Regional fertility data analysis: 
A small area Bayesian approach

Eduardo A. Castro, Zhen Zhang, 
Arnab Bhattacharjee, José M. Martins, and Taps Maiti

Please see the SEEC website (http://seec.hw.ac.uk/) for more information.
Regional fertility data analysis: A small area Bayesian approach

Eduardo A. Castro, # Zhen Zhang, + Arnab Bhattacharjee, †
José M. Martins, # and Taps Maiti +, † *

November 22, 2013

Abstract

Accurate estimation of demographic variables such as mortality, fertility and migrations, by age groups and regions, is important for analyses and policy. However, traditional estimates based on within cohort counts are often inaccurate, particularly when the sub-populations considered are small. We use small area Bayesian statistics to develop a model for age-specific fertility rates. In turn, such small area estimation requires accurate descriptions of spatial and cross-section dependence. The proposed methodology uses spatial clustering methods to estimate an adjacency matrix that captures such dependence more adequately. The model is then used to estimate age-specific fertility rates and total fertility rates at the regional NUTS III area level for Portugal. The paper makes important contributions to small area Bayesian statistics in a spatial domain focusing on estimation of fertility rates. The estimates obtained are more accurate and adequately represent uncertainty in the estimates, and are therefore very useful for demographic policy in Portugal.

*Affiliations:* # Department of Social, Political and Territorial Sciences, University of Aveiro, Portugal; + Department of Statistics and Probability, Michigan State University, East Lansing MI, USA; † Department of Economics and Spatial Economics and Econometrics Centre (SEEC), Heriot-Watt University, UK.
Address for correspondence: Arnab Bhattacharjee, Heriot-Watt University, Spatial Economics & Econometrics Centre (SEEC), Room 1.06, Mary Burton Building, Edinburgh EH14 4AS, Scotland, United Kingdom.
E-mail: a.bhattacharjee@hw.ac.uk.

Acknowledgements: The authors acknowledge financial support provided by the research project DEMO-SPIN (PTDC/CS-DEM/100530/2008), from the Portuguese Foundation for Science and Technology, the Operational Programme ‘Thematic Factors of Competitiveness’ of the EU Community Support Framework, and from the European Regional Development Fund. The usual disclaimer applies.
1 Introduction

We use small area Bayesian statistics to develop a model for age-specific fertility rates. The model is then used to estimate age-specific fertility rates and total fertility rates at the regional NUTS III area level for Portugal. The paper makes important contributions to small area Bayesian statistics in a spatial domain focusing on estimation of fertility rates.

Micro-demographic variables such as mortality, fertility and migrations by age group are not only the main drivers of demographic dynamics but are also key elements to describe the behaviour of populations, both at national and regional levels. Both regional population forecasting and inter-regional comparisons to support policy making require the accurate data on such variables, and are strongly sensitive to the quality of estimates. In turn, data quality depends essentially on the survey and data collection processes, on the size of the analysed population and on the forecast time horizon under study. Quality issues are particularly a matter of great concern in the analyses of regional data, with small populations and relatively rare events such as births, deaths and migrations in specific age groups. This is the basic problem addressed within the literature on small area statistics as well as its application to demographic estimates and projections; see, for example, White (1954), Isserman (1977), Keyfitz (1981), Smith (1987), and Schmitt and Crossetti (1988) as some classic studies reflecting the importance of these issues in the area of demography. In turn, small area statistics involves combining evidences across space or other cohorts, and therefore requires adequate description of spatial and cross-section dependence. We develop a model for small area estimation that uses spatial clustering methods to estimate the adjacency matrix. The estimates obtained are more accurate and adequately quantify the degree of uncertainty involved in estimation. Therefore, these estimates are very useful for demographic policy in Portugal.

The remainder of the paper is organised as follows. In section 2, we start with a short overview of the literature on small areas statistics applied to demography. Particular emphasis is placed on estimation where instability arises from the size of the analysed units rather than from poor quality surveys. Section 3 presents a spatial random effects Poisson model for fertility outcomes, together with priors, Bayesian implementation and statistics for model comparison. This is followed (section 4) by Bayesian inferences on a spatial clustering model to estimate an adjacency matrix. In section 5, the above methods are applied to the estimation of age group fertility rates for Portuguese NUTS III regions. This model assumes that each particular fertility rate is a random variable with a Poisson distribution, which in turn is defined by the combination of age group fixed effects with spatially structured random effects described by a Conditional Auto-Regressive (CAR) model. Fertility rates are thus estimated using a hierarchical Bayesian methodology and a Markov Chain Monte Carlo (MCMC) computational procedure. The results show that estimates for small regions and low fertility age groups are considerably different from the observed values. Such age groups are important for several types of demographic policies, independently of their contribution
to the total number of births (TNB). However, this contribution is expected to have an increasing importance in ageing populations, where the older fertile age groups are relatively large. Therefore, the paper ends with a comparison between TNB estimates and observed values, followed by concluding comments (section 6).

2 Theoretical and Methodological Background

We start with a short review of the literature on small area statistics in demography. This is followed, by way of motivation for the current research, by a discussion of the challenges in applying existing methods to estimation of regional age-specific fertility rates in Portugal.

2.1 Smoothing Techniques Applied to Demographic Analysis

In order to obtain more efficient estimates for sub-national domains, smoothing techniques aimed either at correcting inaccurate data or at finding substitutes for missing values are attracting a growing interest in the fields of demography and planning; see, for example, Smith and Sincich (1988), Bongaarts and Bulatao (2000), Festy et al. (2002), Alkema et al. (2008), and Raftery et al. (2009). Such techniques estimate demographic variables either by using covariates, or data for individual units assumed to have a similar behaviour. In the first case, missing or inaccurate data is estimated assuming that the variables under analysis vary according to fixed effects related to a set of explanatory variables (percentage of rural population, per capita GDP, etc). These indirect estimation methods are widely used by demographers, mainly for population predictions for developing countries (Raftery, 1995; Potter et al., 2002; Alkema et al., 2011). In the second case, smoothing is performed by using any combination of demographic data for other age groups and total population in the same region and year (Retherford, 2010; Zhao and Guo, 2011), for the same age group and region in different years (Potter et al., 2010), or for areas expected to be equivalent, either because they are geographical neighbours or because they have similar socio-economic patterns (Assunção et al., 2005; Potter et al., 2010).

Even with good quality data, smoothing is essential when estimates are unstable and show excessive variation over space or population cohorts. This can occur either because the demographic behaviour is affected by qualitative changes or because the analysed areas or cohorts are too small. A good example of the first case is the rapid change in fertility rates, with heterogenous effects in the different age groups and regions, which occurred in Portugal in the final decades of the previous century; this evidence is also in line with observed variation across other developed countries (Guinnane, 2011). The second case, which is the main focus of this paper, corresponds to structural instability, independent of data accuracy and transitional changes, and is particularly important in regional statistics where the number of occurrences is reduced, because both the base population and the per capita frequency of
such occurrences are small. This structural instability cannot always be corrected for by the use of non-demographic covariates, particularly when these other covariates are affected by the same sort of problems as well. Therefore, the appropriate alternative is the application of shrinkage techniques where each particular estimate is improved by using information concerning the same variable, either in related observational units (Assunção et al., 2005) or in the overall set of estimates (Cavenaghi et al., 2004).

Borrowing strength from related observational units or overall averages is done by assuming that the variable under analysis has a given probability distribution which produces the individual observations and is described by a set of parameters $\theta$, which in turn can be generated by specific distributions described by hyper-parameters. Shrinkage, the correction of observed values taking into account the distribution from which they are generated, can be implemented through various Bayesian methods, namely: i) Hierarchical Bayesian approaches, applied to demography by Borgoni and Billari (2003) and Divino et al. (2009), where the model is described at two or more levels, the first typically at the individual unit level highlighting heterogeneity across these individuals, and the second at a broader regional or cohort level explaining the reasons for such heterogeneity; and ii) Empirical Bayesian approaches, applied by Cavenaghi et al. (2004) and Assunção et al. (2005), where the complete definition of the priors is substituted by inferences provided by observed data. See, for example, Carlin and Louis (2000) for a full description of both approaches.

When probability distributions are related to a spatial structure, a common case in demographic variables, this spatial structure must be adequately described by a model. Typically, either the conditional autoregressive model (CAR) (Besag, 1974) or the spatial autoregressive model (SAR) (Whittle, 1954) describes such spatial dependence. Observationally, these two models are very similar (Wall, 2004), even if the interpretation of the models is somewhat different. SAR models were adopted for example by Borgoni and Billari (2003) and Divino et al. (2009). In our work we use a CAR dependence structure, which is better suited for the adopted estimation methodology. First, it offers interpretation in terms of conditional distributions of fertility rates across different regions. Second, the CAR spatial model is better suited to Bayesian modeling, and better adapted for interpretation within a Bayesian model. This general overview of the application of small area estimation techniques to demography is the basis for the presentation, in the following section, of our methodology for empirical analysis.

### 2.2 Estimation of Fertility Rates for Portuguese NUTS III Regions

Now, we turn to the context of estimating fertility rates for Portuguese regions. Mainland Portugal is divided in 28 NUTS III regions with 2011 census population varying between 2.04 million inhabitants (Greater Lisbon) and slightly more than 40,000 inhabitants (Pinhal...
In the smaller regions, the annual number of births for some age groups are very small, which renders the estimates of the corresponding fertility rates unstable, thus justifying the adoption of small area estimation techniques. We apply these techniques to estimate the fertility rates for 28 regions (indexed by $i$), 7 five age groups (denoted by $j$: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, and 45-49 years old) and time $t$. The simultaneous consideration of temporal and spatial dimensions is crucial for forecasting. This is beyond the scope of the current paper and retained for future work. In this paper, we focus on spatial cross-section estimates and comparisons, setting $t$ fixed at the year 2009.

As referred to above, the adopted estimation technique would ideally use a selected set of covariates defining how fertility is affected by factors such as economic and social well-being or cultural attitudes driving the willingness to produce and raise children, factors and attitudes which are expected to explain a substantial part of heterogeneity in child-bearing behaviour. However, Portugal has been subject to a rapid process of demographic transition, which changed drastically the fertility rates of all age groups and its spatial distribution (Castro et al., 2012). Births are now concentrated in the age groups between 25 and 35 years old, being very low in the extreme age groups, while the highest fertilities moved from the rural areas and northern regions to the urban regions. This means that the spatial effect of the covariates changed over time, following a process yet to be stabilised, which renders use of cross-section covariates problematic.

In the absence of suitable covariates, we assume that each particular value $\hat{Y}_{ijt}$ (the observed fertility rate) is a singular event determined by a given probability distribution, and subject to a statistical error term with both a structural and an idiosyncratic element. The structural element, in turn, is the outcome of fixed effects (the age group specific fertility rate) and random effects arising from spatially structured geographical heterogeneity. In other words, it is assumed that the error term in each individual estimate reflects heterogeneity and has a pattern of autocorrelation with the neighbour regions defined by a CAR model. We assume that the Poisson distribution as the chosen model to describe the observed pattern of fertility rate variation. Though the binomial distribution is an alternative option often adopted in demographic studies (see, for example, Borgoni and Billari, 2003), it is not recommended in our particular case, where we deal with events with very low frequency.

The importance of the simultaneous consideration of age effects and spatially autocorrelated heterogeneity is highlighted by the highly significant values of the Moran’s $I$ measure of spatial autocorrelation (Moran, 1950) index for the regional values of fertility rates in six of the seven age groups (all but $j = 7$). The spatial patterns of observed fertility rates across the territory of Portugal for the different age groups (Figure 1) shows also the same evidence. At the same time, fertility rates across the different age groups are also considerably different; hence the disregard of age fixed effects would generate serious errors.

The parameters defining the Poisson distribution and the respective hyperparameters which define the fixed effects and the CAR model are estimated using the Hierarchical
Bayesian methodology and a Markov chain Monte Carlo (MCMC) posterior sampling algorithm. The MCMC is implemented by running 3 chains each with 6,000 iterations. In addition to the estimation of fertility rates for all the age groups it is useful to estimate the total number of births for each region and to check how much they differ from the observed values. For this we run a Monte Carlo simulation for each region where all the vectors of fertility rates obtained for the iterations are multiplied by the vector of female population for each age group, providing thereby estimates of the total number of births. A full description of the statistical model is presented in the following section.

3 Spatial Random-effects Poisson model

3.1 Model Specification

Suppose we have data containing count response $Y_{ij}$ observed for $i = 1, 2, \ldots, N$ sites and $j = 1, 2, \ldots, J$ groups. Let $E_{ij}$ be the expected value, or the population size. Consider the Spatial Random-effect Poisson model (SRP)

$$Y_{ij} \sim \text{Poisson}(E_{ij} \xi_{ij})$$

$$\log(\xi_{ij}) = \eta_{ij} \sim \mathcal{N}(\mu + \alpha_i + \theta_j, \delta^2)$$

where we treat the group effect $\theta_j = (\theta_1, \theta_2, \ldots, \theta_J)'$ as fixed with $\theta_1 \equiv 0$ since the grand mean $\mu$ is included, and assume the site-specific random effect $\alpha = (\alpha_1, \cdots, \alpha_N)'$ admits the Conditional Autoregressive (CAR) structure

$$\alpha \sim \mathcal{N}(0, \tau^2 D(\gamma)), \quad D(\gamma) = (I - \gamma M W)^{-1} M.$$ 

$W$ is the adjacency matrix with $w_{ij} = 1$ if site $i$ and site $j$ are neighbors and 0 otherwise, with $w_{ii} \equiv 0$ for $1 \leq i, j \leq N$. For each site $i$, define $w_i+ \equiv \sum_{j=1}^{N} w_{ij}$ to be the sum that represents the total number of neighbors of site $i$. $M$ is the diagonal matrix with $m_{ii} = 1/w_i+$. 

Figure 1: Observed Fertility Rate by quinquennial age group $j$ ($j = 1, \ldots, 7$, ages 15 to 49 year old)
Let \( \eta = (\eta_1, \cdots, \eta_J) \) where each \( \eta_j \) is the \( N \times 1 \) response for \( j \)-th group. We employ \( 1_{m,n} \) to denote a \( m \times n \) matrix with all entries 1, and \( I_m \) a \( m \times m \) identity matrix. Accordingly, letting \( \theta = (\mu, \theta_2, \cdots, \theta_J)' \) with corresponding design matrix \( X \), and \( Z = 1_{J,1} \otimes I_N \), we can write the model as

\[
\eta \sim \mathcal{N}(X\theta + Z\alpha, \delta^2 I_{NJ})
\]  

(4)

To perform Bayesian inference for the parameters, we choose the following non-informative priors to allow the posterior to be more data driven:

\[
\pi_0(\theta_j) \propto 1, \quad \pi_2(\gamma) = \text{Uniform}(\lambda_1^{-1}, \lambda_2^{-1}), \quad \pi_1(\delta^2) \propto \delta^{-2}, \quad \pi_1(\tau^2) \propto \tau^{-2}
\]  

(5)

where \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \) are ordered eigenvalues of \( MW \). The condition \( \gamma \in (\lambda_1^{-1}, 1) \) is required to ensure positive definiteness of \( D(\gamma) \). The upper limit of the interval is one since the row sum of \( MW \) is one and \( \lambda_N^{-1} < 0 \) since the trace of \( MW \) is zero.

### 3.2 Implementation

We obtain the posterior samples of the parameters \{\( \theta, \alpha, \delta^2, \tau^2, \gamma \)\} via Gibbs sampling. Under the choice of prior in (5), the conditional distributions of fixed-effects and random-effects are

\[
\pi(\theta|\cdots) = \mathcal{N}(\mu_{\theta}, \Sigma_{\theta}) \quad \left\{ \begin{array}{l}
\Sigma_{\theta} = \delta^2 (X^TX)^{-1} \\
\mu_{\theta} = (X^TX)^{-1}X^T(\eta - Z\alpha)
\end{array} \right.
\]  

(6)

\[
\pi(\alpha|\cdots) = \mathcal{N}(\mu_{\alpha}, \Sigma_{\alpha}) \quad \left\{ \begin{array}{l}
\Sigma_{\alpha} = (\delta^{-2}JM + \tau^{-2}D(\gamma)^{-1})^{-1} \\
\mu_{\alpha} = \delta^{-2}\Sigma_{\alpha}Z^T(\eta - X\theta)
\end{array} \right.
\]  

(7)

The conditional distributions of variance components and spatial dependence are

\[
\begin{align*}
\pi(\delta^2 | \cdots) &= \text{igamma}(NJ/2, \epsilon^T\epsilon/2), \text{ with } \epsilon = \eta - X\theta - Z\alpha \\
\pi(\tau^2 | \cdots) &= \text{igamma}(N/2, \alpha^TD(\gamma)^{-1}\alpha/2) \\
\pi(\gamma | \cdots) &\propto |D(\gamma)|^{-1/2} \exp\{\gamma\alpha^TW\alpha/(2\tau^2)\} \cdot I(\gamma \in (\lambda_1^{-1}, \lambda_N^{-1}))
\end{align*}
\]

where \( \text{igamma}(a, b) \) denotes the Inverse Gamma density with shape \( a \) and scale \( b \), i.e., \( p(x) \propto x^{-a-1} \exp(-b/x) \). Note that the full conditional distribution of \( \gamma \) is not known but can be sampled by numerically evaluating the density over its support.

Finally, assume a clustering partition \( \cup_{r=1}^d C_r \) with corresponding means \( \mu_1, \mu_2, \cdots, \mu_d \). To update \( \eta_{ij} \) for \( i \in C_r \) and \( j = 1, 2, \cdots, J \), we have

\[
\log(\eta_{ij} | \cdots) = \text{Const.} - \eta_{ij}^2/(2\delta^2) + \eta_{ij}(Y_{ij} + (\mu_c + \theta_j)/\delta^2) - E_{ij} \exp\{\eta_{ij}\}
\]
3.3 Model Comparisons

To compare models and validate the importance of the spatial models, we use Deviance Information Criterion (DIC) for mixed-effects model, DIC\(_4\) (Celeux et al 2006) based on complete likelihood

\[
\text{DIC}_4 = -4E_{\theta, \alpha} \left[ \log f(Y, \alpha|\theta) \right] + 2E_{\alpha} \left[ \log f(Y, \alpha|E_{\theta}[\theta|Y, \alpha]) \right] Y
\]

\[\triangleq -4E_1 + 2E_2,
\]

where \(E_2 = E_{\theta}[\theta|Y, W]\) can be evaluated by sampling \(\theta\) for each posterior sample of \(W\) and obtain the mean. We also report \(D(\theta) = -2E_1\) as the posterior expected value of the joint deviance, and \(pD_4 = D(\theta) + 2E_2\) as a measure of model dimensionality. Generally, a smaller DIC\(_4\) indicates better predictive power.

4 Spatial Clustering Model for Estimating the Adjacency Matrix

The above method is based on a spatial clustering model that we use to estimate our adjacency matrix.

4.1 Model Specification

The spatial dependence \(\gamma\) is modeled to improve the predictive power, and can rely on the definition of the adjacency matrix \(W\). For small area data where sampling unit sites are represented as polygons, a natural choice is to define \(w_{ij} = 1\) if site \(i\) and site \(j\) share overlapping boundary, and 0 otherwise. However it is possible to estimate \(W\) from the data, using spatial clustering techniques, and this can potentially further improve the prediction. Specifically, consider a spatial clustering configuration \(\varpi = (d, G_d)\), where \(d\) is the number of clusters, \(G_d = (g_1, \cdots, g_d)\) are the cluster centers, and the memberships of the \(N\) sites are determined according to minimal distance criterion based on a distance measure \(D(i, j)\) between each pair of sites. We choose \(D(i, j)\) to be the smallest number of boundaries crosses is going from from site \(i\) to site \(j\).

Under the clustering partition \(\bigcup_{r=1}^d C_r\) from \(\varpi\), each cluster \(C_r\) with \(n_r\) member sites and \(\sum_{r=1}^d n_r = N\), consider the Spatial Clustering Poisson Model (SCPM) for each \(i \in C_r, r = 1, \cdots, d\) that

\[
Y_{ij} \sim \text{Poisson}(E_{ij}\xi_{ij}) \quad \text{(8)}
\]

\[
\log(\xi_{ij}) = \eta_{ij} \sim \mathcal{N}(\mu_r + \theta_j, \delta^2) \quad \text{(9)}
\]

for \(\theta_1 \equiv 0\). Comparing with the Spatial Random-effect Poisson model (SRP) in (1), since the adjacency matrix is considered as unobserved and the spatial dependence mechanism is
unknown, we cluster the $\mu_i = \mu + \alpha_i$ into $\mu_r$ that is shared by all members in cluster $C_r$. Letting $\mu = (\mu_1, \cdots, \mu_d)$ and $\theta = (\theta_2, \cdots, \theta_J)$, the parameters associated with SCPM are \{\varpi, \mu, \theta, \eta, \delta^2\}.

Next, we consider prior elicitation. For the spatial clustering configuration $\varpi$, a prior model considered by Knorr-Held and Raßer (2000) is $\pi(\varpi) = \pi(G_d|d)\pi(d|\kappa)\pi(\kappa)$, with $\pi(d|\kappa) \propto (1 - \kappa)^{d-1}$ for $d = 1, \cdots, N_0 < N$ and $\pi(\kappa) \sim \text{Unif}[0, 1]$. Here, the hyperparameter $\kappa$ controls the penalty of model complexity for large $d$. The prior for $d$ receives power decay from this elicitation and when $\kappa$ is larger, then $d$ is more likely to be smaller, while when it is fixed at a very small value, $d$ is almost uniformly distributed over \{1, 2, \cdots, N_0\} to indicates weak prior information. However, when the clustering parameters have high dimensionality, a power decay can be insufficient.

Instead, we consider a exponential decay: $\pi(d|\kappa) \propto \exp(-\kappa d)$ and when $\kappa$ is fixed to be (# of parameters under $\varpi$) $\log(NJ)/2$, it is closely related to the Bayesian Information Criterion (BIC). To allow the posterior of $d$ to be more data driven, we fix $\kappa$ at 0 such that the prior for $d$ is non-informative and all possible $d$’s receive equal probability. Conditional on $d$, we again consider the non-informative prior for centers $\pi(G_d|d) = (N - d)!/N!$ where the $(\binom{N}{d})d!$ possible $G_d$’s receive equal probability of being the cluster centers. We then elicit non-informative prior for the remaining parameters for $r = 1, \cdots, d$ and $j = 2, \cdots, J$ as:

$$
\pi_0(\mu_r|\varpi) \propto 1, \quad \pi_1(\theta_j) \propto 1, \quad \pi_2(\delta^2) \propto \delta^{-2}.
$$

### 4.2 Implementation

To obtain posterior samples for \{\varpi, \mu, \theta, \eta, \delta^2\}, the main challenge is that the model dimensionality changes when the clustering configuration $\varpi$ is updated. We therefore construct the reversible jump MCMC (Green, 1995) updating procedure for (\varpi, \mu, \delta^2) given (\theta, \eta), and then update (\theta, \eta) from the full conditional distribution.

The steps of the algorithm are described as follows.

(i) **Update** (\varpi, \mu, \delta^2): Let the $n_rJ$-vector $\Delta_r = (\Delta_{ij})_{i \in C_r, 1 \leq j \leq J}$ with $\Delta_{ij} = \eta_{ij} - \theta_j$, and the $NJ$-vector $\Delta = (\Delta_r)_{1 \leq r \leq d}$. Also let $\Delta_r$ be the first moment of $\Delta_r$. From the current state (\varpi, \mu, \delta^2), we propose a new state (\varpi*, \mu*, \delta^2*) where \varpi* vary from \varpi, and define the auxiliary variable $U = \psi^* = (\mu^*, \delta^2*)$ and let $\psi = (\mu, \delta^2) = U^*$, the corresponding invertible map $q : (\psi, U)$ to ($\psi^*, U^*$) is one-to-one with Jacobian 1. We propose $U \sim h(\|\varpi, \mu, \varpi^\ast\)$. We accept \varpi* with probability

$$
\min \left\{1, \frac{g(\varpi^\ast|\varpi^\ast) \times \pi(\varpi^\ast, \psi^\ast|\Delta) \times h(u^\ast|\varpi^\ast, \psi^\ast, \varpi^\ast)}{g(\varpi^\ast|\varpi) \times \pi(\varpi, \psi|\Delta) \times h(u|\varpi, \psi, \varpi^\ast)} \times 1 \right\}
$$

We choose the full conditional distribution of $\psi^\ast$ as the proposal, i.e., $h(u|\varpi, \mu, \varpi^\ast) = \pi(\mu^\ast, \delta^2^\ast|\varpi^\ast, \Delta) = \pi(\delta^2^\ast|\varpi^\ast, \Delta) \times \pi(\mu^\ast|\varpi^\ast, \delta^2^\ast, \Delta)$. Specifically, we first generate $\delta^2^\ast$ from

$$
\pi(\delta^2^\ast|\varpi^\ast, \Delta) \propto \pi_2(\delta^2^\ast) \int \pi(\Delta|\mu, \delta^2^\ast) \pi_0(\mu) \, d\mu
$$

(11)
which is an inverse-Gamma density with shape \((N - d)J/2\) and scale \(\sum_{r=1}^{d} \sum_{i \in C_r,j} (\Delta_{ij} - \bar{\Delta}_r)^2 / 2\). Next, we generate \(\mu^*\) from the full conditional distribution given \(\delta^2\):

\[
\pi(\mu^* | \varpi^*, \delta^2, \Delta_r) \sim N(\bar{\Delta}_r, (n_r J)^{-1} \delta^2^*) , \quad r = 1, 2, \ldots, d
\]

(12)

Under this choice of proposal density, letting \(m(\Delta)\) be the normalizing constant which is finite by the posterior propriety that is evident from the following context, we can substitute

\[
\pi(\varpi, \psi | \Delta) = \pi(\Delta | \varpi, \mu, \delta^2) \pi(\mu, \delta^2 | \varpi, \Delta) \pi(\varpi) / m(\Delta)
\]

into (10), and the Metropolis-Hasting ratio in (10) becomes

\[
\frac{g(\varpi|\varpi^*)}{g(\varpi^*|\varpi)} \times \frac{\pi(\Delta | \varpi^*, \mu^*, \delta^2^*) \pi(\mu^*, \delta^2^* | \varpi^*) \pi(\mu, \delta^2 | \varpi, \Delta)}{\pi(\Delta | \varpi, \mu, \delta^2) \pi(\mu, \delta^2 | \varpi, \Delta)} \times \frac{\pi(\varpi^*)}{\pi(\varpi)}
\]

(13)

Using the fact that

\[
\pi(\Delta | \varpi) = \frac{\pi(\Delta | \varpi, \mu, \delta^2) \pi(\mu, \delta^2 | \varpi)}{\pi(\mu, \delta^2 | \varpi, \Delta)}
\]

the ratio (13) reduces to

\[
\frac{g(\varpi|\varpi^*)}{g(\varpi^*|\varpi)} \times \frac{\pi(\Delta | \varpi^*)}{\pi(\Delta | \varpi)} \times \frac{\pi(\varpi^*)}{\pi(\varpi)}
\]

(14)

When the prior and proposal density of \(\varpi\) are non-informative, the acceptance rate is dominated by the marginal likelihood ratio \(r = \pi(\Delta | \varpi^*) / \pi(\Delta | \varpi)\) with

\[
\pi(\Delta | \varpi) = \int \left( \prod_{r=1}^{d} \int \pi(\Delta_r | \delta^2, \mu_r) \pi_0(\mu_r | \varpi) \pi_2(\delta^2) d\mu_r \right) d\delta^2
\]

(15)

Letting \(n^* = (N - d)J/2\), we can simplify it into

\[
\log(\pi(\Delta | \varpi)) = -n^* \log(\pi) - \frac{1}{2} \sum_{r=1}^{d} \log(n_r J) + \log(\Gamma(n^*) - n^* \log \sum_{r=1}^{d} \sum_{i \in C_{r,j}} (\Delta_{ij} - \bar{\Delta}_r)^2
\]

(16)

The posterior propriety for \(\Delta\) immediately follows that \(\pi(\Delta | \varpi) < \infty\) for \(d < N\) and the collection of all \(\varpi^*\)'s is finite.

Next, we specify the uniform proposal density \(g(.)\) for \(\varpi^*\) given current clustering configuration \(\varpi\) by searching its neighbor point in the parameter space of \(\varpi\), using the proposals in Knorr-Held and Raßer (2000). We consider the new \(\varpi^*\) under one of the following scenarios according to different move of \(d\), with the respective proposal probability \(P(\text{Growth}) = P(\text{Merge}) = 0.4\) and \(P(\text{Shift}) = 0.2\).

1. Growth step \((d, G_d) \rightarrow (d + 1, G_{d+1})\): We create a new cluster. We first draw a random variable uniformly distributed on the \(N - d\) non-center sites, to determine the new cluster \(C^*\) with center \(g^*\). Secondly, we draw another random variable \(r\) uniformly
distributed on \( \{1, \ldots, d+1\} \) to determine the position of \( g^* \) in \( G^*_{d+1} \). The \( n^*_\lambda \) sites that have minimal distance from \( g^* \) then automatically enter \( C^* \). In this case,

\[
g(\varpi|\varpi^*) = \frac{P(\text{Merge})(N-d)}{P(\text{Growth})}, \quad \pi(\varpi^*) = \frac{\exp(-(d+1)\kappa)(N-d-1)!}{\exp(-d\kappa)(N-d)!}
\]

and the acceptance rate is \( \min\{1, \exp(-\kappa) \times \frac{P(\text{Merge})}{P(\text{Growth})} \times r\} \).

2. Merge step \((d+1, G_{d+1}) \rightarrow (d, G^*_d)\): We delete one existing cluster and merge its members into other existing clusters. First, generate a random variable uniformly \( r \) distributed on \( \{1, \ldots, d+1\} \), which determines the cluster \( C_r \) with center \( g_r \) to be removed with all its members merging into one of the remaining clusters by the minimal distance criterion. The acceptance is the reciprocal of that in growth step.

3. Shift Step: \((d, G_d) \rightarrow (d, G^*_d)\): We adopt a shift step for moving one cluster center to its non-center neighborhood when \( d \) is invariant. For each site \( s \), we define all sites that are directly connected (i.e., via no third site) to it by latitude or longitude as its neighbors. Among \( d \) current cluster centers there are \( n(G_d) \) cluster centers that have at least one non-center neighbors. Draw \( r \sim \text{Uniform}\{1, \ldots, n(G_d)\} \) to obtain one such cluster center \( g_r \) with \( m(g_r) \) non-center neighbors. Secondly, draw \( l \) from \( \{1, \ldots, m(g_r)\} \) uniformly. The \( l \)th non-center neighbor becomes the new cluster center \( g^*_r \) that replaces \( g_r \) in \( G_d \). The acceptance probability is \( \min\{1, \frac{n(G_d)\min\{m(g_r)\}}{\max\{m(g_r)\} \times \pi(\varpi^*)} \}

(ii) Update \((\theta, \eta)\): The full conditional distribution of \( \theta_j \) is \( \mathcal{N}\left(\frac{\sum_{i=1}^{N}(\eta_{ij} - \mu_r)}{N}, \sigma^2/N\right) \) for \( i \in C_r, j = 2, \ldots, J \). Finally, to update \( \eta_{ij} \) for \( i \in C_r \) and \( j = 1, 2, \ldots, J \), we have

\[
\log(\eta_{ij} | \cdots) = \text{Const.} - \frac{\eta^2_{ij}}{(2\sigma^2)} + \eta_{ij}(Y_{ij} + (\mu_r + \theta_j)/\sigma^2) - E_{ij} \exp(\eta_{ij})
\]

### 4.3 Posterior Inference

To estimate the adjacency matrix \( \hat{W} \), once we obtain posterior samples with size \( B \), consider \( w_b(i, j) = 1 \) if site \( i \) and \( j \) share the same clustering membership in the \( b \)-th sample, and 0 otherwise. Let \( w_{ij} = \sum_{b=1}^{B} w_b(i, j)/B \in [0, 1] \) and \( w_{ii} \equiv 0 \) be the estimated adjacency between site \( i \) and \( j \). We can then fit SRP with the estimate \( \hat{W} = (w_{ij})_{1 \leq i,j \leq N} \). To further obtain a central clustering configuration \( \varpi^* \) from the posterior samples \( \varpi^1, \cdots, \varpi^B \), consider the dissimilarity measure \( Diss(i, j) = 1 - w_{ij} \) based on which an agglomerative clustering algorithm is performed with number of cluster to be the posterior mode of \( d \).

### 5 Application and Posterior Inference

For application of the above methods to fertility data from Portugal, we treat the group \( E_1 \) as the baseline group and compare the remaining 6 groups, \( j = 2, \cdots, J \) to its mean.
level. SRP with $\gamma = 0$ (non-spatial) and $\gamma$ updated (spatial) is implemented by running 3 Monte Carlo Markov Chains (MCMC), each with 6,000 iterations. The convergence is well committed after 5,000 iterations in that the Potential Scale Reduction Factor $\sqrt{R} < 1.2$ for all parameters (Brooks and Gelman, 1998). The final 1,000 samples for each chain are then used as posterior samples.

Next, we fit SCPM using 4 MCMC runs with a total of 55,000 iterations, and sample every 5-th iteration of the last 5,000 samples after convergence, and obtain a total of 4,000 posterior samples for analysis. The convergence is validated using the criterion $\sqrt{R} < 1.2$ by monitoring $(\delta^2, \theta, \eta)$, the Gaussian likelihood $f(\eta|\mu, \theta, \delta^2)$ and the marginal likelihood $\pi(\Delta|\omega)$ in (16). The posterior distribution of $d$ turns out to be highly concentrated on $d = 2$ with probability 0.9985, and the central clustering configuration based on $Diss(i, j)$ is shown in Figure 2. We then re-fit SRP using estimated $\hat{W}$ with the same setup as all other SRP implementations.

The parameter estimates and model assessment for SRP with non-spatial case, spatial case with the natural adjacency matrix $W$, and spatial case with estimated adjacency matrix $\hat{W}$ from SCPM, are summarized in Table 1.

From the output the spatial dependence parameter $\gamma$ is highly significant with credible interval deviated from 0. The spatial case of SRP with estimated $\hat{W}$ from SCPM turns out to have the smallest DIC$_4$, and is hence preferable. Consequently, the histograms of the 3,000 posterior samples of parameters from the most favorable model is shown in Figure 2. We also plot the estimated mean fertility rate $\xi_{ij}$’s by groups and regions in Figure 3, and the comparison with observed $\xi_{ij} = Y_{ij}/E_{ij}$ in Figure 5.

6 Final Results and Conclusions

Figures 4 and 5 show that fertility in Portugal is heavily concentrated between 25 and 35 years old age groups, the corresponding fertility rates being higher and spatially homogeneous.
Figure 3: Histogram of the posterior samples for SRP with estimated $\hat{W}$ from SCPM.

Figure 4: Estimated Fertility Rate by quinquennial age group $\phi_j$ ($j = 1, \ldots, 7$, ages 15 to 49 year old).
Figure 5: Observed vs. Estimated Fertility Rate by quinquenial age group.
Another remarkable feature of Portuguese fertility landscape is the incidence of higher rates both in rural and poorer northern regions (Tâmega) and the more urbanized and richer areas around Lisbon (Península de Setubal, Grande Lisboa). This means that the demographic transition, moving higher fertility from north to south and from rural to urban places, is still occurring. As we argued above, this hinders the use of covariates to effectively explain fertility spatial heterogeneity, and justifies our methodological approach.

Finally, it is interesting to see how variation in age specific fertility groups generates variation in the total number of births and total fertility rates. The observed and estimated total number of births and fertility rates are shown in Figure 6 and Figure 7, respectively. The close correspondence between observed and estimated total number of births across all regions highlights the fit achieved by our modeling approach. Likewise the mean estimated
fertility rates in each region is also in close alignment with the corresponding observed rates, even if the prediction intervals indicate substantial uncertainty and spatial variation.

In summary, our proposed model and methodology is effective in estimation of age-specific regional fertility rates, using small area estimation methods to smooth out large instability and capture the spatial structure in an appropriate way. The estimates obtained thereby are in good alignment with observed values, at the same time as they are robust to sampling fluctuations due to small sample sizes in some regions and age groups.

References


Figure 7: Observed vs. Estimated total fertility rates.


criteria for missing data models. *Bayesian Analysis* 1(4), 651-673.

Italy: A Bayesian analysis. *Demographic Research* 20(18), 435-466.


criteria for missing data models. *Bayesian Analysis* 1(4), 651-673.


17–23.


