of two terms. The first term, denoted by G, can be interpreted as a goodness-of-fit measure and the second term, denoted by P, as a penalty term for underfitted as well as overfitted models. The respective expressions for G and P are given by: $G = \sum_{i=1}^m \sum_{t=1}^n (y_{it} - E(y_{it}^{\text{rep}}|M))^2$ and $P = \sum_{i=1}^m \sum_{t=1}^n V(y_{it}^{\text{rep}}|M)$, where the expectations and the variances are with respect to the posterior predictive distribution associated with a future observation (y_{it}^{rep}) of y_{it} generated under the assumed model (M). According to this criterion, the smaller its value, the better the model. As can be seen in Table 6, the EPD criterion slightly favors the Spatial model.

4.6 Analysis of the results

The most disaggregated level for which the PNAD provides precise estimates is the metropolitan region, which is a set of contiguous municipalities. In order to validate the results obtained with the spatial model, population estimates for the greater São Paulo metropolitan region were compared to the official statistics projections. The posterior distribution of $\mu_t = \sum_{i=1}^{r} \pi_{it} * A_i$ is easily obtained by adding $\mu_t^{(l)} = \sum_{i=1}^{r} \pi_{it}^{(l)} * A_i$ to the MCMC algorithm, where μ_t represents the total population of the metropolitan region at time t and r is the number of municipalities belonging to that metropolitan region.

Table 6
Measures for selecting models for demographic density

Model	*G*	*P*	EPD
Hierarchical	1.37E+09	6.14E+09	7.51E+09
Spatial	1.05E+09	6.19E+09	7.24E+09

Figure 7 compares the population estimates (μ_t) of the São Paulo metropolitan region obtained by the Spatial model and the official statistics. The solid lines represent the limits of the 95% credible intervals of μ_t, while the dotted line shows the respective point estimates. The symbol (+) represents the observed official statistics. It is noteworthy that some official statistics projections are outside of the credibles inferior limit (including the 1991 Census). This indicates that further investigations should be made in order to find out the reasons for these discrepancies. However, when we compare them at municipality level, the overall conclusion is that the model predictions and official statistics reasonable agree. The 95% credible intervals contains 92.4% of the official statistics projections. The average of the absolute relative error (ARE) between the estimated population density and the official statistics projection are 3%. These ARE measures are on average nearly the same for selected and non-selected municipalities.

Figure 8 compares the point estimates of the population sizes (μ_{it}) with the official projection statistics and the official census population sizes for a sampled municipality. The official projection methodology assumes that a set of small areas and a larger area, which contains them, have the same population growth rate pattern. The population of the larger area is projected by a component method and then proportionally allocated to the small areas. The component method uses data from the most recent census as well as the number of births and deaths and net migrations obtained from administrative records. The component method projects the population for a time t by adding the population in a previous time with the number of births and net migrations and subtracting the number of deaths in the same time interval.

The solid lines represent the 95% credible intervals for μ_{it} obtained by the Spatial model, while the dotted line shows the respective posterior means. The symbol (+) represents the official population projection for the inter-census period and the observed population in the census years. It is noteworthy that the point estimates are relatively close to the official projection statistics and the population obtained in the census year. This indicates that the use of the proposed model yields reliable estimates at municipality levels, with the extra advantage of providing a measure of the respective error.

We also analyze the estimates obtained for some municipalities not sampled in the PNAD. Figure 9 shows the model predictions, the 95% credible intervals, the official projection statistics and the observed population values in the censuses for a non-sampled municipality (+). It can be seen that the predictions obtained by the Spatial model reasonably agree with the official figures.

5. Final remarks

The model used in this article identifies the population growth trend of the municipalities. Reasonable estimates of the municipal populations are obtained for years with survey data, as well as for the years where census data are available. The point estimates have good precision and reasonably agree with estimates obtained for larger areas using other technique. The past information can be updated as soon as estimates become available from a new census or survey. Furthermore, the proposed approach provides the probability distribution of the quantity of interest, aiding the decision-making process.

Further work should be done in order to allow for autocorrelation of the parameters of interest over time. Extra information about the sampling variance estimates of the direct estimators could also be regarded as additional data. The assumption that the census coverage error is distributed symmetrically around zero could be relaxed by assigning a non-symmetric distribution to it. However a good knowledge of the shape of the distribution is required, which might be difficult in practice.

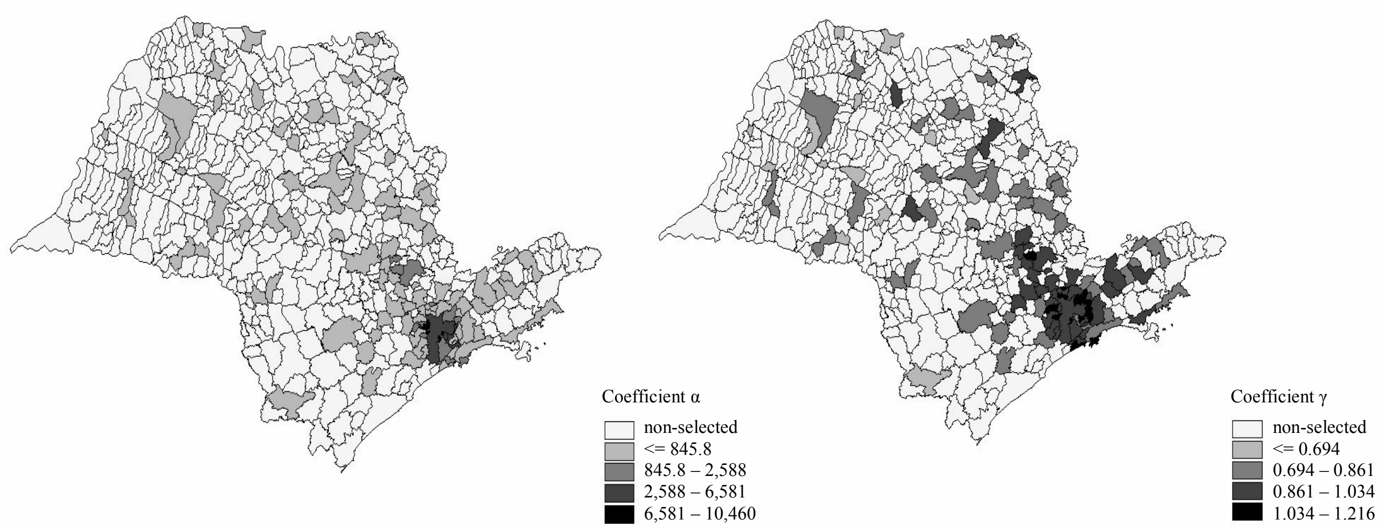

Figure 6 Posterior means of the parameters α and γ obtained by the hierarchical model

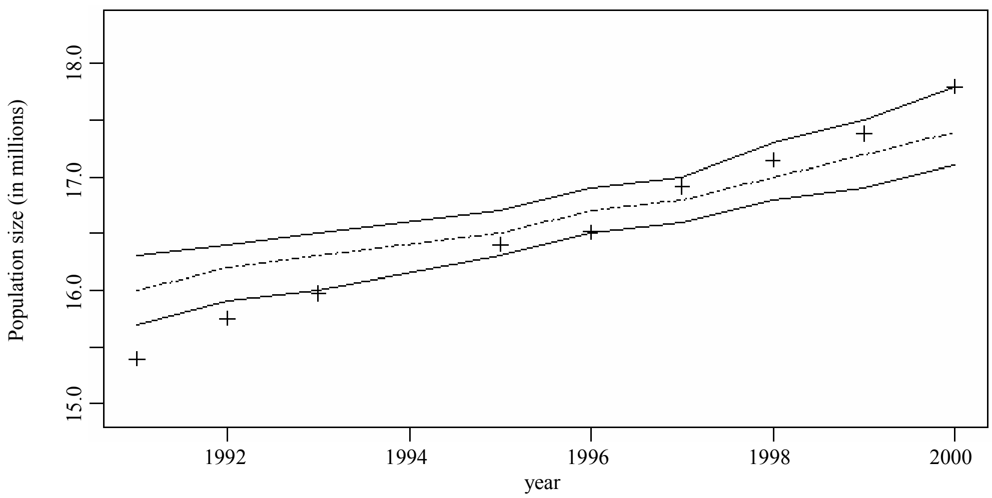

Figure 7 Comparison between the population sizes predicted by the spatial model and the official statistics (+) for the metropolitan region

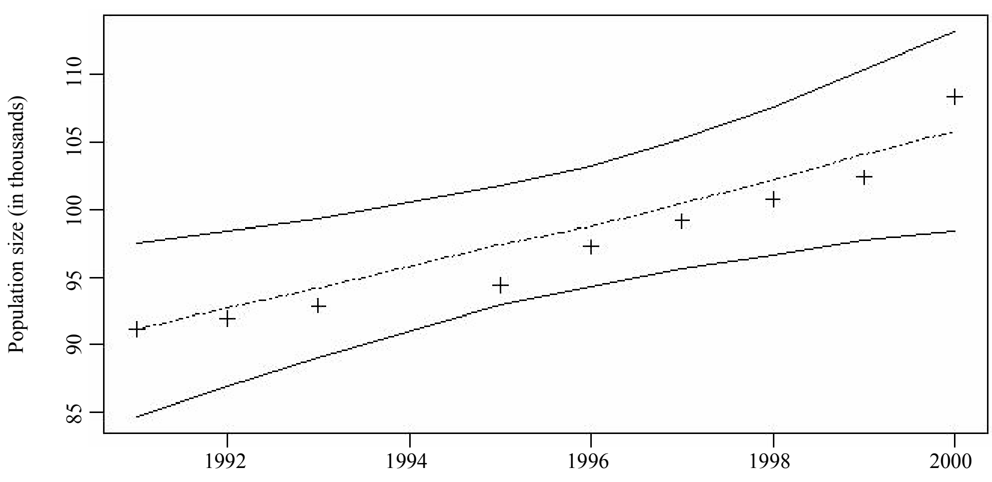

Figure 8 Comparison between the population sizes predicted by the spatial model and the official statistics (+) for a sampled municipality

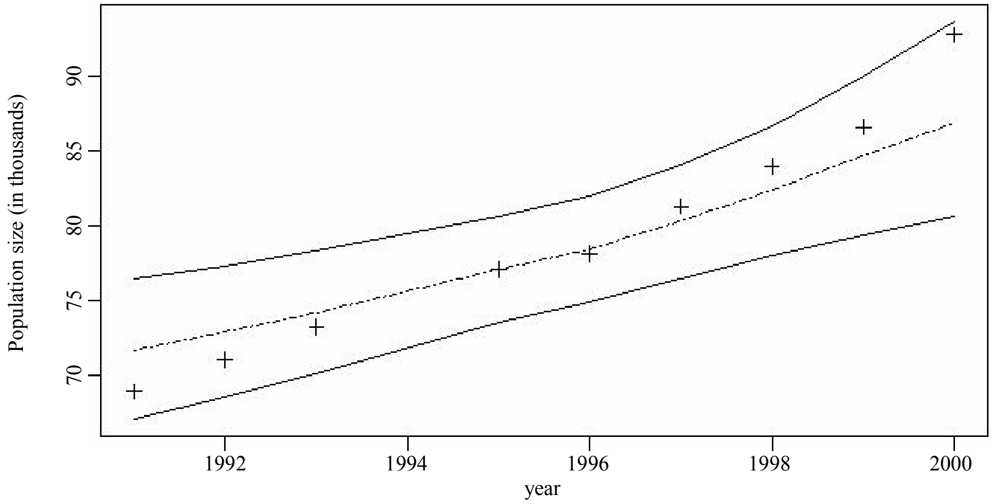

Figure 9 **Population sizes predicted by the spatial model and the official statistics (+) for a non-sampled municipality**

Acknowledgements

The authors thank an associate editor and two reviewers for their constructive comments and suggestions. The work of Fernando Moura and Helio Migon was funded in part by a research grant from the Brazilian National Council for the Development of Science and Technology (CNPq).

References

Besag, J., and Kooperang, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82, 733-746.

Brooks, S.P., and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 4, 434-455.

Gelfand, A.E., and Ghosh, S.K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85, 1, 1-11.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 3, 515-533.

Migon, H., and Gamerman, D. (1993). Generalized exponential growth models: A bayesian approach. *Journal of Forecasting*, 12, 573-584.

Mollié, A. (1996). Bayesian mapping of disease. In: Gilks, W.R.; Richardson, S.; Spieghalter, D.J. Markov Chain Monte Carlo in Practice. New York: Chapman & Hall, 359-379.

Moura, F.A.S., and Migon, H.S. (2002). Bayesian spatial models for small area estimation of proportions. *Statistical Modeling*: *An International Journal*, 2, 3, 183-201.

Pfeffermann, D. (2002). Small area estimation - New developments and directions. *International Statistical Review*, 70, 1, 125-143.

Pfeffermann, D., Moura, F.A.S. and Silva, P.L.N. (2006). Multi-level modeling under informative sampling. *Biometrika*, 93, 4, 943-959.

Pfeffermann, D., and Sverchkov, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *Journal of American Statistical Association*, 102, 1427-1439.

Rao, J.N.K. (2003). *Small Area Estimation*. Wiley Series in Survey Methodology.

Souza, D.F. (2004). Estimação de População em Nível Municipal via Modelos Hierárquicos e Espaciais. Unpublished master's dissertation. Universidade Federal do Rio de Janeiro.

Spiegelharter, D.J., Thomas, A., Best, N. and Lunn, D. (2004). WinBUGS User Manual Version 1.4. MRC Biostatistics Unit, Cambridge.