Spatio-Temporal Modeling and Spatial Clustering of Curves

Spatio-Temporal Patterns in Portuguese Regional Fertility Rates: A Bayesian Approach for Spatial Clustering of Curves

Arnab Bhattacharjee,[$] Eduardo A. Castro,[&] Tapabrata Maiti[+,$] and Zhen Zhang[#] *

Abstract

It is important for demographic analyses and policy-making to obtain accurate models of spatial diffusion, so that policy experiments can reflect endogenous spatial spillovers appropriately. Likewise, it is important to obtain accurate estimates and forecasts of demographic variables such as age-specific fertility rates, by regions and over time, as well as the uncertainty associated with such estimation. Here, we consider Bayesian hierarchical models with separable spatio-temporal dependence structure that can be estimated by borrowing strength from neighbouring regions and all years. Further, we do not consider the adjacency structure as a given, but rather as an object of inference. For this purpose, we use the local similarity of temporal patterns by developing a spatial clustering model based on Bayesian nonparametric smoothing techniques. The Bayesian inference provides the uncertainty associated with the clustering configurations which is typically lacking in classical analyses of large data sets where a unique clustering representation can be insufficient. The proposed model is applied to 16-year data on age-specific fertility rates observed over 28 regions in Portugal, and provides statistical inference on the number of clusters, and local scaling and shrinkage levels. The corresponding central clustering configuration is able to capture spatial diffusion that have key demographic interpretations. Importantly, the exercise aids identification of peripheral regions with poor demographic prospects and development of regional policy for such places.

Keywords: Spatio-temporal modeling; Conditional Autoregressive Model; Spatial clustering; Bayesian Wavelet Smoothing; Bayesian Hierarchical Model; Age-specific fertility rates.
JEL Classification: C23, J11, C11, C14, C53.

*Correspondence: A. Bhattacharjee, Heriot-Watt University, Spatial Economics and Econometrics Centre (SEEC), Room 1.06, Mary Burton Building, Edinburgh EH14 4AS, Scotland, United Kingdom. E-mail: a.bhattacharjee@hw.ac.uk.

Affiliations: [$] Spatial Economics & Econometrics Centre (SEEC), Heriot-Watt University, U.K.; [&] Department of Social, Political & Territorial Sciences, University of Aveiro, Portugal; [+] Department of Statistics & Probability, Michigan State University, East Lansing, MI, U.S.A.; and [#] Department of Statistics, University of Chicago, Chicago IL, U.S.A.

"[D]emographers and population scientists were pioneers in the study of neighborhoods and health ... Putting people into place means explaining behavior and outcomes in relation to a potentially changing local context. A more dynamic conceptualization is needed that fully incorporates human agency, integrates multiple dimensions of local social and spatial context, develops the necessary longitudinal data, and implements appropriate tools" (Entwisle, 2007:687).

1 Introduction

The above quote highlights the key idea behind this paper – to emphasize spatial and temporal dynamics as key elements of demographic study. The complexity of spatio-temporal patterns is perhaps most evident in fertility, resulting as it were from a combination of fertility behavior and health outcomes, both of which are spatially contingent and exhibit large spatio-temporal diffusion. Further, fertility behavior is endogenously related to regional economic performance, being both affected by the economy and influencing the economy, and the regional economy in turn exhibits substantial spatial concentration. However, quantitative study of spatio-temporal dynamics in fertility is challenging, and empirical examination has been lacking. We develop a new framework and methods for spatial clustering of time trends to estimate the spatio-temporal pattern of fertility across small areas or sub-national regions. The results are found useful for understanding spatial diffusion and for regional demographic and economic policies for the peripheral regions of Portugal.

The paper complements and extends current quantitative research in demography along several dimensions. First, the literature acknowledges substantial spatial diffusion in fertility; see, for example, Tolnay (1995) and Weeks et al. (2000). Long run dynamics in fertility have traditionally been explained by the "structural" demographic transition model (Thompson, 1929; Notestein, 1945; Blacker, 1947), describing the transition from high towards low birth and death rates as part of the economic development of a country or region. Under this theoretical premise, macrostructural characteristics of society – such as urbanization, industrialization, and education – determine the demand for children by influencing their economic value to their households. However, following evidence uncovered by the Princeton European Fertility Project (Coale and Watkins, 1986), the "diffusion" perspective on variation and change in fertility levels, emphasizing cultural influences on fertility, has become prominent. Specifically, European fertility transition has been more temporally concentrated and less concentrated in space than what would be expected from a strictly structural explanation of fertility decline, and further, changes in important structural factors such as literacy and industrialization are found only to be weakly related to the timing of fertility decline.

Likewise, Tolnay (1995) found evidence of spatial patterning in twentieth-century US fertility, which is related to diffusion of reproductive patterns implied by endogenous social network formation. Two sources of such diffusion are commonly discussed: (a) information diffusion for example on birth control measures; and (b) diffusion of ideas relating to marriage, family life, and sexuality through social communication leading to an ideational shift. This places societal culture at the cen-

tre of demographic study of fertility trends and changes: *"belief systems (values, norms, language, religion, ideologies), cultural traditionalism, cultural homogeneity, and socialization of individual actors"* (Wejnert, 2002). Together, spatial mechanisms such as proximity and diffusion have also been used to explain spatial patterns in contraceptive choice (Entwisle et al., 1997), crime and violence (Tolnay et al., 1996; Morenoff and Sampson, 1997), religious practise (Land et al., 1991) and poverty and inequality (Lobao and Saenz, 2002; Gundersen and Ziliak, 2004; Flippen, 2010). Such spatial patterns have important implications for policy (Case et al., 2003; Gundersen and Ziliak, 2004; de Castro, 2007). Arguably, one of the main reasons that family planning programs have not had their expected impact on fertility levels is that political decisions, and not spatial analysis, typically determine the location of pre-natal and reproductive clinics (Fuller, 1974). However, it may be noted that, while discussion has largely focused on diffusion, similar patterns can equally result from the spatial concentration of macrostructural characteristics. Indeed, it has been observed that inclusion of structural variables reduces the power of spatial effects in predicting diffusion (Berry and Berry, 1992).

More importantly in the context of this paper, the recent literature has modeled diffusion effects as a function of geographical proximity or distance between social units. At the same time, given the very nature of diffusion, its natural driver should be socio-cultural distances, not geographic proximity. In fact, fertility patterns in Europe follow religious and linguistic contours (Coale and Watkins, 1986), and therefore may vary significantly over geographically proximate countries or regions. Perhaps one reason for lack of focus on socio-cultural distances is that they are more difficult to measure. This paper contributes to a recent strand within the literature in spatial econometrics and statistics that treats such drivers of spatial diffusion as unknown parameters and has developed methods for inference on socio-cultural diffusion; see, for example, Bhattacharjee and Holly (2013), Bhattacharjee and Jensen-Butler (2013) and Bailey et al. (2015). Based on such a perspective, estimated diffusion often reflects drivers that may not be closely related to geographic distances and contiguity, but socio-cultural, ideational or core-periphery relations emerge as equally important influences. In this paper, we extend the above literature in two ways. First, we develop a new framework and methods to conduct Bayesian inferences on an unobserved contiguity (adjacency) matrix which in turn drives spatial diffusion in fertility. This allows explanations of diffusion to extend beyond the restrictive constraint of geographic distances. Second, we consider not only spatial diffusion, but spatio-temporal dynamics. This is very important in the current context, since previous literature has highlighted the fact that the onset of demographic transition in different places can be different, even if they are geographically proximate, and even if their macrostructural caharacteristics are relatively similar (Coale and Watkins, 1986; Watkins, 1987).

A second relevant line of research relates to the influential literature, starting from Seiver (1985), that explores seasonal variation in fertility and the spatial clustering of such seasonal patterns. While within-the-year seasonal variation in fertility has declined over the past 50 years (Seiver, 1985; Lam and Mirron, 1991), the key insight from the above research, that temporal patterns in fertility exhibit significant clustering, is equally prominent over longer periods of time. This is particularly

true in advanced countries, where fertlity has steadily declined as these societies progressed through the third phase of demographic transition, and these declining trends are significantly spatially clustered. In this paper, we develop methodology for spatial clustering of time-trends in age-specific fertility rates over longer periods of time.

A third strand of the current literature has focussed on small area statistical methods to obtain improved estimates and projection for micro-demographic variables such as fertility and migration. Forecasts of population and its micro-demographic components, notably fertility, and quantification of the uncertainty surrounding such projection, are very important for policy. At the same time, it is extremely challenging to obtain reliable forecasts for such quantities (Raftery et al., 2009). In particular, accurately estimating regional fertility rates by age groups is difficult not only for less-developed countries, where there are often problems of the scarcity and quality of demographic data, but also for more-developed countries, particularly at a fine spatial level (Bongaarts and Watkins, 1996; Potter et al., 2002). Assunção et al. (2005) developed empirical Bayes methods for small area estimation, where the data from spatially neighbouring regions is used to improve the precision of estimates, particularly for a region with small sample size. Castro et al. (2015) developed Bayesian small-area inference by spatial clustering of observed fertility rates. The fertility rates were modeled as determined by a collection of covariates, plus a region-specific random effect that has spatial dependence following a conditional autoregressive (CAR) model (Besag, 1974; Mardia, 1988). Since fertility decline is spatially heterogenous and exhibit complex spatial dynamics, based potentially on endogenous socio-cultural diffusion rather than exogenous geographical contiguity, the adjacency matrix is not assumed known, but inferred from the data by spatial clustering. The estimated spatial structure is then used to construct several alternate scenarios, which aid policy-makers and experts in identifying peripheral regions with adverse demographic prospects and designing policy to regenerate such areas. This approach is also related to Billari et al. (2012, 2014) who developed methods for Bayesian stochastic population forecasting based on expert evaluations. In this paper, we consider the approach of Castro et al. (2015) and extend this by developing methods for spatial clustering of curves, applied to time trends of fertility rates across different regions.

The above discussion highlights four main issues that empirical modeling of regional age-specific fertility rates must address: first, the heterogeneity of fertility behavior across age-groups (age-specific fertility rates); second, the territorial patterns and differences across spatial units or regions (spatial heterogeneity); third, socio-cultural and spatial diffusion, potentially subject to spatial non-stationarity (or, strong spatial dependence); and fourth, the time trends of fertility rates, which are heterogeneous in terms of regions and age groups, but typcally exhibit strong spatial clustering. In particular, regional data shows that, in addition to spatial heterogeneity due to varying macrostructures, trends in fertility rates are also subject to complex spatial diffusion that may not be captured by restrictive dynamics based on geographic distances. For example, two nearby regions with relatively similar economic and demographic macrostructure may often exhibit dramatically different temporal patterns of fertility decline (Potter et al., 2002). These arguments justify the integration of the above four features in modeling regional age-specific fertility rates,

to perform accurate estimation, but more importantly to understand spatial diffusion and design demographic policy. This in turn requires new framework and methodologies to encompass the high level of complexity. In this paper, we develop Bayesian methodology to model variations in time trends of fertility and spatial clustering of temporal curves of fertility across different regions. This allows for very rich spatio-temporal dynamics, and improves the precision of demographic forecasts substantially. Most importantly, it enables policy aimed at identifying depressed peripheral regions with poor demographic dynamics and the design of regional policy for such regions.

The rest of the paper is organized as follows. In Section 2, we briefly review spatial smoothing methods in demography. Section 3 describes the Portuguese demographic context, the data structure and exploratory analysis for the age-specific Portuguese NUTS III regions, which motivates our methodology and modeling strategies. Section 4 discusses our methodology. We propose a new spatio-temporal mixed-effects model that jointly considers spatial, temporal, and age-group effects for estimating and predicting the fertility rates, and that allows inference on spatial diffusion (Section 4.1). In Section 4.2, we develop a spatial functional clustering methods based on Bayesian wavelets to estimate the unobserved spatial structure that leads to diffusion. We apply the proposed methodologies to Portuguese NUTS III regional fertility rates data and summarize the results in Section 5, together with policy implications. Section 6 presents a short evaluation of forecast performance, and Section 7 collects conclusions.

2 Smoothing and Spatial Methods in Demography

The methods for spatial clustering developed in this paper are primarily used for understanding the process of fertility diffusion. However, the methods also provide small area estimates by smoothing. Smoothing, spatial modeling and spatial clustering are areas that have witnessed substantial methodological advances over recent times.

2.1 Smoothing and small area estimation

Frequentist smoothing methodologies are popular for estimating cross-classified counts for subpopulations with small sample size or sparse observations. Prime examples are small area estimation (SAE) methods that borrow strength from related characteristics to enhance precision of such count estimates. These methods have been used for measurement of labor force characteristics, poverty and deprivation, and incidence of adverse health outcomes; see for example, Zhang and Chambers (2004), Molina et al. (2007), Ferrante and Trivisano (2010) and López-Vizcaíno et al (2013).

Smoothing techniques aimed either at correcting inaccurate data or obtaining efficient estimates for sub-national domains are also attracting a growing interest in the field of demography; see, for example, Smith and Sincich (1988), Bongaarts and Bulatao (2000), Festy et al. (2002), Alkema et al. (2008), and Raftery et al. (2009). These methods aid estimation of demographic variables either by using macro-level explanatory variables (covariates) or micro-level data for individuals

assumed to have a similar behavior. In the first case, missing or inaccurate data are estimated assuming that the demographic variables under analysis vary according to fixed effects related to a set of macrostructural variables (percentage of rural population, per capita GDP, education, 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 a suitable 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 similar, 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 behavior is affected by qualitative socio-cultural changes or because the analysed areas or cohorts are too small. Both these issues are of key concern in this paper. 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 20th century; this evidence is also in line with observed variation across other developed countries (Guinnane, 2011). The second case corresponds to structural instability, independent of data accuracy and transitional changes, and is particularly important in regional statistics on counts where incidence is reduced, because both the base population and the per capita frequency of such occurrences are small. This structural instability cannot always be adjusted by using non-demographic covariates, particularly when these other covariates are affected by the same sort of issues as well. Therefore, an appropriate alternative is the application of shrinkage techniques where each specific (small area) estimate is improved by using information on the same variable, either from related observational units (Assunção et al., 2005) or from 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 θ, 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:

1. 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

2. Empirical Bayesian approaches, applied by Cavenaghi at 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 these approaches, which are both very flexible and general. Nevertheless, by jointly considering multiple factors and dependence structures, the complexity of statistical inference increases, which in turn can create challenges for efficiently estimating and describing the uncertainty of an increasing number of modeling parameters. However, because of rapid development of scientific computation in both equipment and techniques over the past several decades, these issues can now be well addressed by Bayesian hierarchical models and Markov Chain Monte Carlo (MCMC) techniques.

2.2 Spatial and spatio-temporal modeling in demography

Together with developments in smoothing methods, analyses of social processes embedded within a spatial context have also gained prominence in the recent past, with applications to socio-economic and demographic data; see, for example, Sampson et al. (2002). The spatial distribution of inequality has gained attention (Harvey 1996), and likewise the spatial concentration of poverty and deprivation in urban and peri-urban contexts (Sassen, 1994; Wilson 1987). There is a large literature on spatial inequalities in access to healthcare (Goodchild et al., 2000), and specifically to fertility decisions (Tolnay, 1995) and contraceptive choice (Entwisle et al., 1997), while Chi et al. (2011) use spatial analysis as a tool for small area population forecasting. However, Voss (2007) argues that demographic research has moved from the traditional spatial view where place had a central position to one where increasing attention is offered to the individual as the agent of demographic action. In fact, apart from migration studies (Johnson et al., 2005; Wachter, 2005), there is very little spatial thinking in demography. As Weeks (2004) writes: *"[s]patial analysis has thus far played only a small role in the development and testing of demographic theory."* Based on our discussion earlier, this trend in the literature is not unexpected; while demographic processes such as fertility are driven by diffusion and partly by macrostructural features, geographic distances and contiguity provide at best imperfect measures of the strength and direction of such diffusion.

However, there are recent exciting developments within the spatial econometrics and statistics literatures that may help bridge this gap. The traditional literature assumed an *a priori* known structure of spatial spillovers in terms of a well-specified spatial weights (or, social interactions) matrix, usually measured by geographic distances and contiguity. Then, spatial dependence and spatial heterogeneity are examined within a spatial context implied by the pattern of these spatial weights. In sharp contrast, this new branch of the literature treats these weights as unknown and potentially endogenous, and in itself an object of inference. There are two sub-branches within this literature. Contributions such as Getis and Alstadt (2004), Alstadt and Getis (2006), Folmer and Oud (2008) and Mur and Paelinck (2011) focus on improved measurement of spatial weights using information from observed spatial behaviour. Other contributions take inference a step further by proposing estimation of spatial weights; see Bhattacharjee and Jensen-Butler (2013), Bhattacharjee and Holly (2013) and Bailey et al. (2015) for a representative selection, and Bhattacharjee et al. (2014, 2015) for reviews. Castro et al. (2015) propose Bayesian estimation of spatial weights, or

equivalently the spatial contiguity matrix, by spatial clustering of fertility rates in a specific time-point. This paper extends the methodology by developing Bayesian methods for spatial clustering of curves, which are then applied to time trends of fertility rates in different regions.

Thus, this paper also falls within the domain of spatio-temporal modeling under the Bayesian framework, which has a relatively independent literature. One traditional approach is to extend the well-developed time series models and spatial models for the spatio-temporal data by modeling the dependence structure as an interaction of two random processes that admits temporal and spatial dependence structure, respectively. Inference and prediction are then made based on the calibrated standard regression techniques with composite variance and dependence components.

Alternatively, one can view temporal data as a function of time to study its behavior; see, for example, Clyde and George (2000) and Morris and Carroll (2006). Moreover, one can regionally group such functions (curves) that arise spatially and share common patterns. Under this approach, in addition to fitting and predicting the data, one can identify local similarity and regional effects of temporal trends by conducting spatial cluster analysis, where we seek clusters of curves that are similar for spatially contiguous regions. This approach is closely related to Seiver (1985), where seasonal patterns in fertility are analyzed for a collection of regions to help identify clusters of similar trends. The key difference in our approach is that we do not consider the contiguity (adjacency) matrix as precisely known *a priori*. Diffusion in fertility behavior is based on endogenous social networks and may follow complex spatial patterns that are not necessarily closely related to geography. By allowing diffusion to be based on a variable number of spatial clusters whose boundaries are estimated from the data, our methodology allows socio-cultural diffusion to shape fertility outcomes in relatively unrestricted manner. In turn, this enriches spatial analysis in a way that allows social and cultural diffusion processes embedded in a spatial context to be accurately modeled.

2.3 Spatial clustering of curves

Bayesian approaches for clustering of curves have received considerable attention during the past decade due to their capacity to measure the uncertainty associated with the clustering structure; this is crucial for clustering high-dimensional data where a unique clustering representation can be insufficient. One key feature of the Bayesian clustering methods is to assume the clustering structure, including the number of clusters, is in itself random, and draw posterior realizations of the underlying structure using Bayesian techniques such as Dirichlet process methodologies (Ferguson, 1973; Ray and Mallick, 2006) and reversible jump MCMC algorithm (Green, 1995) for variable-dimensional exploration of the candidate model space; see, for example, Knorr-Held and Raßer (2000). The functional clustering techniques based on Bayesian nonparametric smoothing techniques, such as the wavelet-based approach in Ray and Mallick (2006), do not consider cluster-specific shrinkage levels which may characterize important features of the clusters. It is therefore less restrictive and potentially useful to assume varying shrinkage levels and selection probabilities for

the wavelets across functional clusters. However, issues for hyper-parameters across fixed groups, as addressed in Clyde and George (2003), need to be overcome in the more intractable case of random groups. Furthermore, by extending the spatial clustering technique in Knorr-Held and Raßer (2000) to functional data, the functional clustering technique tends to produce spatially-connected clusters that have meaningful interpretations and provide further insight into regional effects, local similarity and diffusion that accounts for demographic variables of interests. On the other hand, harnessing wavelet smoothing techniques on the spatial clustering method for functional data is necessary for addressing issues of model complexity when there are more than 1 clusters, particularly when the data are high dimensional.

When probability distributions are related to a spatial structure, as anticipated for 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) is used to describe 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, and spatial methods and clustering, to demography is the basis for the development, in the following section, of our methodology for empirical analysis.

3 Data and the Portuguese Context

3.1 Portuguese regional demographic context

The methodology developed in this paper is applied to the context of Portugal which is a good example of how important it is to analyze, in a spatio-temporal framework with flexible spatial diffusion, regional trends in fertility. Though Portugal follows the general trend of the developed countries well into the third phase of demographic transition, the spatial-temporal evolution of the fertility rate by age-group has some important particularities. In the past few decades Portugal has undergone substantial political and socio-economic transformations with considerable impacts on the fertility rate, leading to a late and rapid fertility decline: total fertility rates significantly decreased from 3.19 to 1.35, between 1960 and 2011, and the population replacement threshold (2.07 children by women) has been consistently breached since 1982. In November 2007, Portuguese President Cavaco Silva famously said: *"A country without children is a nation without a future"* (Press Report A.). With the economic crisis, things got worse, with the number of new births falling to a 60-year low in June 2012, a 19 percent fall year-on-year. The situation is so acute that the OECD (2011) expressed strong concerns: *"For over a generation of families fertility rates*

in Portugal have been falling. Today, only one other OECD country has a lower fertility rate (Korea). Sustainable fertility is important to ensure dependency ratios don't threaten Portugal's welfare systems and future productivity. Total fertility rates are sensitive to income shocks such as the global financial crisis; a further fertility dip is evident in Portugal since the onset of the financial crisis."

Even if these changes occurred in the entire country, the intensity of the impacts was not similar in all regions and all age groups. The decrease in total fertility rate was the combined result of a declining fertility rate in younger age groups and an increase in the older age groups. Moreover, the higher levels of total fertility rate moved from the northern regions to the southern and metropolitan areas. This occurred because fertility in younger groups had a sharper decrease in the north while the increase in older groups was stronger in the south and in the Metropolitan Area of Lisbon. Together with fertility rates, there were large variations in demographic change which in turn impacted upon spatial variations in skills supply (OECD, 2014). Over the period 1991 to 2012, Algarve had the fastest rate of population growth at 1.5 percent, followed by Lisbon (0.5 percent), the north, Madeira and Açores in the middle (around 0.2 percent), while Alentejo had a shrinking population and central Portugal was stagnant. Then, this translates into high skills equilibrium in Lisboa and Alentejo and low skills equilibrium in central Portugal, while the north and Algarve have developed a surplus of skills.

The spatial heterogeneity described above is combined as a high level of spatial clustering and diffusion, shaping a complex pattern which can only be fully revealed when the units of analysis are small and representative of local contexts. This means that modeling and estimation of fertility rates should aim to use shrinkage methods, borrowing strength from spatially related regions, consecutive years, and relevant age groups. This is very important for estimation of fertility rates in extreme age groups, which are sensitive to small changes and subject of various types of distortions. Bigotte et al. (2014) study the relationship between population dynamics and the hierarchy of urban centers in Portugal, using geographical contiguity to measure diffusion. However, as discussed above, such reliance on geography as the sole driver of spatial spillovers and concentration is not adequate for demographic variables subject to spatial diffusion. Additionally, the rapid changes in the spatial patterns recently occurred imply that spatial dependence should be captured by estimated spatial weight matrices, rather than by coefficients previously defined in an ad hoc manner. Castro et al. (2015) examine the short term dynamics of fertility rates driven by spatially complex contexts of economic growth and decline, as evidenced in the recent fall in Portuguese fertility rates induced by the economic crisis. They developed Bayesian small-area inference by spatial clustering of observed fertility rates across the Portuguese regions in 2009. Here, the fertility rates are modeled as determined by a collection of covariates, plus a region-specific random effect that has spatial dependence following a CAR model, and where the adjacency matrix of spatial diffusion is not assumed a priori, but estimated from the data by spatial clustering. This allows for rich spatial dynamics and improves the precision of demographic forecasts substantially.

More importantly, such flexible modeling of spatial diffusion allows identification of regions and clusters where policy initiatives need to be placed. Regeneration of the peripheral depressed regions in Portugal, through integrated demographic and economic policies, is a key priority identified by researchers and policy makers. Economic growth in economically depressed or peripheral regions depends on the ability of these places to revert population decline by increasing fertility rates and attracting migrants, which then influences the ability of firms to locate there and create employment opportunities (Castro et al., 2013). The model of endogenous economy-demography linkages in Castro et al. (2013) together with spatial small area estimation in Castro et al. (2015) has been actively used in devising regional policy for the Portuguese regions; see, for example, Castro et al. (2013) and Press reports (B, C and D).

However, the trajectories of fertility decline display both substantial spatial clustering as also heterogeneity in the trends. Hence, careful spatio-temporal modeling is required. In this paper, we develop a framework and methodology for spatial clustering of curves and apply this to data on trajectories of fertility rates across the Portuguese regions. The methods aid understanding the nature of spatial diffusion in fertility and identification of regions where policy initiative need to be focussed.

3.2 Portuguese NUTS III regional fertility data

The Portuguese NUTS III geographical regions ($N = 28$) regions as shown in Figure 1, with shading based on two clusters identified by spatial clustering using data for the year 2009 (Castro et al., 2015). The clusters split the country into two halves, North and South. This is consistent with the discussion above, where the southern regions have higher fertility rates compared with the northern regions. However, this static view is somewhat simplistic, because the highest fertility rates have moved from the rural and northern regions to the urban areas and teh south, as part of the demographic transition in recent decades. Being based on one year, temporal patterns were ignored in the previous analysis and now forms the basis for the current paper.

For each region, we have observed the number of births B_{ijt} and the number of women E_{ijt} for $i = 1, 2, \ldots, N = 28$ regions, $j = 1, 2, \ldots, J = 7$ quinquennial age groups from 15 to 50 years old, and $t = 1, 2, \ldots, T = 19$ time points (years) from year 1991 to 2009. We are interested in modeling the fertility rates $F_{ijt} = B_{ijt}/E_{ijt}$. One common approach for modeling count responses B_{ijt} in Bayesian small area estimation (SAE) is to use spatial Poisson regression; see, for example, Castro et al. (2015) as a previous study of the data in a single year. However, Gaussian models are among the most popular models of flexibly handling spatial and temporal dependence structure due to computational efficiency in particular for high-dimensional data. In this paper, we model transformed fertility rates Y_{ijt} as a Gaussian mixed-effects model. We consider the Freeman-Tukey (square-root) transformation with variance stabilization (Cressie and Chan, 1989)

$$Y_{ijt} = \sqrt{10^3 E_{ijt}^{-1}} \left(\sqrt{B_{ijt}} + \sqrt{B_{ijt} + 1} \right) \tag{1}$$

to obtain the transformed variable Y_{ijt} for fertility rates that meets the Gaussian assumption,

expressed formally below in (3), and avoid additional computational burden introduced by Poisson regression. Note that when a predicted value of the transformed variable $\widehat{Y}_{ijt}$ is obtained, one can obtain an estimate $\widehat{F}_{ijt}$ of the fertility rate by transforming $\widehat{Y}_{ijt}$ back using

$$\widehat{F}_{ijt} = H\left(\widehat{Y}_{ijt}\sqrt{10^{-3}E_{ijt}}\right)/E_{ijt} \tag{2}$$

where $H(x) = \left((x^4+1)/(2x^2)-1\right)/2$ is the inverse function of the Freeman-Tukey transformation $y = \sqrt{x} + \sqrt{x+1}$ on the count response x. Therefore the transformation in (1) does not cause any loss of information in interpretation.

The Freeman-Tukey transformed fertility rates by the quinquennial age groups are shown in Figure 2, with a different trend line for each of the $N = 28$ regions.. Overall, the fertility rates in the South area (in shades of blue), particularly around capital Lisboa, are higher than that in the North area (green and yellow). This difference is particularly evident for the age group 15-19 years. However for 20-24 years old age group the region Tâmega (yellow) in the north has fertility rates above other regions in years before 2005. Generally there is no visible heteroscedasticity over time and between age groups after the variance stabilization in (1); however, the mean levels and time trends are clearly varying across age groups. The age groups with age below 29 years generally exhibit a decreasing trend over the 19 years, while the age groups above 30 has an increasing trend, except that the final age group (45-49 years old) where fertility shows a flat pattern.

Based on the figures, we consider addressing the following two major concerns:

1) Estimating and predicting the age-specific fertility rates based on a model that can accommodate spatial, temporal, and age-group effects. Spatial dependence is assumed to be based on the adjacency of regions.

2) Investigating socio-cultural fertility diffusion, where for each age group the temporal trends display spatial concentration, but the gradient of change across regions is not necessarily driven by geographical distances or contiguity. Rather, one can estimate the adjacency of regions by borrowing strength from the data, and use the estimated adjacency to infer on social and cultural diffusion.

In particular for 2), Figures 1 and 2 jointly suggest that there is a potential local similarity in curve trends and mean levels. Therefore, estimating diffusion patterns by clusters that tend to be spatially connected is meaningful for estimating regional adjacency, but can also capture diffusion patterns.

4 Proposed Spatio-Temporal Models

There are two components to our spatio-temporal model for fertility rates. First, we develop a spatio-temporal mixed effects model assuming knowledge of the drivers of spatial diffusion, followed by estimation of the diffusion patterns by functional spatial clustering.

Figure 1: Map of the $N = 28$ Portuguese NUTS III regions.

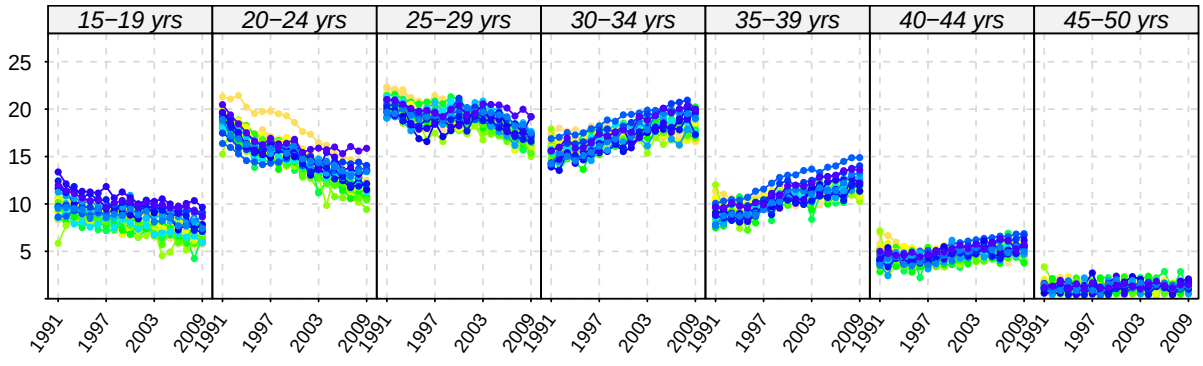

Figure 2: Transformed fertility rates for the $J = 7$ quinquennial age over $T = 19$ years.

4.1 Spatio-Temporal Mixed-effects (STM) model

To model the transformed fertility rates Y_{ijt} by incorporating the spatial, temporal and group effects, we consider the following Spatio-Temporal Mixed-effects model (STM)

$$Y_{ijt} \sim \mathcal{N}\left(\alpha_{it} + \mu_j + \beta_j t,\, \delta^2\right) \tag{3}$$

We assume the fixed age-group effects in both the intercept and time trend, and the spatiotemporal random effect $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', \cdots, \boldsymbol{\alpha}_T')'$, where $\boldsymbol{\alpha}_t = (\alpha_{1t}, \cdots, \alpha_{Nt})'$ is the vector of random effects at time t. The age-group effect does not interact with spatial and temporal dependences.[1] In other words, the spatio-temporal dependence does not vary across age groups, so that it can be quantitatively measured by borrowing strengths from observations in all age groups. Moreover, we assume (i) Markov property; and (ii) Stationarity for both spatial and temporal dependences by specifying

$$\boldsymbol{\alpha}_t = \Phi \boldsymbol{\alpha}_{t-1} + \boldsymbol{\epsilon}_t \tag{4}$$

where $\Phi = diag_{1\le s\le N}\{\phi_s\}$ with $\phi_s \in (-1, 1)$ measures the temporal dependence for each region and $\boldsymbol{\epsilon}_t \sim_{ind} \mathcal{N}\left(\mathbf{0}_N,\, \tau_t^2 \boldsymbol{D}(\gamma_t)\right)$ with γ_t measures the spatial dependence at each time. Note that non-stationary temporal dynamics can be accommodated by letting Φ be the identity matrix. It is also very common to assume the spatial and temporal dependences are separable, for example, setting $\phi_s \equiv \phi$, that is, $\Phi = \phi I_N$, and $\tau_t^2 \equiv \tau^2$ and $\gamma_t \equiv \gamma$; then (4) becomes

$$\boldsymbol{\alpha} \sim \mathcal{N}\left(0,\, \tau^2 \boldsymbol{A}(\phi) \otimes \boldsymbol{D}(\gamma)\right) \tag{5}$$

where " $\otimes$ " denotes the Kronecker product of two matrices; We have Autoregressive (AR) type structure of order 1 for the temporal dependence with covariance matrix $\boldsymbol{A}(\phi) = \left[\phi^{n-m}\right]_{1\le m,n\le T}$, and Conditional Autoregressive (CAR) structure for the spatial dependence with covariance matrix $\boldsymbol{D}(\gamma) = (\boldsymbol{M}^{-1} - \gamma \boldsymbol{W})^{-1}$, where both matrices have inverses in closed form which brings computations efficiency. $\boldsymbol{W}$ is the adjacency matrix with $w_{ij} = 1$ if region i and region j are adjacent and 0 otherwise, and with zero diagonal, that is $w_{ii} \equiv 0$ for $1 \le i, j \le N$. For each region i, define $w_{i+} \equiv \sum_{j=1}^{N} w_{ij}$ as the row sum that represents the total number of neighbors of region i. $\boldsymbol{M}$ is a diagonal matrix with diagonal entries $m_{ii} = 1/\max\{1, w_{i+}\}$. In the above specification, $\phi \in (-1, 1)$ measures temporal dependence and $\gamma \in (e_N^{-1}, e_1^{-1})$ measures spatial dependence, where e_1 and e_N are the largest and smallest eigenvalues of $\boldsymbol{M}\boldsymbol{W}$, which ensures that $\boldsymbol{D}(\gamma)$ is a valid covariance matrix.

Let I_M denote the identity matrix of dimension M, and $\mathbf{1}_M$ denote the column vector of dimension M and all entries 1. By stacking data in the order of geographical regions, time points (years), and then groups, we can write the model in the canonical form of the mixed-effects model

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{Z}\boldsymbol{\alpha},\, \delta^2 I_{NTJ}), \tag{6}$$

[1] This assumption is adequate in our application, but may be strong in other contexts. It can be relaxed.

where the fixed effects $\boldsymbol{\theta} = (\mu_1, \cdots, \mu_J, \beta_1, \cdots, \beta_J)'$ have the design matrix $\boldsymbol{X} = [\boldsymbol{X}_0 \,|\, \boldsymbol{X}_1]$ with $\boldsymbol{X}_0 = I_J \otimes \mathbf{1}_{NT}$ and $\boldsymbol{X}_1 = I_J \otimes ((1, 2, \cdots, T)' \otimes \mathbf{1}_N)$. The design matrix for the spatio-temporal random effects is $\boldsymbol{Z} = \mathbf{1}_J \otimes I_{NT}$.

To proceed with Bayesian estimation of the parameters, we choose the following priors:

$$\begin{aligned} \pi(\theta_k) &\propto 1, \quad \pi(\gamma) = \text{Uniform}(e_N^{-1}, e_1^{-1}), \quad \pi(\phi) = \text{Uniform}(-1, 1), \\ \pi(\delta^2) &= \texttt{igamma}(a_\delta, b_\delta), \quad \pi(\tau^2) = \texttt{igamma}(a_\tau, b_\tau). \end{aligned} \tag{7}$$

For the prior densities that involve hyperparameters, we choose inverse gamma (`igamma`) densities with shape parameters $a_\delta = a_\tau = 2$ and scale parameters $b_\delta = b_\tau = 0.01$. This yields dispersed prior distributions, involves less subjectivity and allows estimates to be more data-driven.

4.1.1 STM Implementation

We obtain the posterior samples of the parameters $\{\boldsymbol{\theta}, \boldsymbol{\alpha}, \delta^2, \tau^2, \gamma\}$ via Gibbs Sampler. Under our choice of the prior densities in (7), the full conditional distributions given the data and the remaining parameters for the fixed-effects and random-effects are, respectively

$$\pi(\boldsymbol{\theta} \,|\, \boldsymbol{\eta}, \boldsymbol{\alpha}, \delta^2) = \mathcal{N}(\mu_{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}}) \quad \begin{cases} \Sigma_{\boldsymbol{\theta}} = \delta^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} \\ \mu_{\boldsymbol{\theta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' (\boldsymbol{\eta} - \boldsymbol{Z}\boldsymbol{\alpha}) \end{cases}$$

$$\pi(\boldsymbol{\alpha} \,|\, \boldsymbol{\eta}, \boldsymbol{\theta}, \delta^2, \tau^2, \gamma, \phi) = \mathcal{N}(\mu_{\boldsymbol{\alpha}}, \Sigma_{\boldsymbol{\alpha}}) \quad \begin{cases} \Sigma_{\boldsymbol{\alpha}} = \left(\delta^{-2} J I_N + \tau^{-2} \boldsymbol{A}(\phi)^{-1} \otimes \boldsymbol{D}(\gamma)^{-1}\right)^{-1} \\ \mu_{\boldsymbol{\alpha}} = \Sigma_{\boldsymbol{\alpha}} \boldsymbol{Z}' (\boldsymbol{\eta} - \boldsymbol{X}\boldsymbol{\theta}) / \delta^2 \end{cases}$$

The conditional distributions of the variance components and spatio-temporal dependences are

$$\begin{aligned} \pi(\delta^2 \,|\, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\theta}) &= \texttt{igamma}\left(a_\delta + NJT/2,\, b_\delta + \boldsymbol{\epsilon}'\boldsymbol{\epsilon}/2\right), \text{with } \boldsymbol{\epsilon} = \boldsymbol{\eta} - \boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{Z}\boldsymbol{\alpha} \\ \pi(\tau^2 \,|\, \boldsymbol{\alpha}, \gamma, \phi) &= \texttt{igamma}\left(a_\tau + NT/2,\, b_\tau + \boldsymbol{\alpha}'(\boldsymbol{A}(\phi)^{-1} \otimes \boldsymbol{D}(\gamma)^{-1})\boldsymbol{\alpha}/2\right) \\ \pi(\gamma \,|\, \boldsymbol{\alpha}, \tau^2, \phi) &\propto |\boldsymbol{D}(\gamma)|^{-T/2} \exp\{\gamma \boldsymbol{\alpha}'(\boldsymbol{A}(\phi)^{-1} \otimes \boldsymbol{W})\boldsymbol{\alpha}/(2\tau^2)\} \cdot \boldsymbol{I}(\gamma \in (e_N^{-1}, e_1^{-1})) \\ \pi(\phi \,|\, \boldsymbol{\alpha}, \tau^2, \gamma) &\propto |\boldsymbol{A}(\phi)|^{-N/2} \exp\{-\boldsymbol{\alpha}'(\boldsymbol{A}(\phi)^{-1} \otimes \boldsymbol{D}(\gamma)^{-1})\boldsymbol{\alpha}/(2\tau^2)\} \cdot \boldsymbol{I}(\phi \in (-1, 1)) \end{aligned}$$

For CAR structure we have a convenient form for the precision matrix $\boldsymbol{D}(\gamma)^{-1} = \boldsymbol{M}^{-1} - \gamma \boldsymbol{W}$ with UL-decomposition $\boldsymbol{D}(\gamma)^{-1} = L_D' L_D$, and for AR(1) structure we have $\boldsymbol{A}(\phi)^{-1} = L_A' L_A$ where

$$L_A = \frac{1}{\sqrt{1-\phi^2}} \begin{pmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 \\ -\phi & 1 & 0 & \cdots & 0 \\ 0 & -\phi & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\phi & 1 \end{pmatrix}.$$

Hence $|\boldsymbol{A}(\phi)| = (1-\phi^2)^{T-1}$. Sampling τ^2 and ϕ requires efficient computation of $\boldsymbol{\alpha}' V \boldsymbol{\alpha}$ where $V = \boldsymbol{A}(\phi)^{-1} \otimes \boldsymbol{D}(\gamma)^{-1} = (L_A' L_A) \otimes (L_D' L_D) = (L_A' \otimes L_D')(L_A \otimes L_D) = L_V' L_V$ with $L_V = L_A \otimes L_D$. Hence $\boldsymbol{\alpha}' V \boldsymbol{\alpha} = \|\tilde{\boldsymbol{\alpha}}\|_2^2$ with $\tilde{\boldsymbol{\alpha}} = L_V \boldsymbol{\alpha}$ and can be efficiently evaluated by exploiting the sparseness of L_A.

4.1.2 STM Model Comparisons

To evaluate the importance of the spatio-temporal dependence, we consider the full implementations of the unrestricted model (3) and the following 3 sub-models under different restrictions:

1. Non-random model: fixing $\boldsymbol{\alpha} \equiv 0$, and hence all associated components τ^2, ϕ and γ are zero.
2. Spatial-only model: fixing $\phi \equiv 0$.
3. Temporal-only model: fixing $\gamma \equiv 0$ and $\boldsymbol{M} = I_N$.

For model comparisons, we use Deviance Information Criterion (DIC) for mixed-effects model, DIC_4 (Celeux et al 2006), based on complete likelihood f

$$\begin{aligned} \text{DIC}_4 &= -4\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\alpha}}[\log f(\boldsymbol{Y}, \boldsymbol{\alpha} \,|\, \boldsymbol{\theta}) \,|\, \boldsymbol{Y}] + 2\mathbb{E}_{\boldsymbol{\alpha}}[\log f(\boldsymbol{Y}, \boldsymbol{\alpha} \,|\, \mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta} \,|\, \boldsymbol{Y}, \boldsymbol{\alpha}]) \,|\, \boldsymbol{Y}] \\ &\triangleq -4\mathbb{E}_1 + 2\mathbb{E}_2 \end{aligned}$$

where the conditional expectation $\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta} \,|\, \boldsymbol{Y}, \boldsymbol{\alpha}]$ can be evaluated by sampling $\boldsymbol{\theta}$ for each posterior sample of $\boldsymbol{\alpha}$ and obtain the mean. We also report $\overline{D(\boldsymbol{\theta})} = -2\mathbb{E}_1$ as the posterior expected value of the joint deviance, and $p_{D4} = \overline{D(\boldsymbol{\theta})} + 2\mathbb{E}_2$ as a measure of model dimensionality. Generally, a smaller DIC_4 indicates better predictive perfomance.

4.2 Spatial Functional Clustering (SFC) Model

Our STM model above assumes the specification of an adjacency matrix $\boldsymbol{W}$. However, as discussed before, spatial diffusion in fertility is complex and need not be adequately described by geographical proximity or contiguity. At the same time, socio-cultural diffusion should capture the phenomenon of spatial concentration observed in data on fertility rates and their evolution over time. In this section, we consider grouping the observed curves in Figure 2 according to the local similarity and regional effects to capture the common trends in fertility rates, and thereby estimate the adjacency matrix $\boldsymbol{W}$ from the data.

We adopt the functional mixed-effects model approach by regularizing the curves (temporal observations) using basis functions, such as wavelets, for reducing the number of parameters proliferated by the clustering structure. Since different age groups can have different clustering configurations, we conduct the clustering analysis for each group. We therefore suppress the group index j in the following model description. Importantly, the number of clusters is not assumed known, but is itself an object of inference.

4.2.1 SFC Model Specification

Consider the functional response $Y_i(t)$ over $[0, 1]$ for region $i = 1, 2, \ldots, N$, with $k = 1, 2, \ldots, p$ functional covariates $\tilde{X}_i(t)$ including the intercept. Our goal is to spatially cluster these regions according to the regional impact of $\tilde{X}_i(t)$'s. In general, the $\tilde{X}_i(t)$'s may capture macrostructural

characteristics, and spatial diffusion may reflect the variation in the socio-cultural effects of these macrostructural features on fertility. When $\tilde{X}_i(t)$ includes only the intercept, with $p = 1$, we would cluster $Y_i(t)$ by its trajectory as ordinary functional clustering procedure.

Suppose we have data $Y_i(t)$ observed at $t = t_1, t_2, \ldots, t_T$ over a set of regions, $i \in S = \{1, 2, \ldots, N\}$, which can be partitioned into d clusters, $S = \cup_{r=1}^{d} C_r$; C_r's are mutually exclusive and each contains n_r regions and we have $\sum_{r=1}^{d} n_r = N$. Consider the model:

$$Y_i = \tilde{X}_i \tilde{\boldsymbol{\beta}}_r + \epsilon_i \tag{8}$$

for $i \in C_r, r = 1, 2, \ldots, d$, where Y_i is the $T \times 1$ functional response for i-th region, and ϵ_i is the $T \times 1$ error vector and which follows i.i.d. $\mathcal{N}(0, \sigma^2)$; $\tilde{\boldsymbol{\beta}}_r$ is the $pT \times 1$ vector that represents the impact of $\tilde{X}_i$ on Y_i in the C_r cluster, with the $T \times pT$ design matrix $\tilde{X}_i$ defined correspondingly. The model can be written with data pooled by region as

$$\boldsymbol{Y} = \tilde{\boldsymbol{X}} \tilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon} \tag{9}$$

where $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1', \cdots, \tilde{\boldsymbol{\beta}}_d')'$ is $dpT \times 1$ vector of slope coefficients organised as

$$\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_{11}(\boldsymbol{t})', \ldots, \tilde{\beta}_{1p}(\boldsymbol{t})', \ldots, \tilde{\beta}_{d1}(\boldsymbol{t})', \ldots, \tilde{\beta}_{dp}(\boldsymbol{t})')' \tag{10}$$

with $\tilde{\beta}_{rk}(\boldsymbol{t}) = (\tilde{\beta}_{rk}(t_1), \ldots, \tilde{\beta}_{rk}(t_T))', r = 1, \ldots, d, k = 1, \ldots, p$.

It is evident that model complexity is greatly increased by the clustering structure. To avoid potential redundancy of parameters, we consider Bayesian wavelet smoothing techniques for dimension reduction. Specifically, let $\boldsymbol{\beta} = \boldsymbol{Q}\tilde{\boldsymbol{\beta}}$ with the block-diagonal matrix $\boldsymbol{Q} = \bigoplus_{i=1}^{dp} Q$, where each block matrix Q is the Discrete Wavelet Transform (DWT) matrix; that is, we reparametrize $\tilde{\boldsymbol{\beta}}$ by Wavelet transforming each of its T-functional components. Then the slope vector reduces to $\boldsymbol{\beta} = (\beta_{11}', \cdots, \beta_{1p}', \cdots, \beta_{d1}', \cdots, \beta_{dp}')'$ with $T \times 1$ wavelet coefficients β_{rk} for $\tilde{\beta}_{rk}(\boldsymbol{t})$. Furthermore, without loss of generality, by assuming $T = 2^L$, we have the explicit regularization form

$$\tilde{\beta}_{rk}(t) = \xi_{rk,00}\phi_{00}(t) + \sum_{l=1}^{L} \sum_{m=1}^{2^{l-1}} \zeta_{rk,lm}\psi_{lm}(t),$$

where $\phi_{00}(t)$ and $\psi_{lm}(t)$ are the wavelet basis functions at resolution level $l = 1, 2, \ldots, L$, and $m = 1, 2, \ldots, 2^{l-1}$ denotes the location at the l-th level. Then, $\beta_{rk} = (\xi_{rk,00}, \zeta_{rk,11}, \zeta_{rk,21}, \ldots, \zeta_{rk,L2^{L-1}})'$. Note that, the coefficients in the temporal domain with a single index (t) are regularized in the wavelet domain with double index (lm), with equal length as $1 + \sum_{l=1}^{L} \sum_{m=1}^{2^{l-1}} 1 = 2^L = T$. The model in (9) can now be written as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\boldsymbol{X} = \tilde{\boldsymbol{X}}\boldsymbol{Q}'$. Under the clustering structure, the model can be also written at cluster level as

$$\boldsymbol{Y}_r = \boldsymbol{X}_r \boldsymbol{\beta}_r + \boldsymbol{\epsilon}_r, \quad r = 1, 2, \cdots, d, \tag{11}$$

where $\boldsymbol{Y}_r$ is the $n_r T \times 1$ vector obtained by stacking all the Y_i's for $i \in C_r$. The likelihood of the model is

$$f(\boldsymbol{Y} \,|\, \boldsymbol{\beta}, \sigma^2) = \prod_{r=1}^{d} \mathcal{N}(\boldsymbol{X}_r \boldsymbol{\beta}_r,\, \sigma^2 I_{n_r T}) \tag{12}$$

For the intercept-only clustering model ($p = 1$), each $\tilde{X}_i = I_T$ and hence $X_i = Q'$, the inverse-wavelet transformation matrix. Hence $\boldsymbol{X}_r = (Q, Q, \cdots, Q)' = \mathbf{1}_{n_r} \otimes Q'$ and $\boldsymbol{X}_r'\boldsymbol{X}_r = n_r I_T$ by the orthonormality of Q.

Developing inferences for $\boldsymbol{W}$ is a major methodological innovation in the current paper. Here, we conduct such inferences using spatial clustering of curves, where the number of clusters is not assumed *a priori*, but is also an object of inference.

4.2.2 SFC Prior Specification

Let ϖ be a clustering configuration which consists of two components $\varpi = (d, \boldsymbol{z})$, where d is the number of clusters and $\boldsymbol{z} = (z_1, z_2, \cdots, z_N)$ is the vector containing labels for each region, $z_i \in \{1, 2, \cdots, d\}$. For spatial clustering, we can reduce to $\varpi = (d, G_d)$ where $G_d = (g_1, \cdots, g_d)$ are cluster centers and $g_r \in S$; see, for example, Knorr-Held and Raßer (2000) and Feng et al. (2014). The membership $\boldsymbol{z}$ is determined by the minimal distance criterion, that is, regions with the minimal distance from center g_r constitute the cluster C_r.[2]

A clustering prior model is defined as $\pi(\varpi) = \pi(d)\pi(G_d \,|\, d)$, where $\pi(d)$ is a prior for the number of clusters, d, and $\pi(G_d \,|\, d)$ is a prior for the cluster centers, G_d. We specify the minimal cluster size n_0 such that $n_r \geq n_0 \geq 1$ for all C_r's; consequently, d is bounded above, say $d \leq N_0 \leq N$.[3] Knorr-Held and Raßer (2000) considered a power penalty on the number of clusters d by eliciting $\pi(d \,|\, \alpha) \propto (1 - \alpha)^d$ so that $\pi(d + 1 \,|\, \alpha)/\pi(d \,|\, \alpha) = 1 - \alpha$, although an exponential penalty can mimic the well-known criterion such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) for model comparisons. Under their choice, when d is bounded above by N_0, we explicitly have $\pi(d \,|\, \alpha) = C(\alpha)(1 - \alpha)^{d-1}$ for $d = 1, \cdots, N_0$ with $C(\alpha) = \alpha/(1 - (1 - \alpha)^{N_0})$ and $\pi(\alpha) \sim \text{Unif}[0, 1]$. When α is larger, d receives higher penalty for increasing, while when α is fixed at a tiny value, say $\alpha = 0$, d becomes uniformly distributed over $\{1, 2, \ldots, N_0\}$ and all possible values of d receives the same prior weight. Conditioning on d, we choose the prior for centers to be $\pi(G_d \,|\, d) = (N - d)!/N!$ and the $\binom{N}{d} d!$ possible G_d's receive equal prior weight for being the cluster centers.

Although it is common to assume certain penalty for higher-dimensional models by penalizing d, it is generally more reasonable to penalize the associated model parameters themselves. Popular choices are L-2 penalty through a normal prior, or L-1 penalty through a logistic density prior on $\boldsymbol{\beta}$ which is equivalent to LASSO estimates (Park and Casella, 2008). More specifically, for the

[2] For areal data, the minimum number of boundaries that need to be crossed to go from one region to another (the order of a neighbor), can be taken as a measure of distance. This can be obtained from the natural adjacency matrix which is typically available for areal data. For point reference data, one can alternatively use the Euclidean distance between coordinates as the distance. Consequently G_d is non-ordered to avoid misspecification under minimum distance criterion since there are distance ties naturally associated with the order of neighbor as distance.

[3] Note that the minimal cluster size is not really a constraint, since this can be as small as 1. However, the researcher is offered the flexibility to choose any other minimal size as might be relevant to a specific application.

fixed-effects function $\boldsymbol{\beta}$, we assume each element $\beta_{rk}(lm)$ of $\boldsymbol{\beta}$ is i.i.d.

$$\beta_{rk}(lm) \,|\, \varpi \sim \mathcal{N}(0,\, \sigma^2 \lambda_{rkl} \gamma_{rk}(lm)), \qquad \gamma_{rk}(lm) \sim \text{Bernoulli}\,(p_{rkl}). \tag{13}$$

Note when $\gamma_{rk}(lm) = 0$, $\pi(\beta_{rk}(lm) \,|\, \varpi)$ is a point mass at 0. The clustering configuration ϖ and the wavelet shrinkage indicator $\gamma_{rk}(lm)$'s jointly determine the effective size of $\boldsymbol{\beta}$. This prior specification (13) introduces a set of hyperparameters: the signal-to-noise ratio λ_{rkl}, and the amount of shrinkage p_{rkl}, which can be estimated using an empirical Bayes procedure as described by Clyde et al. (2000). However, empirical Bayesian estimation procedures become intractable when the random clustering structure is introduced. Furthermore, the clustering results can be sensitive to the selection of hyperparameters. Therefore, we update these hyperparameters during the posterior sampling procedure.

Accordingly, we elicit a Beta prior $p_{rkl} \sim \text{Beta}\,(a_{0,kl}, b_{0,kl})$, that is, $\pi(p_{kl}) \propto p_{rkl}^{a_{0,kl}-1}(1 - p_{rkl})^{b_{0,kl}-1}$, where the case $a_{0,kl} = b_{0,kl} = 1$ corresponds the non-informative uniform prior on $[0, 1]$, and we assume Inverse Gamma prior for $\lambda_{rkl} \sim \texttt{igamma}(a_{1,kl}, b_{1,kl})$ with shape and scale parameters ($a_{1,kl}$ and $b_{1,kl}$ respectively) selected to specify a rather dispersed distribution to indicate our weak prior information. For the variability parameter for $\boldsymbol{\epsilon}$, we consider the inverse-Gamma prior $\sigma^2 \sim \texttt{igamma}(a_\sigma, b_\sigma)$, where we choose $a_\sigma = 2$ and $b_\sigma = 0.01$. The hyperparameters $a_{0,kl}$, $b_{0,kl}$, $a_{1,kl}$ and $b_{1,kl}$ are estimated from the posterior samples by fitting a model with no clusters $d = 1$ which gives reasonable prior information of the scaling and shrinkage levels of the wavelet coefficients.

For notational simplicity, we use $(\boldsymbol{a}_{kl})_{k,l}$ to denote a vector $\boldsymbol{a}$ with elements indexed by double index $k = 1, 2, \ldots, K$ and $l = 1, 2, \ldots, L$, and stacked by the order of the subscripts, that is, $(\boldsymbol{a}_{kl})_{k,l} = (a_{1,1}, \ldots, a_{K,1}, \ldots, a_{1,L}, \ldots, \alpha_{K,L})$; we use a similar notation for more than two indexes. For example, we have $\boldsymbol{\beta} = (\beta_{rk}(lm))_{m,l,k,r}$. Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \cdots, \boldsymbol{\gamma}_d')'$ where $\boldsymbol{\gamma}_r = (\gamma_{rk}(lm))_{m,l,k}$, $\boldsymbol{\lambda}_r = (\lambda_{rkl})_{l,k,r}$ and $\boldsymbol{p} = (p_{rkl})_{l,k,r}$. Then, the prior (13) can be re-written as $\boldsymbol{\beta}_r | \varpi \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \Lambda_r)$ where Λ_r is a diagonal matrix with $\boldsymbol{\lambda}_r \boldsymbol{\gamma}_r$ as diagonal entries. We also employ $\boldsymbol{a}[c]$ to denote the c-th element (column) of the vector (matrix) $\boldsymbol{a}$, and $\boldsymbol{a}[-c]$ denotes the complement of $\boldsymbol{a}[c]$.

4.2.3 SFC Implementation

The Bayesian estimation of parameters $\{\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \sigma^2, \varpi\}$ involves variable model dimensions. We consider an embedded reversible jump MCMC within the conventional Gibbs sampler that updates a set of parameters given the rest, using the full conditional distributions. More specifically, we consider following steps within one Gibbs circle:

(1) Update ϖ: The model parameters are clustering parameters $\varpi = (d, G_d)$, cluster-specific parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ with dimension determined by ϖ, variance component σ^2 and hyperparameters $(\boldsymbol{\lambda}_r, \boldsymbol{p}, \alpha)$. We construct a reversible jump MCMC algorithm (Green 1995) to adjust the sampling procedures to the situation where the dimensionality of models change. For a current state $(\varpi, \boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta})$, if we propose a new state $(\varpi^*, \boldsymbol{\gamma}^*, \sigma^{2*}, \boldsymbol{\beta}^*)$ where ϖ^* vary from ϖ, we define the

auxiliary $U = \boldsymbol{\theta}^* = (\sigma^{2*}, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*)$ and let $\boldsymbol{\theta} = (\sigma^2, \boldsymbol{\beta}, \boldsymbol{\gamma}) = U^*$, the corresponding invertible map $\boldsymbol{q} : (\boldsymbol{\theta}, U)$ to $(\boldsymbol{\theta}^*, U^*)$ is one-to-one with Jacobian 1. We propose $U \sim h(\cdot \,|\, \varpi, \boldsymbol{\theta}, \varpi^*)$ for a certain proposal density h. The acceptance ratio for the proposal is

$$\min\left\{1,\ \frac{g(\varpi \,|\, \varpi^*)}{g(\varpi^* \,|\, \varpi)} \times \frac{\pi(\varpi^*, \boldsymbol{\theta}^* \,|\, \boldsymbol{Y})}{\pi(\varpi, \boldsymbol{\theta} \,|\, \boldsymbol{Y})} \times \frac{h(u^* \,|\, \varpi^*, \boldsymbol{\theta}^*, \varpi)}{h(u \,|\, \varpi, \boldsymbol{\theta}, \varpi^*)} \times 1\right\} \tag{14}$$

We choose $h(u \,|\, \varpi, \boldsymbol{\theta}, \varpi^*) = \pi(\boldsymbol{\gamma}^* \,|\, \varpi^*) \times \pi(\sigma^{2*} \,|\, \varpi^*, \boldsymbol{\gamma}^*, \boldsymbol{Y}) \times \pi(\boldsymbol{\beta}^* \,|\, \varpi^*, \boldsymbol{\gamma}^*, \sigma^{2*}, \boldsymbol{Y})$. More specifically, we first generate $\boldsymbol{\gamma}^*$ from its prior distribution, that is, $\gamma_{rk}(lm)$ from Bernoulli(p_{rkl}) for the new cluster only. Note for given ϖ^* and $\boldsymbol{\gamma}^*$, the effective dimension of $\boldsymbol{\beta}^*$ is determined. Let $\boldsymbol{\beta}^*_{r,1}$ be the non-zero portion of each $\boldsymbol{\beta}^*_r$ with dimension q_r, the corresponding design matrix $\boldsymbol{X}_{r,1}$, and the prior covariance $\sigma^2\boldsymbol{\Lambda}_{r,1}$ determined by $\boldsymbol{\gamma}^*_r$. We first sample σ^{2*} from

$$\pi(\sigma^{2*} \,|\, \varpi^*, \boldsymbol{\gamma}^*, \boldsymbol{Y}) \propto \pi(\sigma^{2*}) \prod_{r=1}^{d} \int \pi(\boldsymbol{Y}_r \,|\, \boldsymbol{\beta}_r, \sigma^{2*})\, \pi(\boldsymbol{\beta}_r \,|\, \boldsymbol{\gamma}^*_r)\, d\boldsymbol{\beta}_r \tag{15}$$

which is an inverse-Gamma density with shape parameter $a_\sigma + NT/2$ and scale parameter $b_\sigma + \sum_{r=1}^{d} \boldsymbol{Y}_r^T V_r \boldsymbol{Y}_r/2$ with $V_r = I_{n_r T} - \boldsymbol{X}_{r,1}(\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + \boldsymbol{\Lambda}^{-1}_{r,1})^{-1}\boldsymbol{X}'_{r,1}$. Next, we sample the non-zero portion $\boldsymbol{\beta}^*_{r,1}$ from $\pi(\boldsymbol{\beta}^*_{r,1} \,|\, \varpi^*, \sigma^{2*}, \boldsymbol{Y}_r, \boldsymbol{\gamma}^*_r)$, which is in turn a Gaussian density with mean $(\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + \boldsymbol{\Lambda}^{-1}_{r,1})^{-1}\boldsymbol{X}'_{r,1}\boldsymbol{Y}_r$ and covariance matrix $\sigma^{2*}(\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + \boldsymbol{\Lambda}^{-1}_{r,1})^{-1}$. Under this choice of proposal density, we can substitute

$$\pi(\varpi, \boldsymbol{\theta} \,|\, \boldsymbol{Y}) = \pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\theta})\pi(\boldsymbol{\beta}|\varpi, \boldsymbol{\gamma}, \sigma^2)\pi(\sigma^2 \,|\, \varpi, \boldsymbol{\gamma})\pi(\boldsymbol{\gamma} \,|\, \varpi)\pi(\varpi)/\boldsymbol{m}(\boldsymbol{Y})$$

and cancel out the prior and proposal part $\pi(\boldsymbol{\gamma}^* \,|\, \varpi)$. Then, the Metropolis-Hasting ratio in (14) becomes

$$\frac{g(\varpi \,|\, \varpi^*)}{g(\varpi^* \,|\, \varpi)} \times \frac{\pi(\boldsymbol{Y} \,|\, \varpi^*, \boldsymbol{\beta}^*, \sigma^{2*}, \boldsymbol{\gamma}^*)\, \pi(\boldsymbol{\beta}^*, \sigma^{2*} \,|\, \varpi^*, \boldsymbol{\gamma}^*)\, \pi(\boldsymbol{\beta}, \sigma^2|\varpi, \boldsymbol{\gamma}, \boldsymbol{Y})}{\pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma})\, \pi(\boldsymbol{\beta}, \sigma^2 \,|\, \varpi, \boldsymbol{\gamma})\, \pi(\boldsymbol{\beta}^*, \sigma^{2*} \,|\, \varpi^*, \boldsymbol{\gamma}^*, \boldsymbol{Y})} \times \frac{\pi(\varpi^*)}{\pi(\varpi)} \tag{16}$$

Using the fact that

$$\pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\gamma}) = \frac{\pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma})\, \pi(\boldsymbol{\beta}, \sigma^2 \,|\, \varpi, \boldsymbol{\gamma})}{\pi(\boldsymbol{\beta}, \sigma^2 \,|\, \varpi, \boldsymbol{\gamma}, \boldsymbol{Y})},$$

the ratio (16) reduces to

$$\frac{g(\varpi \,|\, \varpi^*)}{g(\varpi^* \,|\, \varpi)} \times \frac{\pi(\boldsymbol{Y} \,|\, \varpi^*, \boldsymbol{\gamma}^*)}{\pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\gamma})} \times \frac{\pi(\varpi^*)}{\pi(\varpi)}. \tag{17}$$

When the prior and proposal density of ϖ are both diffuse (such as under fixed tiny α and uniform proposal), the acceptance rate is mainly determined by the marginal likelihood ratio, the middle term in (17). We have the marginal likelihood ratio $\boldsymbol{r} = \pi(\boldsymbol{Y} \,|\, \varpi^*, \boldsymbol{\gamma}^*)/\pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\gamma})$ with

$$\pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\gamma}) = \int \left(\prod_{r=1}^{d} \int \pi(\boldsymbol{Y}_r \,|\, \sigma^2, \boldsymbol{\beta}_r, \varpi, \boldsymbol{\gamma})\, \pi(\boldsymbol{\beta}_r \,|\, \sigma^2, \varpi, \boldsymbol{\gamma})\, \pi(\sigma^2)\, d\boldsymbol{\beta}_r \right) d\sigma^2 \tag{18}$$

which simplifies to a multivariate t-density

$$\pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\gamma}) = \frac{\Gamma(a_\sigma + NT/2)\boldsymbol{\Delta}^{-(a_\sigma + NT/2)}}{\Gamma(a_\sigma)(2\pi b_\sigma)^{NT/2}} \prod_{r=1}^{d} |\, \boldsymbol{\Lambda}_{r,1}\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + I_{q_r}|^{-\frac{1}{2}}, \tag{19}$$

where $\boldsymbol{\Delta} = 1 + \sum_{r=1}^{d} \boldsymbol{Y}'_r(I_{n_rT} - \boldsymbol{X}_{r,1}(\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + \boldsymbol{\Lambda}_{r,1}^{-1})^{-1}\boldsymbol{X}'_{r,1})\boldsymbol{Y}_r/(2b_\sigma)$. See Appendix A for a detailed proof. Since the marginal likelihood can be sensitive to the choice of a_σ and b_σ, one can choose the non-informative Jeffery's prior ($a_\sigma = b_\sigma = 0$) when prior information for the noise level is not available. Then, the marginal likelihood would no longer be multivariate t-distributed as in (19), but rather

$$\pi(\boldsymbol{Y} \,|\, \varpi, \boldsymbol{\gamma}) = \frac{\Gamma(NT/2)\boldsymbol{\Delta}_0^{-(NT/2)}}{(2\pi)^{NT/2}} \prod_{r=1}^{d} |\boldsymbol{\Lambda}_1^r \boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + I_{q_r}|^{-\frac{1}{2}} \tag{20}$$

where $\boldsymbol{\Delta}_0 = \sum_{r=1}^{d} \boldsymbol{Y}'_r(I_{n_rT} - \boldsymbol{X}_{r,1}(\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + \boldsymbol{\Lambda}_{r,1}^{-1})^{-1}\boldsymbol{X}'_{r,1})\boldsymbol{Y}_r/2$.

Next we consider how to propose ϖ^* which involves varying model dimensions. The efficiency of implementing the reversible jump MCMC algorithm strongly relies on constructing different move types towards different neighbors of the current state ϖ in the model space. This requires specification of the probabilities for each type of move. For instance, in a variable selection problem, one can assign equal probability of making a birth move (add a new variable) and death move (remove an existing variable). For the spatial clustering model, we need to consider not only the cases of creating one more cluster and deleting an existing cluster, but also a move to a different clustering configuration with the same number of clusters but different cluster centers, particularly when distances are discrete and can therefore introduce ties. Specifically, to propose a new ϖ^* at each iteration, we consider the following moves with the corresponding proposal probabilities P(Growth) = P(Merge) = 0.4 and P(Shift) = P(Switch) = 0.1; this specification is similar to the choice in Knorr-Held and Raßer (2000). The motivation underlying the choice of the prescribed probabilities, which gives higher chance of proposing a new clustering configuration with a different number of clusters d, is to fully explore the uncertainty associated with d. For completeness, we provide the details of the 4 steps above together with the respective acceptance probabilities under our choice of the prior densities and h; see Appendix B.

(2) Update $(\boldsymbol{\beta}, \boldsymbol{\gamma})$: The independence of posterior distribution of $\boldsymbol{\beta}_r$'s and $\boldsymbol{\gamma}_r$'s follows the independence structure in both likelihood and prior distributions. Hence, we update $(\boldsymbol{\beta}_r, \boldsymbol{\gamma}_r)$ independently for each $r = 1, \cdots, d$ through $\pi(\boldsymbol{\beta}_r, \boldsymbol{\gamma}_r \,|\, \boldsymbol{Y}_r, \varpi, \boldsymbol{\eta}^2, \boldsymbol{p}) = \pi(\boldsymbol{\beta}_r \,|\, \boldsymbol{\gamma}_r, \boldsymbol{Y}_r, \boldsymbol{\eta}^2, \varpi) \times \pi(\boldsymbol{\gamma}_r \,|\, \boldsymbol{Y}_r, \boldsymbol{p}, \varpi)$. We update each pair $(\boldsymbol{\beta}_r[c], \boldsymbol{\gamma}_r[c])$ given others $(\boldsymbol{\beta}_r[-c], \boldsymbol{\gamma}_r[-c])$, where $c = 1, 2, \ldots, pT$. We first sample $\boldsymbol{\gamma}_r[c]$ through

$$\begin{aligned} \pi(\boldsymbol{\gamma}_r[c] \,|\, \boldsymbol{Y}_r, \boldsymbol{p}) &= \int \pi(\boldsymbol{\gamma}_r[c], \boldsymbol{\beta}_r[c] \,|\, \boldsymbol{Y}_r, \boldsymbol{\beta}_r[-c], \varpi, \sigma^2, \boldsymbol{\lambda}_r, \boldsymbol{p})\, d\boldsymbol{\beta}_r[c] \\ &= \frac{\pi(\boldsymbol{\gamma}_r[c] \,|\, \boldsymbol{p})}{\boldsymbol{m}(\boldsymbol{Y}_r)} \int \pi(\boldsymbol{Y}_r \,|\, \boldsymbol{\beta}_r[c], \boldsymbol{\beta}_r[-c], \sigma^2)\, \pi(\boldsymbol{\beta}_r[c] \,|\, \boldsymbol{\gamma}_r[c], \boldsymbol{\eta}^2)\, d\boldsymbol{\beta}_r[c]. \end{aligned}$$

Let M_s denotes the likelihood with $\boldsymbol{\beta}_r[c]$ marginalized out conditioning on $\boldsymbol{\gamma}_r[c] = s$ for $s = 0, 1$. Then M_1 is the density of $\mathcal{N}\left(\boldsymbol{X}_r[-c]\boldsymbol{\beta}_r[-c], \sigma^2(I_{n_rT} + \boldsymbol{\lambda}_r[c]\boldsymbol{X}_r[c]\boldsymbol{X}_r[c]')\right)$. Similarly, M_0 is the density of $\mathcal{N}\left(\boldsymbol{X}_r[-c]\boldsymbol{\beta}_r[-c], \sigma^2 I_{n_rT}\right)$. Assume the index c corresponds to the k-th covariate and l-the resolution level, then the posterior odds of $\boldsymbol{\gamma}_r[c]$ being 1 is $O_\gamma = p_{rkl}M_1/\left[(1 - p_{rkl})M_0\right]$ and we sample $\boldsymbol{\gamma}_r[c] \sim$Bernoulli$(O_\gamma/(O_\gamma + 1))$.

Next, we sample $\boldsymbol{\beta}_r[c]$ from posterior distribution given the updated $\boldsymbol{\gamma}_r[c]$. If $\boldsymbol{\gamma}_r[c] = 0$, we have $\boldsymbol{\beta}_r[c] = 0$; otherwise, we sample $\boldsymbol{\beta}_r[c]$ through

$$\pi(\boldsymbol{\beta}_r[c] \,|\, \boldsymbol{Y}_r, \boldsymbol{\lambda}_r, \boldsymbol{\beta}_r[-c], \boldsymbol{\gamma}_r[c] = 1) \propto \pi(\boldsymbol{Y}_r \,|\, \boldsymbol{\beta}_r[c], \boldsymbol{\beta}_r[-c], \sigma^2)\, \pi(\boldsymbol{\beta}_r[c] \,|\, \boldsymbol{\gamma}_r[c] = 1, \boldsymbol{\lambda}_r),$$

which is a Gaussian density with variance $\delta^2 = \sigma^2 \left(\boldsymbol{\lambda}_r[c]^{-1} + \boldsymbol{X}_r[c]'\boldsymbol{X}_r[c]\right)^{-1}$ and mean $\mu = (\boldsymbol{\lambda}_r[c]^{-1} + \boldsymbol{X}_r[c]'\boldsymbol{X}_r[c])^{-1}\boldsymbol{X}_r[c]'(\boldsymbol{Y}_r - \boldsymbol{X}_r[-c]\boldsymbol{\beta}_r[-c])$. Note that the Bayes factor M_1/M_0 for calculating posterior odds O_γ has the simple expression

$$\log(M_1/M_0) = -\frac{1}{2}\log(1 + \boldsymbol{\lambda}_r[c]\boldsymbol{X}_r[c]'\boldsymbol{X}_r[c]) + \frac{\mu^2}{2\delta^2}.$$

(3) Update $(\sigma^2, \boldsymbol{\lambda}_r, \boldsymbol{p}, \alpha)$**:** During this step, we update $(\boldsymbol{p}, \sigma^2, \alpha)$ under current $(\varpi = (d, G_d), \boldsymbol{\beta}, \boldsymbol{\gamma})$. Although σ^2 is updated when proposing ϖ^*, we further update it given the clustering configuration to obtain its sample within the Gibbs circle at every iteration even when a new configuration ϖ^* is not accepted during the Metropolis step.

The conditional distribution of the shrinkage probability p_{rkl} is given by

$$p_{rkl} \,|\, \boldsymbol{\gamma} \sim \text{Beta}\left(a_{0,kl} + \sum_{m=0}^{2^l-1} \gamma_{rk}(ml),\; b_{0,kl} + \sum_{m=0}^{2^l-1}(1 - \gamma_{rk}(lm))\right) \tag{21}$$

for $k = 1, \ldots, p$, and $l = 1, \ldots, L$.

The conditional distribution of residual variance σ^2 is

$$\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\lambda} \sim \texttt{igamma}(a^*_\sigma, b^*_\sigma) \tag{22}$$

with $a^*_\sigma = a_\sigma + \left(NT + \sum_{r,k,l,m} \gamma_{rk}(lm)\right)/2$, and
$b^*_\sigma = b_\sigma + \left((\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \sum_{r,k,l,m} \beta^2_{rk}(lm)/\lambda_{rkl}\right)/2$.

The conditional distribution of signal-to-noise ratio $\boldsymbol{\lambda}_r$ is

$$\lambda_{rkl} \mid \boldsymbol{\beta}, \sigma^2 \sim \texttt{igamma}(a^*_{1,rkl}, b^*_{1,rkl}) \tag{23}$$

where $a^*_{1,rkl} = a_{1,kl} + \sum_m \gamma_{rk}(lm)/2$ and $b^*_{1,rkl} = b_{1,kl} + \sum_m \beta_{rk}(lm)^2/(2\sigma^2)$.

Updating α is optional. Under the power penalty of d, the posterior $\pi(\alpha \,|\, \varpi) \propto \pi(\alpha)\pi(d \,|\, \alpha) \propto \alpha(1-\alpha)^{d-1}/(1-(1-\alpha)^{N_0})$ with $\alpha \in (0, 1)$, and can be sampled using the Griddy-Gibbs sampler (Ritter and Tanner, 1992).

4.2.4 SFC Posterior Inference

The prediction for the above spatial clustering model is based on averaging over all models ϖ's using the posterior samples, which is shown to have better predictive power compared with the prediction from a single model (Raftery et al., 1997). To estimate the adjacency matrix $\boldsymbol{W}$, once we have obtained posterior samples of size B, consider $w_b(i_1, i_2) = 1$ if regions i_1 and i_2 share the same

clustering membership in the b-th sample, and 0 otherwise. Let $w_{i_1 i_2} = \sum_{b=1}^{B} w_b(i_1, i_2)/B \in [0,1]$ and $w_{i_1 i_1} \equiv 0$ be the estimated adjacency between regions i_1 and i_2. We can then fit the SRP with our estimate $\widehat{\boldsymbol{W}} = ((w_{i_1 i_2}))_{1 \le i_1, i_2 \le N}$. To further obtain a central clustering configuration ϖ^* from the posterior samples $\varpi^1, \cdots, \varpi^B$, consider the dissimilarity measure $Diss(i_1, i_2) = 1 - w_{i_1 i_2}$ based on which an agglomerative clustering algorithm is performed with the number of cluster $\widehat{d}$ set to the posterior mode of d. Once the memberships for each region under the central configuration are determined, the inference on the cluster-specific effects $\beta_r(t) \,|\, \varpi^*$ can be made via the posterior samples of $\beta_s^{(b)}(t)$ for all $s \in C_r$ with size n_r. For example, the mean estimates $\widehat{\beta_r}(t) = \sum_{b=1}^{B} \sum_{s \in C_r} \beta_s^{(b)}(t)/(n_r B)$ and the associated 95% confidence interval can be drawn using the $(2.5, 97.5)\%$ quantiles accordingly. Similarly, we can obtain the posterior samples of $\widehat{Y}_{ijt}$ based on $\beta_s^{(b)}(t)$'s for each age group j, and transform it back using (2) to obtain the posterior estimates of the fertility rates $\widehat{F}_{ijt}$ together with 95% confidence bands.

5 Analysis of Portuguese Regional Fertility Rate Evolution

Using Portuguese data for $N = 28$ (NUTS III) regions,[4] $T = 19$ years from 1991 to 2009, and $J = 7$ groups,[5] we fit the spatio-temporal mixed-effects model (3) by proceeding with 3 Markov Chain Monte Carlo (MCMC) runs with distinct initial values, and $6,000$ total iterations per chain. The convergence is well committed after $5,000$ iterations in that the Potential Scale Reduction Factor (PSRF) $\sqrt{R} < 1.2$ for all parameters (Brooks and Gelman, 1998). The final $1,000$ samples for each chain are then used as posterior samples.

Further, we implement the spatial functional clustering model (8) in the intercept only case ($p = 1$) for each of the $J = 7$ age groups. Since the implementation of wavelet smoothing using Haar basis requires the length of the curve to be a power of 2, we choose the first $T = 16$ years out of the 19, which are representative of the overall pattern and temporal evolution that is required for clustering (Figure 2); the remaining 3 years data are retained for evaluation of out of sample forecasts. Then we conduct 5 MCMC runs with a total of $60,000$ iterations per chain for each

[4] The Nomenclature of Units for Territorial Statistics (NUTS) is the geocode standard for referencing the subdivisions of countries for statistical purposes, developed and regulated by the European Union (EU). It covers the member states of the EU and is used in the European Union's Structural Fund delivery mechanisms. For each EU member country, a hierarchy of three NUTS levels is established by Eurostat (the statistical agency for the EU). The current NUTS classification lists 97 regions at NUTS I, 270 regions at NUTS II and 1294 regions at NUTS III level; NUTS III represents regional classification at the finest spatial level. Our regions refer to the 28 NUTS III statistical regions in Portugal. Namely: Alentejo Central, Alentejo Litoral, Algarve, Alto Alentejo, Alto Minho, Alto Trás-os-Montes, Ave, Baixo Alentejo, Baixo Mondego, Baixo Vouga, Beira Interior Norte, Beira Interior Sul, Cávado, Cova da Beira, Dão-Lafões, Douro, Entre Douro e Vouga, Grande Lisboa, Grande Porto, Lezíria do Tejo, Médio Tejo, Oeste, Península de Setúbal, Pinhal Interior Norte, Pinhal Interior Sul, Pinhal Litoral, Serra da Estrela, and Tâmega.

[5] Seven quinquennial age-groups of women in their reproductive ages: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44 and 45-50 years old. Childbirth is a rare event, and the highest fertility rates occur in age-groups 25-29 and 30-34 years olds, followed by 20-24 years old (see Figure 2). Hence these three age-groups are the most crucial in determining the number of children born.

age group $j = 1, 2, \ldots, J = 7$. The 5 chains start with distinct values of d and corresponding randomly selected cluster centers, which yield distinct clustering configurations and the associated parameters. Each chain with the total of $60,000$ iterations takes around 9 minutes to complete on a regular 32MB RAM machine. The convergence of all 5 chains is well committed after $50,000$ iterations by monitoring the PSRF for parameters and the deviance statistics based on both the marginal likelihood and full model likelihood. We then stack the final $10,000$ samples for each of the 5 chains, and sample at every 10 iterations to obtain a total of $B = 5,000$ samples for posterior inference.

5.1 Estimates

Table 1 reports the posterior distribution of the number of clusters d from the total of $5,000$ posterior samples. The estimated clustering configuration varies somewhat across the $J = 7$ age groups in terms of both the posterior mode and spread of d. Figure 3 plots the central clustering configuration $\widehat{\varpi}_j$ based on the estimated adjacency matrix $\widehat{W}_j$ for the j-th age group. Correspondingly, the observed curves of fertility rates F_{ijt} for each cluster, together with the estimated mean curve of fertility rates $\widehat{F}_{ijt}$ and the 95% confidence bands for each cluster, are shown in Figures 4 and 5, with cluster labels and colors matched to the central clustering configurations in Figure 3. For each cluster in each age group in Figures 4 and 5, we also report the estimated Variation Rate (VR)

$$\text{VR} = \left(\widehat{F}_{ijT} - \widehat{F}_{ij1}\right) / \widehat{F}_{ij1} \times 100\%,$$

the percentage change in estimated fertility rates in each cluster over the 16 years 1991-2006. There is substantial spatial variation in the trends of fertility change over time, reflected in the optimal cluster configuration reflecting between 4 to 8 clusters in most age-groups, except the final age-group, $j = 7$: 45 to 50 years old, which has the smallest fertility rates and exhibits a flat pattern over time. The posterior mass of d centers at 2 for this age group and the configuration suggests that the observed curves can be generally grouped by North and South regions. The fitted curves suggests there is a slightly decreasing trend for the North regions, while a slightly increasing trend for the South regions, although the overall trends are flat for both groups.

The optimal representation of temporal trends generally corresponds to a number of clusters d varying from 2 to 8. The evolution of fertility rates by age group ($j = 1, 2, 3, 4, 5, 6$), over the period between 1991 and 2006, as shown in Figures 4 and 5, generally follows two types of dynamics: i) a decreasing trend of the fertility rates in the lower age groups, which encompass women from 15 to 29 years old ($j = 1, 2, 3$); and ii) an increasing trend of fertility rates in the higher age groups, in particular those between 30 to 44 years old ($j = 4, 5, 6$). The final age group ($j = 7$), is an exception presenting two clusters with opposite trends. For all the 7 age groups, the clustering configuration can be generally summarized as South and North areas, which again coincides with the general regional economic conditions. In particular, the most representative fertility age groups with the richest fertility rates, $j = 2, 3, 4$ from age 20 to 34 years old, all exhibit clusters that center around

the capital Grande Lisboa and Península de Setubal, including more urbanized, richer areas with better economic conditions. For women who are aged below 30, decreasing trends for the regions in South area around the capital are more steady. For example, for the two richest fertility groups 20-24 and 25-29 in Figure 4 for these regions, both labeled as cluster 4, the reduction in fertility rates from 1991 to 2006 is smaller compared to other clusters, plus there is a visible bump near 1999 and 2000 which is manifested in both the fitted cluster-specific mean curves. On the other hand, for the age group above 30, the increasing trend for ages above 30 for these areas is steeper, particularly for the 35-39 age group where Grande Lisboa (clsuter 1) and Península de Setubal (cluster 2) form clusters distinct from the southern area cluster due to their sharp increments. For the 40-44 age group, Grande Lisboa constitutes a distinct cluster due to its steady increasing trend after 1998. Such varying fertility trends are well detected by our functional clustering model in that they typically form a singleton cluster. Another example is Serra da Estrela (cluster 3) for age group 15-19, which has distinct temporal trends compared to the other regions.

Table 1. Posterior probabilities of the number of clusters d for the functional clustering model for each age group. The bold value corresponds to the posterior mode, an estimates of d used for extracting the central clustering configuration.

d	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-50**
1							0.006
2							**0.828**
3		0.003	0.012	0.185		0.074	0.142
4		**0.525**	0.305	**0.413**	0.056	0.111	0.017
5		0.307	**0.594**	0.045	**0.741**	**0.532**	0.004
6	0.041	0.164	0.089	0.174	0.197	0.131	0.002
7	0.440	0.000		0.103	0.006	0.066	0.001
8	**0.493**			0.078	0.000	0.081	
9	0.026			0.002		0.005	
10	0.000					0.001	

Since the functional clustering model is applied to the first 16 years of data for the j-th age group, the results are not exactly comparable with the spatio-temporal mixed-effects model (STM) in (3) which is applied to the full data observed over 19 years for all $J = 7$ age groups. We plug in the estimated adjacency matrix $\widehat{W}_j$ from each age group j to investigate if it can potentially improve the STM model fit but more importantly, offer key insights into fertlity diffusion. The parameter estimates and model assessment for STM are summarized in Tables 2 and 3. Several case are presented: the non-random case (with no spatial or temporal dynamics), spatial-only case, temporal-only case, full model with the natural adjacency matrix and estimated adjacency $\widehat{W}_j$ matrices from each sub age-group $j = 1, 2, \cdots, J = 7$ using the functional clustering techniques. The estimated $\widehat{W}_2$ based on SFC for $j = 2$: 20-24 offers the best model fit, that is, the STM with the lowest DIC_4.

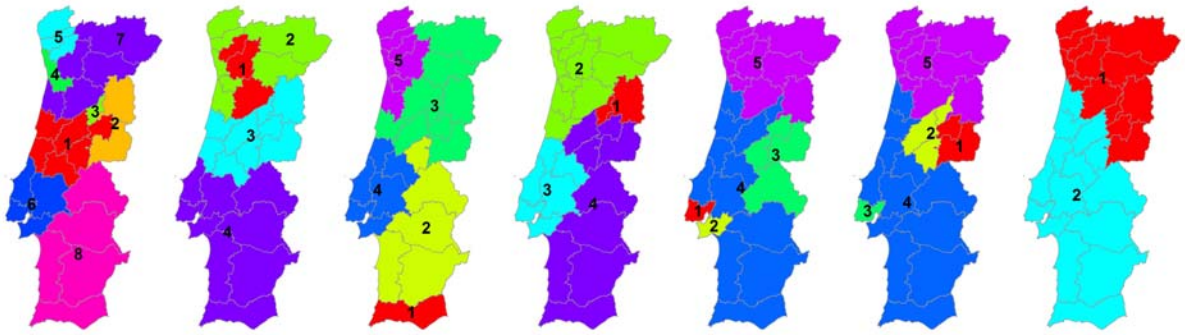

Figure 3: Estimated central clustering configurations by quinquennial age group j ($j = 1, \ldots, 7$: ages 15 to 49 years old, from left to right), with cluster labels.

5.2 Fertility diffusion

Moving now to a more detailed discussion of the estimated patterns of spatial diffusion (Figures 3, 4 and 5), a sharp contrast between the North and South of Portugal is immediately evident across all the different age-groups. Some further features of these cluster configurations are described below.

- 15-19 age group (8 clusters)
  All clusters show a decrease in the fertility rates, with large variation rates (a decrease around 40%) in the northern region (clusters 1, 2, 5 and 7) and southern region (cluster 8), while there is a more stable pattern (variation rates between 10% and 20%) in metropolitan Areas of Porto and Lisbon (clusters 4 and 6) and around 20% for Serra da Estrella (cluster 3). This spatial structure shows a sharp decrease of fertility rates for teenagers somewhat balanced by fertility behavior of immigrants concentrated in metropolitan areas.

- 20-24 age group (4 clusters)
  In clusters 1 to 3 (northern region) the fertility continuously decreased (more than 50%), whereas in cluster 4 (metropolitan Area of Lisbon and the south) the decline is moderated by some stability in the middle period (1995 to 2000).

- 25-29 age group (5 clusters)
  This age group is characterized by initially the highest fertility but moderate decline over the period, ranging between 8% and 28% over the 5 clusters. There was a stronger decline in the initial and final years, while in the quinquennium between 1995 and 2000 the evolution of fertility rates was steadier. The overall decline was between 7% and 16% in the south and the metropolitan area of Lisbon (clusters 1, 2 and 4), where the decline in initial and final years aws puncutated by a period in the middle (1995-2000) with a marginal increase in fertility. By contrast, the north (clusters 3 and 5) recorded sharper decline between 20% and 30%.

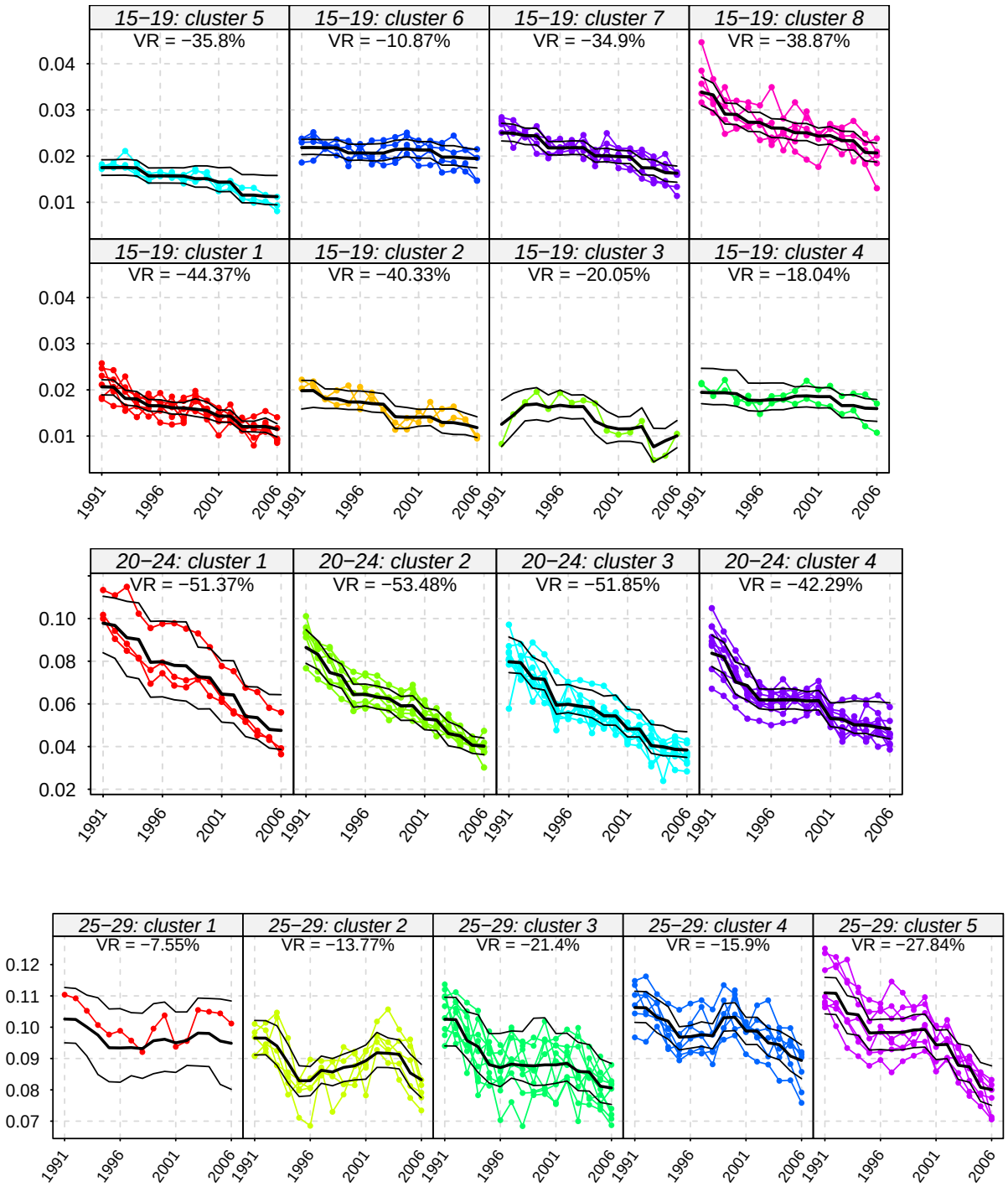

Figure 4: Functional clustering groups of fertility rates by quinquennial age group $j = 1, 2, 3$: 15 to 29 years old, together with variation rate (VR).

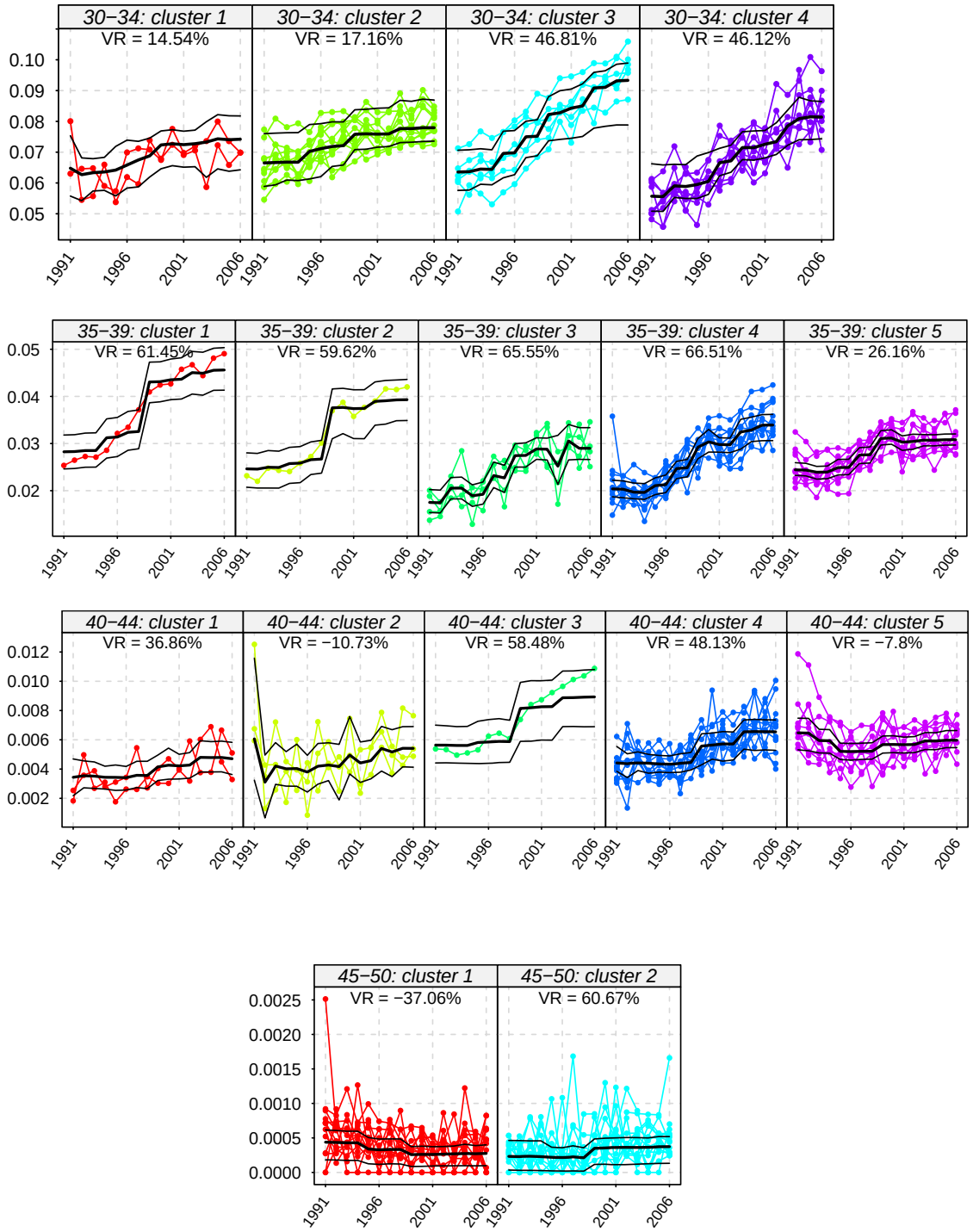

Figure 5: Functional clustering groups of fertility rates by quinquennial age group $j = 4, 5, 6, 7$: 30 to 50 years old, together with variation rate (VR).

- 30-34 age group (4 clusters)
  In this age group, the increase of the fertility rates between 1991 and 2006 was by and large continuous. However two distinct dynamic patterns can be clearly identified: clusters 1 and 2 (northern regions) with a more modest fertility rate increase (between 14% and 18%); and clusters 3 (Lisbon metropolitan Area) and 4 (the south together with the interior central region) with a stronger increase (growth rates around 46%). However, the growth in Lisbon was on a higher base fertility rate than teh south (cluster 4).

- 35-39 age group (5 clusters)
  Comparing the 5 clusters in this age group, besides a general increase of the fertility rates, the behavior of this evolution is clearly different. Clusters 1 and 2, both in the metropolitan area of Lisbon, are characterized by higher fertility rates together with rapid increase of around 60%. By contrast, the north (cluster 5) showed a relatively flat pattern with increase of 26%. The south (cluster 3) and the interior central (cluster 4) showed a considerable growth over the period (65-67%), but on a lower base than clusters 1 and 2.

- 40-44 age group (5 clusters)
  In this age group there is a general increase of the fertility rate with the exception of cluster 5 (North region) which has a slight decrease (8%). Cluster 2 (central region) also has a negative VR value, however this is caused by the large initial fertility rate in 1991, while overall it shows a steadily increasing trend after a dramatic drop in 1992. The metropolitan area of Lisbon (cluster 3) and the south region (cluster 4) are where the growth rate were higher (58% and 48%, respectively).

- 45-49 age group (2 clusters)
  Finally, the last age group defined by two distinct clusters, is characterized by lower fertility rates that are relatively homogenous across regions. Nevertheless, northern regions (cluster 1) had a slight decrease, while in the southern regions (cluster 2) the fertility rate in this age group was stable with an upward step in the middle of the period.

The above results offer very interesting insights on the patterns of spatial diffusion in fertility behavior. First, the spatio-temporal mixed-effects model (STM) based on a natural weights matrix based on spatial contiguity of the regions offers good description of regional fertility variation and time trends; this model fits the data better than a model with no dynamics, but also models with only spatial dynamics or only temporal dynamics. However, the estimated weights matrix $\widehat{W}_2$, based on spatial functional clustering for the age-group 20-24 years old ($j = 2$) offers the best fit to the data. This shows that spatial diffusion can be better explained by an estimated weights matrix, than by an adjacency matrix based solely on geographical contiguity.

Table 2. Posterior inference for STM with the 3 partial implementations and estimated adjacency matrix $\widehat{W}_j$ for $j = 1$. Parameter estimates include posterior mean (95% highest probability density set) from the 3,000 samples.

	No dynamics	**Spatial-only**	**Temporal-only**	**Spatio-Temporal**	**STM($\widehat{W}_1$)**
μ_1	9.92 (9.76,10.09)	9.94 (9.67,10.21)	10.04 (9.82,10.25)	10.01 (9.50,10.57)	10.00 (9.41,10.53)
μ_2	18.51 (18.35,18.68)	18.54 (18.27,18.80)	18.64 (18.43,18.85)	18.61 (18.11,19.17)	18.60 (18.02,19.14)
μ_3	20.64 (20.47,20.80)	20.66 (20.38,20.93)	20.76 (20.55,20.97)	20.73 (20.21,21.29)	20.72 (20.14,21.24)
μ_4	15.47 (15.31,15.63)	15.49 (15.22,15.77)	15.59 (15.38,15.79)	15.56 (15.05,16.11)	15.56 (14.96,16.09)
μ_5	8.89 (8.72,9.05)	8.90 (8.63,9.18)	9.01 (8.80,9.23)	8.98 (8.47,9.53)	8.97 (8.37,9.50)
μ_6	4.20 (4.04,4.36)	4.21 (3.94,4.49)	4.32 (4.12,4.53)	4.29 (3.77,4.83)	4.28 (3.69,4.84)
μ_7	1.25 (1.08,1.42)	1.27 (1.00,1.54)	1.37 (1.17,1.59)	1.34 (0.83,1.90)	1.33 (0.73,1.86)
β_1	−0.14 (−0.16,−0.13)	−0.15 (−0.17,−0.13)	−0.16 (−0.17,−0.14)	−0.17 (−0.22,−0.12)	−0.15 (−0.19,−0.11)
β_2	−0.35 (−0.37,−0.34)	−0.36 (−0.38,−0.34)	−0.37 (−0.38,−0.35)	−0.38 (−0.43,−0.33)	−0.36 (−0.40,−0.32)
β_3	−0.17 (−0.18,−0.15)	−0.17 (−0.20,−0.15)	−0.18 (−0.20,−0.16)	−0.19 (−0.24,−0.15)	−0.18 (−0.22,−0.13)
β_4	0.17 (0.16,0.19)	0.17 (0.14,0.19)	0.16 (0.14,0.18)	0.15 (0.10,0.19)	0.17 (0.12,0.21)
β_5	0.18 (0.16,0.19)	0.17 (0.15,0.20)	0.17 (0.15,0.18)	0.15 (0.10,0.20)	0.17 (0.13,0.22)
β_6	0.05 (0.04,0.07)	0.05 (0.02,0.07)	0.04 (0.02,0.06)	0.03 (−0.02,0.07)	0.04 (0.00,0.09)
β_7	−0.00 (−0.01,0.01)	−0.01 (−0.03,0.02)	−0.01 (−0.03,0.00)	−0.02 (−0.07,0.02)	−0.01 (−0.05,0.03)
δ^2	0.86 (0.82,0.89)	0.61 (0.58,0.65)	0.58 (0.55,0.61)	0.57 (0.55,0.60)	0.57 (0.55,0.60)
τ^2	0.00 (−0.08,0.08)	0.42 (0.33,0.51)	0.20 (0.15,0.26)	0.34 (0.24,0.44)	0.92 (0.66,1.18)
ϕ	0.00 (−0.08,0.08)	0.00 (−0.08,0.08)	0.92 (0.89,0.94)	0.93 (0.91,0.95)	0.94 (0.91,0.95)
γ	0.00 (−0.08,0.08)	0.94 (0.90,0.97)	0.00 (−0.08,0.08)	0.97 (0.95,0.99)	0.95 (0.91,0.98)
$\overline{D(\boldsymbol{\theta})}$	9990.89	9192.32	8279.49	7786.06	7942.31
p_{D4}	14.63	16.41	16.19	16.50	16.49
DIC_4	10005.53	9208.73	8295.69	7802.56	7958.80

Second, there has historically been a north-south divide within Portugal, reinforced by religious, cultural and political differences, as well as by regional economies. This difference is strongly emphasized by the clustering pattern. The rural and poorer northern regions (such as Tâmega and Trás os Montes) traditionally had higher fertility, but this feature has over time moved to richer and urban areas around Lisbon (Península de Setúbal, Grande Lisboa). The metropolitan area of Lisbon together with the southern regions constitute the South, and they show depressed

demographic dynamics. At the same time, the northern regions together with the central belt and particularly regions in the interior (Serra da Estrella, Beira Interior Norte, Beira Interior Sul, Cova da Beira, Pinhal Interior Norte and Pinhal Interior Sul) have attained a devastating vicious cycle of low fertility, outward migration and poor economic growth.

Table 3. Posterior inference for STM with estimated adjacency matrix $\widehat{W}_j$ for $j = 2$ to 7. Parameter estimates include posterior mean (95% highest probability density set) from the 3,000 samples.

	STM($\widehat{W}_2$)	**STM**($\widehat{W}_3$)	**STM**($\widehat{W}_4$)	**STM**($\widehat{W}_5$)	**STM**($\widehat{W}_6$)	**STM**($\widehat{W}_7$)
μ_1	10.20 (9.70,10.78)	10.16 (9.66,10.62)	9.98 (9.45,10.55)	10.00 (9.41,10.53)	10.08 (9.53,10.60)	10.13 (9.61,10.70)
μ_2	18.80 (18.30,19.37)	18.76 (18.25,19.23)	18.58 (18.05,19.14)	18.60 (18.02,19.14)	18.68 (18.13,19.19)	18.73 (18.20,19.28)
μ_3	20.92 (20.41,21.50)	20.88 (20.38,21.35)	20.70 (20.17,21.26)	20.72 (20.14,21.24)	20.80 (20.27,21.32)	20.85 (20.32,21.42)
μ_4	15.76 (15.26,16.34)	15.71 (15.21,16.19)	15.54 (14.99,16.10)	15.56 (14.96,16.09)	15.63 (15.10,16.15)	15.69 (15.16,16.24)
μ_5	9.17 (8.66,9.75)	9.13 (8.63,9.59)	8.95 (8.40,9.51)	8.97 (8.37,9.50)	9.04 (8.49,9.56)	9.10 (8.57,9.65)
μ_6	4.48 (3.99,5.05)	4.44 (3.93,4.90)	4.26 (3.71,4.82)	4.28 (3.69,4.84)	4.36 (3.83,4.88)	4.41 (3.89,4.96)
μ_7	1.53 (1.02,2.10)	1.49 (1.00,1.95)	1.31 (0.78,1.88)	1.33 (0.73,1.86)	1.41 (0.87,1.93)	1.46 (0.95,2.02)
β_1	−0.16 (−0.20,−0.12)	−0.16 (−0.21,−0.13)	−0.17 (−0.22,−0.12)	−0.15 (−0.19,−0.11)	−0.16 (−0.21,−0.12)	−0.17 (−0.21,−0.14)
β_2	−0.37 (−0.42,−0.33)	−0.38 (−0.42,−0.34)	−0.38 (−0.43,−0.33)	−0.36 (−0.40,−0.32)	−0.37 (−0.42,−0.33)	−0.38 (−0.42,−0.35)
β_3	−0.19 (−0.23,−0.14)	−0.19 (−0.23,−0.15)	−0.19 (−0.25,−0.14)	−0.18 (−0.22,−0.13)	−0.18 (−0.23,−0.14)	−0.20 (−0.23,−0.16)
β_4	0.15 (0.11,0.20)	0.15 (0.11,0.19)	0.15 (0.10,0.20)	0.17 (0.12,0.21)	0.16 (0.11,0.20)	0.14 (0.11,0.18)
β_5	0.16 (0.12,0.20)	0.16 (0.11,0.20)	0.15 (0.10,0.21)	0.17 (0.13,0.22)	0.16 (0.12,0.21)	0.15 (0.11,0.18)
β_6	0.03 (−0.01,0.08)	0.03 (−0.01,0.07)	0.03 (−0.02,0.08)	0.04 (0.00,0.09)	0.04 (−0.01,0.08)	0.02 (−0.01,0.06)
β_7	−0.02 (−0.06,0.02)	−0.02 (−0.07,0.02)	−0.02 (−0.08,0.03)	−0.01 (−0.05,0.03)	−0.01 (−0.06,0.03)	−0.03 (−0.06,0.00)
δ^2	0.57 (0.54,0.60)	0.57 (0.55,0.60)	0.57 (0.54,0.60)	0.57 (0.55,0.60)	0.57 (0.54,0.60)	0.57 (0.55,0.60)
τ^2	0.52 (0.38,0.67)	0.73 (0.53,0.94)	0.85 (0.61,1.10)	0.92 (0.66,1.18)	0.83 (0.61,1.06)	1.66 (1.20,2.12)
ϕ	0.93 (0.91,0.95)	0.93 (0.91,0.95)	0.94 (0.92,0.96)	0.94 (0.91,0.95)	0.93 (0.91,0.95)	0.94 (0.91,0.95)
γ	0.97 (0.95,0.98)	0.95 (0.92,0.98)	0.96 (0.93,0.99)	0.95 (0.91,0.98)	0.96 (0.93,0.98)	0.96 (0.93,0.99)
$\overline{D(\boldsymbol{\theta})}$	7773.46	7925.02	7878.33	7942.31	7901.21	7923.85
p_{D4}	15.91	17.00	16.34	16.49	16.60	17.07
DIC_4	7789.37	7942.02	7894.68	7958.80	7917.82	7940.92

Third, the clusters offer valuable insights into the nature of spatial diffusion, which in turn is not solely based on geography, but also socio-cultural linkages. As discussed above, the clusters are estimated separately, by spatial functional clustering (SFC) of curves, for each of the seven age-groups. However, since the spatio-temporal mixed-effects model (STM) based on estimated $\widehat{W}_2$ (for the age-group $j = 2$) offers the best fit with the data, and since most of the other clustering profiles

are quite similar, we discuss diffusion based mainly on the age-group 20-24 years. Figure 3 shows that the two major cities, Lisbon and Porto, each drive fertility behavior not only in their own hinterlands but across large areas – cluster 4 in the south and cluster 2 in the north, respectively. Thus, these two cities are extremely important centres, not only economically, but also culturally. This observation has tremendous implications for policy. Then, change in behavior in Porto is likely to have spillover effects across a large part of the north, and likewise Lisbon for the south. The third large cluster (cluster 3) spans the central region, from the Atlantic coast in the west to the Spanish border in the east. The cultural centre for this cluster may be the oldest university city in the country, Coimbra, and perhaps the industrial city Aveiro. Once again, changes made here is expected to have spillovers across the central cluster. The final cluster, cluster 1, encompasses poor rural areas of Dão-Lafões, Tâmega and Ave, and does not seem to have a natural social or clultural centre. However, based on the clustering patterns for $j = 3$ and $j = 4$, one can anticipate that these regions may also be somewhat influenced by Porto. Such insights into the true, rather than assumed patterns of fertility diffusion offer very rich interpretation of spatial dynamics, and offer unique policy implications.

Fourth, an interesting feature is the sharp increase in fertility for older age groups in the Lisbon metropolitan area and the South which occurred around 2000. Though more detailed analysis is required to understand the sources of this change, we can speculate that the phenomenon corresponds to a changing behavior in the middle class older women who decided to have children after their professional career was established. This is an interesting issue for social research and a clear illustration of the power of the presented methodology to detect subtle changes in patterns of fertility behavior.

5.3 Policy implications

The above inferences offer substantial policy implications. Substantial concern has been expressed in policy and government circles on the depressive demographic cycle affecting Portuguese peripheral areas towards the north and the central regions. Castro et al. (2013, 2015) obtain demographic projections that predict that some of these areas, particularly Tâmega and Trás os Montes, may become literally empty by 2030. This is because of fall in fertility rates combined with poor economic opportunities and consequent rapid net out-migration; see also Press Reports (B, C, D). In a strongly worded policy advice, OECD (2011) suggested that *"Portugal should try to invest more of their public family budget towards early year's supports ... Investment on early year's services is essential to enable families to flourish, is essential for future welfare state sustainability and economic growth."* Indeed, Portugues early-, mid- and late- childhood spending lag substantially behind OECD levels. Further, it is suggested that such investment should take the form of childcare support for low-income families and parental leave policies, which are expected to promote sustainable fertility rates, particularly in the Portuguese context of high female labor force participation (over 60% of children live with both parents working full-time).

Our inferences on fertility diffusion can provide valuable insights into how such policies may be implemented. First, the problem clusters are in the north and central regions of the country and this is where investments may be focused. Second, policies may be targeted towards the leading cultural and economic centres within each cluster – Porto in the north, and Coimbra and Aveiro in the central region. Additional focus may be provided to somewhat isolated clusters in specific age-groups, for example, cluster 1 in the north for 20-24 year olds. Likewise, in the eastern periphery of the central region, special attention may be focused on cluster 1 for ages 30-34, cluster 3 for 35-39 and clusters 1 and 2 for 40-45 yer old women. Third, economic opportunities also need to be developed keeping in view the skills availability within these clusters (OECD, 2014). This will reduce out-migration but also potentially encourage migration into the interior of the country.

6 Forecast and Prediction

While our focus in this paper has been on estimation and inference, the ability of the model to generate accurate out-of-sample and in-sample predictions also needs to be examined. Thus, we evaluate the forecast and prediction performance of our proposed approach. We compare the in-sample prediction and out-of-sample forecast performance of our main Spatio-Temporal Mixed-effects model (STM) (Equations 1-3) with two other related models.

6.1 Models

Our main model for the transformed fertility rates Y_{ijt} incorporating the spatial, temporal and group effects is the Spatio-Temporal Mixed-effects model (STM) described in Equations 1-3. This we denote as **Model 1**.

We also include **Model 2** with only temporal variation modelled, that is, $\boldsymbol{D}(\gamma) = I$ in (1) above. Finally, we consider the spatial-temporal (independent random effects) as our **Model 3**:

$$Y_{ijt} \sim \mathcal{N}\left(\alpha_{1,i} + \alpha_{2,t} + \mu_j + \beta_j t,\, \delta^2\right)$$

with spatial random effect $\boldsymbol{\alpha}_1 = (\alpha_{1,1}, \cdots, \alpha_{1,N})'$ and temporal random effect $\boldsymbol{\alpha}_2 = (\alpha_{2,1}, \cdots, \alpha_{2,T})'$:

$$\boldsymbol{\alpha}_1 \sim \mathcal{N}\left(0,\, \tau_1^2 \boldsymbol{D}(\gamma)\right), \quad \boldsymbol{\alpha}_2 \sim \mathcal{N}\left(0,\, \tau_2^2 \boldsymbol{A}(\phi)\right)$$

6.2 Prediction

As discussed above, our data run through 19 years. We fit **Models 1, 2** and **3** to the data on Portuguese age-specific fertility rates $\boldsymbol{Y} = \boldsymbol{\alpha} + \boldsymbol{\mu} + \boldsymbol{\beta t}$ with $t = 1, 2, \cdots, 16$, and make prediction on $\boldsymbol{Y}_0 = \boldsymbol{\alpha}_0 + \boldsymbol{\mu} + \boldsymbol{\beta t}$ for the last 3 time points $t = 17, 18, 19$. Generally, let $\boldsymbol{\theta} = \{\gamma, \phi, \boldsymbol{\mu}, \boldsymbol{\beta}, \tau^2, \delta^2\}$ denote the model parameters, and $\boldsymbol{\alpha}$ be the random effects for each model for fitted data. The

posterior prediction of $\boldsymbol{Y}_0$ is

$$\begin{aligned} p(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}) &= \int p(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{\alpha}_1)\, p(\boldsymbol{\theta}, \boldsymbol{\alpha}_1 \,|\, \boldsymbol{Y})\, d\boldsymbol{\theta}\, d\boldsymbol{\alpha}_1 \\ &= \int \left(\int p(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_0)\, p(\boldsymbol{\alpha}_0 \,|\, \boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{\alpha}_1)\, d\boldsymbol{\alpha}_0 \right) p(\boldsymbol{\theta}, \boldsymbol{\alpha}_1 \,|\, \boldsymbol{Y})\, d\boldsymbol{\theta}\, d\boldsymbol{\alpha}_1 \end{aligned}$$

via composite sampling: with posterior samples $\{\boldsymbol{\theta}^{(b)}, \boldsymbol{\alpha}_1^{(b)}\}_{b=1}^{B}$, we draw $\boldsymbol{Y}_0^{(b)}$ from the inner integration term which can be identified to a $\mathcal{N}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ since $\boldsymbol{\alpha}_0 \,|\, \boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\Sigma}_0 = \Sigma_{00} - \Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10}$ and $\boldsymbol{\mu}_0 = \Sigma_{01}\Sigma_{11}^{-1}\boldsymbol{\alpha}$, and Σ_{ij} is the partition block of the covariance assumed for $(\boldsymbol{\alpha}_0' | \boldsymbol{\alpha}')'$ with $i, j = 0$ for $\boldsymbol{\alpha}_0$ and 1 for $\boldsymbol{\alpha}$. We then have $\boldsymbol{\Sigma}^* = \delta^2 I + \boldsymbol{\Sigma}_0$ and $\boldsymbol{\mu}^* = \boldsymbol{\mu}_0 + \boldsymbol{\mu} + \boldsymbol{\beta t}$.

6.3 Model Comparisons

As before, for model comparisons, we use Deviance Information Criterion (DIC) for mixed-effects model and DIC_4 (Celeux et al 2006) based on complete likelihood. We also report $\overline{D(\boldsymbol{\theta})} = -2\mathbb{E}_1$ as the posterior expected value of the joint deviance, and $p_{D4} = \overline{D(\boldsymbol{\theta})} + 2\mathbb{E}_2$ as a measure of model dimensionality. As discussed in Section 2.4, generally a smaller DIC_4 indicates better predictive power.

To evaluate forecast performance, we also compute the mean squared error (MSE) and relative average deviation (RAD) to evaluate the prediction for Y_k, $k = 1, 2, \cdots, K$, which are, respectively, defined as

$$\text{MSE} = \sum_{k=1}^{K} (\widehat{Y}_k - Y_k)^2 / K, \quad \text{RAD} = \sum_{k=1}^{K} |\widehat{Y}_k / Y_k - 1| / K.$$

6.4 Results

Table 4 reports model estimates and the model comparison statistics for **Models 1, 2** and **3**. The forecast performance of each individual estimated age-sspecific mortality rate for each region are shown graphically in Appendix C. In terms of in-sample predictive densities, as reflected for example in DIC_4, the best performance is obtained for the proposed spatio-temporal mixed-effects model (STM) (**Model 1**). The additive model (**Model 3**) comes out worst of the three models in this measure. This underscores the importance of modeling spatio-temporal dynamics appropriately. In an important departure from the standard literature, we do not take the adjacency structure of the spatial units as a given, but estimate this using spatial clustering; see also Bhattacharjee and Jensen-Butler (2013), Bhattacharjee and Holly (2013) and Castro et al. (2015).

However, exactly the opposite situation emerges when we consider out-of-sample forecast performance. Here, the additive model (**Model 3**) comes best, while the temporal-only model (**Model 2**) is the worst. However, while the additive model has the smallest MSE and RAD in forecasting, the p_{D4} statistic suggests that the model also has the highest complexity. An important insight resulting from this observation is that the most complex over-parametrised model typically has

better forecast performance. At the same time, it is relatively weak in terms of structural interpretation. Hence, such a model may not be readily useful for evaluation of alternate scenerios and policy measures.[6]

Table 4. Posterior inference for the 3 models. Parameter estimates include posterior mean (95% credible set) from the 3,000 samples.

	Full Spatio-temporal **Model 1**	Temporal-only **Model 2**	Additive model **Model 3**
μ_1	$10.24\,(9.64, 10.73)$	$10.00\,(9.77, 10.22)$	$10.13\,(8.46, 12.43)$
μ_2	$18.94\,(18.33, 19.43)$	$18.69\,(18.46, 18.92)$	$18.83\,(17.16, 21.07)$
μ_3	$20.73\,(20.13, 21.21)$	$20.48\,(20.25, 20.71)$	$20.62\,(18.94, 22.86)$
μ_4	$15.72\,(15.11, 16.20)$	$15.48\,(15.25, 15.70)$	$15.62\,(13.95, 17.86)$
μ_5	$9.22\,(8.62, 9.70)$	$8.98\,(8.74, 9.20)$	$9.11\,(7.46, 11.38)$
μ_6	$4.59\,(3.98, 5.06)$	$4.34\,(4.11, 4.57)$	$4.47\,(2.78, 6.69)$
μ_7	$1.63\,(1.03, 2.11)$	$1.39\,(1.17, 1.62)$	$1.52\,(-0.14, 3.78)$
β_1	$-0.16\,(-0.21, -0.13)$	$-0.15\,(-0.17, -0.13)$	$-0.16\,(-0.32, -0.01)$
β_2	$-0.39\,(-0.43, -0.35)$	$-0.37\,(-0.39, -0.35)$	$-0.38\,(-0.54, -0.24)$
β_3	$-0.15\,(-0.19, -0.11)$	$-0.13\,(-0.15, -0.11)$	$-0.14\,(-0.30, 0.00)$
β_4	$0.16\,(0.12, 0.20)$	$0.18\,(0.16, 0.20)$	$0.17\,(0.01, 0.31)$
β_5	$0.15\,(0.11, 0.19)$	$0.17\,(0.15, 0.19)$	$0.16\,(0.00, 0.31)$
β_6	$0.02\,(-0.02, 0.06)$	$0.04\,(0.02, 0.06)$	$0.03\,(-0.13, 0.17)$
β_7	$-0.03\,(-0.07, 0.01)$	$-0.01\,(-0.03, 0.01)$	$-0.02\,(-0.19, 0.12)$
δ^2	$0.56\,(0.53, 0.59)$	$0.57\,(0.54, 0.60)$	$0.62\,(0.59, 0.65)$
τ_1^2	$0.34\,(0.24, 0.46)$	$0.18\,(0.13, 0.24)$	$0.57\,(0.31, 1.01)$
τ_2^2	-	-	$0.49\,(0.02, 1.74)$
γ	$0.97\,(0.94, 0.99)$	$0.00\,(0.00, 0.00)$	$0.77\,(-0.04, 0.99)$
ϕ	$0.93\,(0.91, 0.96)$	$0.91\,(0.88, 0.94)$	$0.88\,(0.38, 0.99)$
$\overline{D(\boldsymbol{\theta})}$	6491.57	6907.71	7452.12
p_{D4}	16.71	16.36	23.49
DIC_4	6508.28	6924.07	7475.62
MSE	1.25	1.36	0.97
RAD	0.16	0.17	0.12

[6] This evidence is also in line with macroeconomic time series data, where structural macroeconomic models, such as DSGE models, typically demonstrate weaker forecast performance compared to more complex models, such as VARs.

7 Conclusion

This paper develops new Bayesian methodology for spatio-temporal modeling of demographic outcomes. The proposed spatio-temporal mixed-effects model (STM) allows for very rich spatial and temporal dynamics, which are assumed separable for mathematical tractability. Importantly, we do not asume spatial diffusion to be driven by geographic proximity, but equally dependent on socio-cultural distances. Inferences on such complex spatial dynamics is based on a spatial functional clustering (SFC) model where time trends of demographic outcomes across regions and groups are modeled; neither the number of clusters nor their boundaries are assumed *a priori*, but inferred from the data.

Applied to Portuguese regional fertility data, the methods uncover exciting insights on the nature of spatial diffusion. This is driven by the socio-economic patterns of historical development of the country and its regions. The results aid identification of peripheral regions with depressive demographic dynamics, and allow for unique design of regional policy for such regions.

The methods extend the recent literature on unknown and endogenous spatial dynamics (Bhattacharjee and Holly, 2013; Bhattacharjee and Jensen-Butler, 2013; Bailey et al., 2015). In line with other recent evidences, for example in Bhattacharjee at al. (2014), we find that estimated spatial diffusion aids the design of effective place based policies, by focusing not only on local benefits from policy, but spillover benefots brought about by diffusion.

References

[1] Alkema, L., Raftery, A., Gerland, P., Clark, S. and Pelletier, F. (2008). Estimating the total fertility rate from multiple imperfect data sources and assessing its uncertainty. Working Paper **89**, Center for Statistics and the Social Sciences, University of Washington.

[2] Alkema, L., Raftery, A., Gerland, P., Clark, S., Pelletier, F., Buettner, T. and Heilig, G. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography* **48**, 815-839.

[3] Alstadt, J. and Getis, A. (2006). Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geographical Analysis* **38**, 327-343.

[4] Assunção, R., Schmertmann, C., Potter, J. and Cavenaghi, S. (2005). Empirical Bayes estimation of demographic schedules for small areas. *Demography* **42**(3), 537-558.

[5] Bailey, N., Holly, S. and Pesaran, M.H. (2015). A Two Stage Approach to Spatio-temporal Analysis with Strong and Weak Cross-sectional Dependence. *Journal of Applied Econometrics* (forthcoming).

[6] Berry, F.S. and Berry, W.D. (1992). Tax innovation in the states: capitalizing on political opportunity. *American Journal of Political Science* **36**, 715-742.

[7] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B* **36**(2), 192-236.

[8] Bhattacharjee, A., Castro, E., Maiti, T. and Marques, J. (2015). Endogenous spatial regression and delineation of submarkets: A new framework with application to housing markets. *Journal of Applied Econometrics* (forthcoming).

[9] Bhattacharjee, A. and Holly, S. (2013). Understanding Interactions in Social Networks and Committees. *Spatial Economic Analysis* **8**(1), 23-53.

[10] Bhattacharjee, A. and Jensen-Butler, C. (2013). Estimation of the Spatial Weights Matrix under Structural Constraints. *Regional Science and Urban Economics* **43**, 617-634.

[11] Bhattacharjee, A., Maiti, T. and Petrie, D. (2014). General equilibrium effects of spatial structure: Health outcomes and health behaviours in Scotland. *Regional Science and Urban Economics* **49**, 286–297.

[12] Bigotte, J.F., Antunes, A.P., Krass, D. and Berman, O. (2014). The Relationship between Population Dynamics and Urban Hierarchy: Evidence from Portugal. *International Regional Science Review* **37**(2), 149-171.

[13] Billari, F.C., Graziani, R. and Melilli, E. (2012). Stochastic population forecasts based on conditional expert opinions. *Journal of the Royal Statistical Society Series A* **175**, 491-511.

[14] Billari, F.C., Graziani, R. and Melilli, E. (2014). Stochastic Population Forecasting Based on Combinations of Expert Evaluations Within the Bayesian Paradigm. *Demography* **51**, 1933-1954.

[15] Blacker, C.P. (1947). Stages in population growth. *Eugenics Review* **39**, 88-101.

[16] Bongaarts, J. and R. Bulatao (1999). Completing the demographic transition. *Population and Development Review* **25**(3), 515-529.

[17] Bongaarts, J. and Watkins, S.C. (1996). Social Interactions and Contemporary Fertility Transitions. *Population and Development Review* **22**, 639-682.

[18] Borgoni, R. and F. Billari (2003). Bayesian spatial analysis of demographic survey data. *Demographic Research* **8**(3), 61-92.

[19] Brooks, S.P. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**(1), 434-455.

[20] Carlin, B.P. and T.A. Louis (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd edition, Chapman and Hall/CRC: Boca Raton FL.

[21] Case, A.C., Lin, I.-F. and McLanahan, S.S. (2003). Explaining Trends in Child Support: Economic, Demographic, and Policy Effects. *Demography* **40**(1), 171-189.

[22] Castro, E.A., Gomes, M.C.S., Silva, C.J. and Martins, J.M. (2012). An inter-regional migration model applied to Portuguese data. Paper presented in the European Population Conference 2012 (Stockholm, Sweden), *Mimeo* (available at http://epc2012.princeton.edu/papers/120824).

[23] Castro, E.A., Zhang, Z., Bhattacharjee, A., Martins, J.M. and Maiti T. (2015). Regional Fertility Data Analysis: A Small Area Bayesian Approach. In: Dey, D.K., Loganathan, A., Singh, U. and Upadhyay, S.K. (Eds.), *Current Trends in Bayesian Methodology with Applications*, Chapter 10, Chapman & Hall/CRC Press (In Press), 205-226.

[24] Cavenaghi, S.M., Potter, J.E., Schmertmann, C.P. and Assunção, R. (2004). Estimating total fertility rates for small areas in Brazil. Paper presented in Population Association of America 2004 Annual Meeting (Boston, MA), Mimeo (available at http://paa2004.princeton.edu/papers/42194).

[25] Celeux, G., Forbes, F., Robert, C.P. and Titterington, D.M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis* **1**(4), 651-673.

[26] Chi, G., Zhou, C. and Voss, P.R. (2011). Small-area Population Forecasting in an Urban Setting: A Spatial Regression Approach. *Journal of Population Research* **28**, 185-201.

[27] Clyde, M. and George, E.I. (2000). Flexible empirical Bayes estimation for wavelets. *Journal Royal Statistical Society Series B* **62**, 681-698.

[28] Clyde, M. and George, E.I. (2003). Discussion on 'Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis' (by J.S. Morris, M. Vannucci, P.J. Brown and R.J. Carroll). *Journal of the American Statistical Association* **98**, 584-585.

[29] Coale, A.J. and Watkins, S.C. (1986). (Eds.) *The Decline of Fertility in Europe*. Princeton, NJ: Princeton University Press.

[30] Cressie, N. and Chan, N.H. (1989). Spatial Modeling of Regional Variables. *Journal of the American Statistical Association* **84**, 393-401.

[31] de Castro, M.C. (2007). Spatial Demography: An Opportunity to Improve Policy Making at Diverse Decision Levels. *Population Research and Policy Review* **26**, 477-509.

[32] Divino, F., Egidi, V. and Salvatore, M.A. (2009). Geographical mortality patterns in Italy: A Bayesian analysis. *Demographic Research* **20**(18), 435-466.

[33] Entwisle, B., Rindfuss, R.R., Walsh, S.J., Evans, T.P. and Curran, S.R. (1997). Geographic information systems, spatial network analysis, and contraceptive choice. *Demography* **34**, 171-187.

[34] Entwisle, B. (2007). Putting people into place. *Demography* **44**(4), 687-703.

[35] Feng, W., Lim, C., Maiti, T. and Zhang, Z. (2014). Spatial Regression and Estimation of Disease Risks: A Clustering based Approach. *Statistical Analysis and Data Mining*, in revision.

[36] Ferguson, T. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics* **1**, 209-230.

[37] Ferrante, M.R. and C. Trivisano (2010). Small area estimation of the number of firms' recruits by using multivariate models for count data. *Survey Methodology* **36**(2), 171-180.

[38] Festy, P. and F. Prioux (2002). *An Evaluation of the Fertility and Family Survey Project*. United Nations: New York and Geneva (available at http://www.unece.org/fileadmin/DAM/pau/_docs/ffs/FFS_2000_Prog_EvalReprt.pdf).

[39] Flippen, C.A. (2010). The Spatial Dynamics of Stratification: Metropolitan Context, Population Redistribution, and Black and Hispanic Homeownership. *Demography* **47**(4), 845-868.

[40] Folmer, H. and Oud, J. (2008). How to get rid of W: a latent variables approach to modelling spatially lagged variables. *Environment and Planning A* **40**, 2526-2538.

[41] Getis, A. and Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. *Geographical Analysis* **36**, 90-104.

[42] Goodchild, M.F., Anselin, L., Appelbaum, R.P. and Harthorn, B.H. (2000). Toward Spatially Integrated Social Science. *International Regional Science Review* **23**(2), 139-159.

[43] Green, P.J. (1995). Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**(4), 711-732.

[44] Guinnane, T.W. (2011). The historical fertility transition: A guide for economists. *Journal of Economic Literature* **49**(3), 589-614.

[45] Gundersen, C. and Ziliak, J.P. (2004). Poverty and Macroeconomic Performance across Space, Race, and Family Structure. *Demography* **41**(1), 61-86.

[46] Harvey, D. (1996). *Justice, nature and the geography of difference*. Oxford: Basil Blackwell.

[47] Isserman, A.M. (1977). The accuracy of population projections for subcounty areas. *Journal of the American Institute of Planners* **43**, 247-259.

[48] Johnson, K.M., Voss, P.R., Hammer, R.B., Fuguitt, G.V. and McNiven, S. (2005). Temporal and spatial variation in age-specific net migration in the United States. *Demography* **42**, 791-812.

[49] Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review* **7**, 579-593.

[50] Knorr-Held, L. and G. Raßer (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **56**(1), 13-21.

[51] Lam, D. and Miron, J.A. (1991). Seasonality of Births in Human Populations. *Social Biology* **38**(5), 1-78.

[52] Land, K.C., Deane, G. and Blau, J.R. (1991). Religious Pluralism and Church Membership: A Spatial Diffusion Model. *American Sociological Review* **56**, 237-249.

[53] Lobao, L. and Saenz, R. (2002). Spatial inequality and diversity as an emerging research area. *Rural Sociology* **67**(4), 497-511.

[54] López-Vizcaíno, E., Lombardía, M.J. and Morales, D. (2013). Multinomial-based small area estimation of labour force indicators. *Statistical Modeling* **13**, 1-26.

[55] Molina, I., Saei, A. and Lombardía, M.J. (2007). Small area estimates of labour force participation under multinomial logit mixed model. *Journal of the Royal Statistical Society Series A* **170**, 975-1000.

[56] Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* **37**(1), 17-23.

[57] Morenoff, J. and Sampson, R.J. (1997). Violent crime and the spatial dynamics of neighborhood transition: Chicago 1970-1990. *Social Forces* **76**, 31-64.

[58] Morris, J.S. and Carroll, R.J. (2006). Wavelet-based functional mixed models. *Journal Royal Statistical Society Series B* **68**, 179-199.

[59] Mur, J. and Paelinck, J.H.P. (2011). Deriving the W-matrix via p-median complete correlation analysis of residuals. *Annals of Regional Science* **47**, 253-267.

[60] Notestein, F.W. (1945). Population: the long view. In: Schultz, T.W. (Ed.), *Food for the World*, University of Chicago Press: Chicago, 36-57.

[61] OECD (2011). Doing Better for Families: Country Report for Portugal. Organisation for Economic Co-operation and Development.

[62] OECD (2014). Job Creation and Local Economic Development: Country Report for Portugal. Organisation for Economic Co-operation and Development.

[63] Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103**, 681-686.

[64] Potter, J., Schmertmann, C., Assunção, R. and Cavenaghi, S. (2010). Mapping the timing, pace and scale of the fertility transition in Brazil. *Population and Development Review* **36**(2), 283-307.

[65] Potter, J., Schmertmann, C. and Cavenaghi, S. (2002). Fertility and development: Evidence from Brazil. *Demography* **39**, 739-761.

[66] Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology* **25**, 111-169.

[67] Raftery, A., Alkema, L., Gerland, P., Clark, S., Pelletier, F., Buettner, T., Heilig, G., Li, N. and Ševčíková, H. (2009). *White Paper: Probabilistic Projections of the Total Fertility Rate for All Countries for the 2010 World Population Prospects*. United Nations (available at http://www.un.org/esa/population/meetings/EGM-Fertility2009/P16_Raftery.pdf).

[68] Raftery, A.E., Madigan, D. and Hoeting, J.A. (1997). Bayesian model averaging for regression models. *Journal of the American Statistical Association* **92**, 179-191.

[69] Ray, S. and Mallick, B.K. (2006). Functional clustering by Bayesian wavelet methods. *Journal of the Royal Statistical Society Series B* **68**, 305-332.

[70] Retherford, R., Ogawa, N., Matsukura, R. and Eini-Zinab, H. (2010). Multivariate analysis of parity progression-based measures of the total fertility rate and its components. *Demography* **47**(1), 97-124.

[71] Ritter, C. and Tanner, M.A. (1992). Facilitating the Gibbs Sampler: the Gibbs Stopper and the Griddy Gibbs Sampler. *Journal of the American Statistical Association* **87**, 861-868.

[72] Sampson, R.J., Morenoff, J.D. and Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology* **28**, 443-478.

[73] Sassen, S. 1994. *Cities in a world economy*. Thousand Oaks, CA: Pine Forge Press.

[74] Schmitt, R.C. and Crosetti, A.H. (1951). Accuracy of the ratio method for forecasting city population. *Land Economics* **27**, 346-348.

[75] Seiver, D.A. (1985). Trend and Variation in the Seasonality of U.S. Fertility, 1947-1976. *Demography* **22**(1), 89-100.

[76] Smith, S. and T. Sincich (1988). Stability over time in the distribution of population forecast errors. *Demography* **25**(3), 461-474.

[77] Smith, S.K. (1987). Tests of forecast accuracy and bias for county population projections. *Journal of the American Statistical Association* **82**, 991-1003.

[78] Thompson, W.C. (1929). Population. *American Journal of Sociology* **34**, 959-975.

[79] Tolnay, S.E. (1995). The Spatial Diffusion of Fertility: A Cross-Sectional Analysis of Counties in the American South, 1940. *American Sociological Review* **60**(2), 299-308.

[80] Tolnay, S.E., Deane, G. and Beck, E.M. (1996). Vicarious violence: Spatial effects on Southern lynchings, 1890-1919. *American Journal of Sociology* **102**, 788-815.

[81] Voss, P.R. (2007). Demography as a Spatial Social Science. *Population Research and Policy Review* **26**, 457-476.

[82] Wachter, K.W. (2005). Introduction to Special feature: Spatial demography. *Proceedings of the National Academy of Sciences* **102**, 15299-15300.

[83] Wall, M.M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference* **121**, 311-324.

[84] Watkins, S.C. (1987). The Fertility Transition: Europe and the Third World Compared. *Sociological Forum* **2**, 645-673.

[85] Weeks, J. (2004). The Role of Spatial Analysis in Demographic Research. In: Goodchild, M.F. and Janelle, D.G. (Eds.), *Spatially Integrated Social Science*, New York: Oxford University Press, 381-399.

[86] Weeks, J.R., Gadalla, M.S., Rashed, T., Stanforth, J. and Hill, A.G. (2000). Spatial Variability in Fertility in Menoufia, Egypt, Assessed Through the Application of Remote Sensing and GIS Technologies. *Environment and Planning A* **32**, 695-714.

[87] Wejnert, B. (2002). Integrating Models of Diffusion of Innovations: A Conceptual Framework. *Annual Review of Sociology* **28**, 297-326.

[88] White, H.R. (1954). Empirical study of the accuracy of selected methods of projecting state population. *Journal of the American Statistical Association* **49**, 490-498.

[89] Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434-449. Reprinted as P. Whittle (2001). On stationary processes in the plane. In D.M. Titterington and D.R. Cox. *Biometrika: One Hundred Years*. Oxford University Press, 293-308.

[90] Wilson,W.J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy.* Chicago: University of Chicago Press.

[91] Zhang, L.-C. and R.L. Chambers (2004). Small area estimates for cross-classifications. *Journal of the Royal Statistical Society* series B **66**, 479-496.

[92] Zhao, Z. and B. Guo (2012). An algorithm for determination of age-specific fertility rate from initial age structure and total population. *Journal of Systems Science and Complexity* **25**(5), 833-844.

Press Reports:

A. *Mail Online* (2012). Portugal's birth rate falls by a FIFTH after cash-strapped country introduces austerity measures. *Daily Mail*, Popular daily newspaper from the United Kingdom. 8 November 2012. http://www.dailymail.co.uk/news/article-2229999/Portugals-birth-rate-falls-lowest-point-60-years.html
Highlights*:*

- *Births in June down by 19 per cent on same time last year*
- *Lowest monthly tally for 60 years*
- *Births on the decline due to economic crisis.*

B. *Expresso* (2013). Interior perderá um terço da população em 2040 (Interior to lose one-third of the population by 2040). *Expresso*, Leading Portuguese newspaper. 26 October 2013. Report by Hugo Franco. http://expresso.sapo.pt/interior-perdera-um-terco-da-populacao-em-2040=f837697
Abstract*: Study predicts a sharp drop of inhabitants in the regions of borders and Trás os Montes, if economic and demographic policies are not changed.*

C. *Público* (2013). Daqui a vinte anos pode ser demasiado tarde para recuperar o interior (Twenty years from now it may be too late to recover the interior of the country). *Público*, a Portuguese daily national newspaper. 13 November 2013.
http://www.publico.pt/sociedade/noticia/daqui-a-vinte-anos-pode-ser-demasiado-tarde-para-recuperar-o-interior-1612415
Abstract*: Study coordinated by the University of Aveiro (UA) indicates that in 2040 the interior of Portugal will have about a third of the current population. The alert comes from a study led by the University of Aveiro (UA): "If there is no migration, the interior will be empty," said the study's coordinator, Eduardo Castro.*

D. *Diário Digital* (2013). Estudo: Interior do país pode perder um terço da população em 30 anos (Study: Interior of the Country can lose a third of the population in 30 years). *SAPO* (Servidor de Apontadores Portugueses), Portugal's largest web portal and internet information company. Report by Por Sandra Gonçalves. 14 November 2013.
http://diariodigital.sapo.pt/news.asp?id_news=668623&fb_action_ids=1020243929845
Abstract*: If the current trend of the evolution of the fertility rate in the country continues and there is no migration, by 2040 the interior of the country ranging from Tras-os-Montes to Alentejo would lose 157,000 inhabitants, about a third of the current population.*

Appendix A: Calculation of the Marginal Likelihood

Consider the marginal likelihood $\pi(\boldsymbol{Y}\,|\,\varpi,\boldsymbol{\gamma})$ in (18), the integrand is a product of the densities: $\mathcal{N}(\boldsymbol{X}_{r,1}\boldsymbol{\beta}_r,\,\sigma^2 I_{n_rT})$'s, $\mathcal{N}(\boldsymbol{0}_{q_r},\,\sigma^2\boldsymbol{\Lambda}_{r,1})$'s and $\texttt{igamma}(a_\sigma,\,b_\sigma)$. After the first integration with respect to $\boldsymbol{\beta}_r$'s, the first two Gaussian densities merge into $\mathcal{N}(\boldsymbol{0}_{q_r},\sigma^2 V_r^{-1})$ with $V_r = I_{n_rT} - \boldsymbol{X}_{r,1}(\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + \boldsymbol{\Lambda}_{r,1}^{-1})^{-1}\boldsymbol{X}'_{r,1}$. By Sherman-Morrison-Woodbury formula and Sylvester's determinant theorem

$$\begin{aligned}(A+UCV)^{-1} &= A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}\\ |A+UCV| &= |C^{-1}+VA^{-1}U||C||A|\end{aligned}$$

and taking $A = I_{n_rT}$, $U = -\boldsymbol{X}_{r,1}$, $C = (\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} + \boldsymbol{\Lambda}_{r,1}^{-1})^{-1}$ and $V = \boldsymbol{X}'_{r,1}$, we have $V_r^{-1} = I_{n_rT} + \boldsymbol{X}_{r,1}\boldsymbol{\Lambda}_{r,1}\boldsymbol{X}'_{r,1}$ with determinant $|I_{n_rT} + \boldsymbol{X}_{r,1}\boldsymbol{\Lambda}_{r,1}\boldsymbol{X}'_{r,1}| = |I_{q_r} + \boldsymbol{\Lambda}_{r,1}\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1}|$. After the second integration with respect to σ^2, which is over $\texttt{igamma}(a_\sigma + NT/2, b_\sigma + \sum_{r=1}^d \boldsymbol{Y}_r V_r \boldsymbol{Y}_r/2)$, we obtain the expression (19), which is recognized as a multivariate t-density for ordered $\boldsymbol{Y}$ by clusters with location $\boldsymbol{0}_{NT}$, scale matrix $(b_\sigma/a_\sigma)\bigoplus_{r=1}^d (I_{n_rT} + \boldsymbol{X}_{r,1}\boldsymbol{\Lambda}_{r,1}\boldsymbol{X}'_{r,1})$ and degree of freedom $2a_\sigma$. Note for intercept-only model, since $\boldsymbol{X}'_{r,1}\boldsymbol{X}_{r,1} = n_r I_{q_r}$ and , we have the simple form $\log\pi(\boldsymbol{Y}\,|\,\varpi,\boldsymbol{\gamma}) = const. - (a_\sigma + NT/2)\log(1 + \sum_{r=1}^d \boldsymbol{Y}_r V_r \boldsymbol{Y}_r/(2b_\sigma)) - \sum_{r=1}^d\sum_{c=1}^T \log(n_r\boldsymbol{\lambda}[c]\boldsymbol{\gamma}[c]+1)/2$.

Appendix B: Construction of Move Types for the Reversible Jump MCMC Algorithm

We consider the following 4 move types for proposing a new spatial clustering configuration ϖ^* given the current configuration ϖ, with the probability of acceptance for ϖ^* under each move type based on our choice of the proposal density function h for the triplet $(\sigma^2, \boldsymbol{\beta}, \boldsymbol{\gamma})$.

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

$$\frac{g(\varpi|\varpi^*)}{g(\varpi^*|\varpi)} = \frac{\mathrm{P(Merge)}(N-d)}{\mathrm{P(Growth)}}, \qquad \frac{\pi(\varpi^*)}{\pi(\varpi)} = \frac{(1-\alpha)^{d+1}(N-d-1)!}{(1-\alpha)^d(N-d)!}$$

   and the acceptance rate is $\min\{1,\, \boldsymbol{r}(1-\alpha)\mathrm{P(Merge)}/\mathrm{P(Growth)}\}$, which is clearly dominated by the marginal likelihood ratio $\boldsymbol{r} = \pi(\boldsymbol{Y}\,|\,\varpi^*,\boldsymbol{\gamma}^*)/\pi(\boldsymbol{Y}\,|\,\varpi,\boldsymbol{\gamma})$, in particular when one specify a tiny α for almost equal prior weights on the number of clusters d, and P(Merge) = P(Growth) for equal chance of proceeding to Growth and Merge step.

2. Merge step $(d+1, G_{d+1}) \to (d, G^*_d)$: We delete one existing cluster and merge its members into other existing clusters. First, generate a random variable uniformly r distributed on $\{1,\cdots,d+1\}$, 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 to potentially improve the mixing performance of the MCMC. For each region s, we define all regions that are directly connected (i.e., via no third region) to it by latitude or longitude as its neighbors. Among d current cluster centers there are $K(G_d)$ cluster centers that have at least one non-center neighbors. Draw $r \sim \text{Uniform}\{1, \cdots, K(G_d)\}$ to obtain one such cluster center g_r with $R(g_r)$ non-center neighbors. Secondly, draw s from $\{1, \cdots, R(g_r)\}$ uniformly. The sth non-center neighbor becomes the new cluster center g_r^* that replaces g_r in G_d. The acceptance probability is $\min\{1, \boldsymbol{r} K(G_d) R(g_r) / \left(K(G_d^*) R(g_r^*)\right)\}$

4. Switch Step: $(d, G_d) \rightarrow (d, G_d^*)$: We adopt a switch step for exchanging the positions of a randomly selected pair of cluster centers, when the choice of the distance measure introduce distance ties, i.e., certain regions have equal distance from more than 1 cluster centers. We assign such regions according to the cluster center that has smaller index in G_d. Consequently, swapping two cluster centers in this step can result in different cluster memberships and hence the marginal likelihood. The acceptance probability is $\min\{1, \boldsymbol{r}\}$.

Appendix C: Forecast performance of Models 1-3

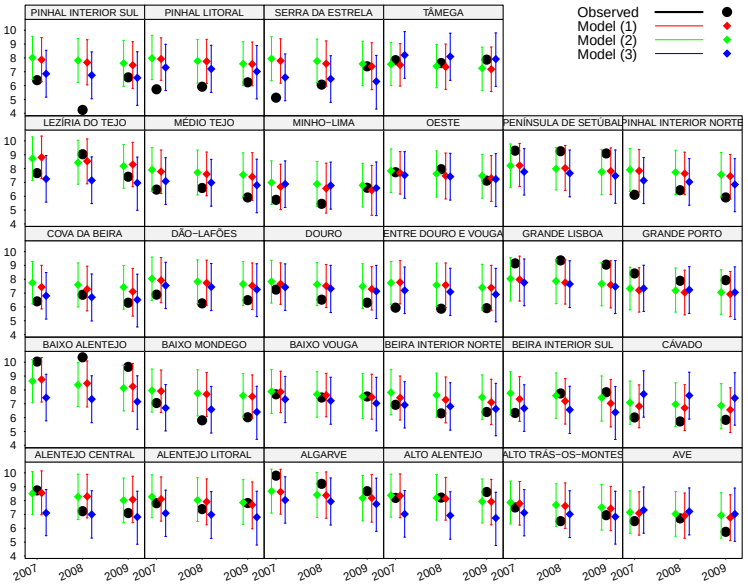

Figure 6: Estimated transformed fertility rates versus observed values for $j = 1$.

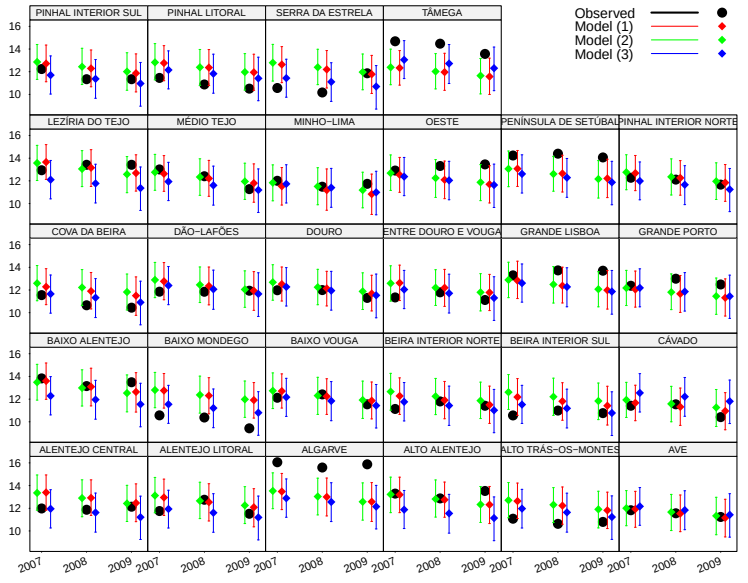

Figure 7: Estimated transformed fertility rates versus observed values for $j = 2$.

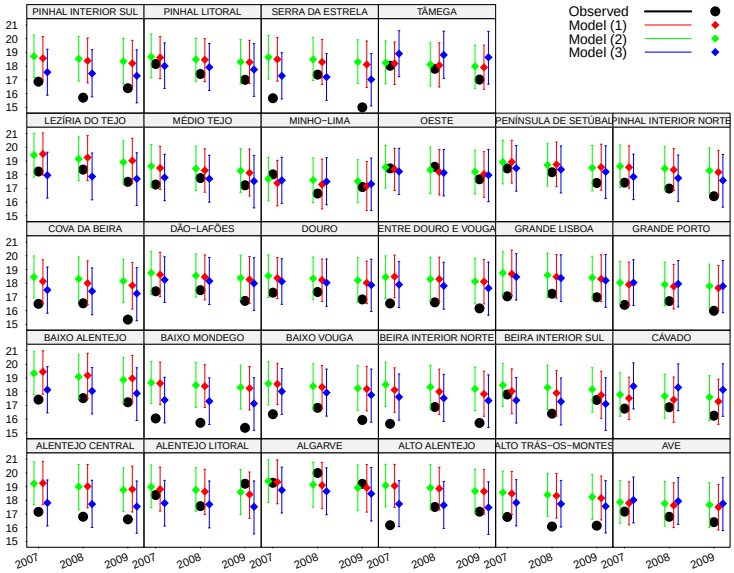

Figure 8: Estimated transformed fertility rates versus observed values for $j = 3$.

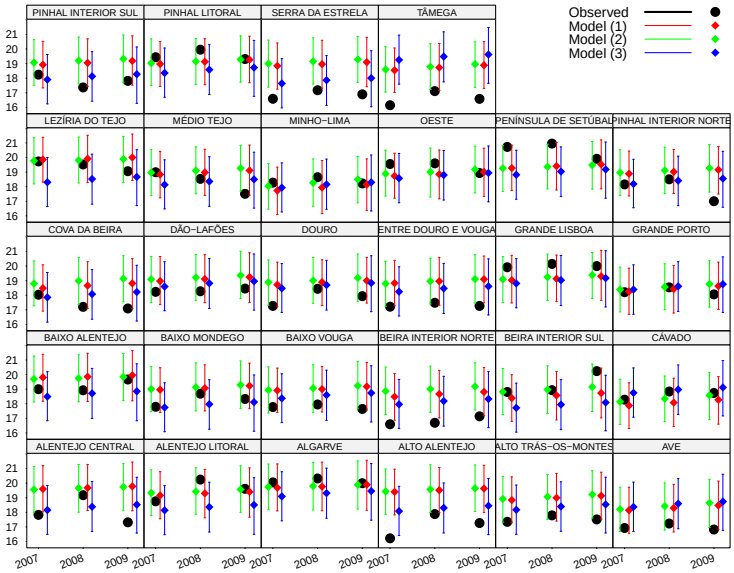

Figure 9: Estimated transformed fertility rates versus observed values for $j = 4$.

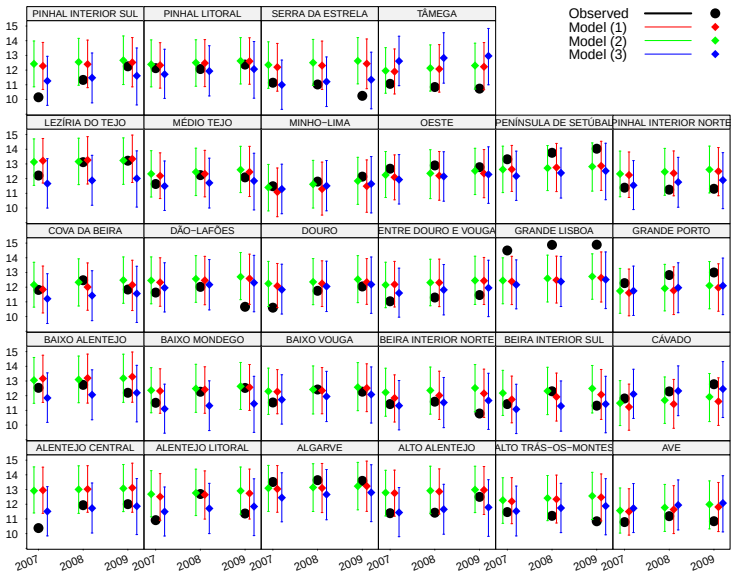

Figure 10: Estimated transformed fertility rates versus observed values for $j = 5$.

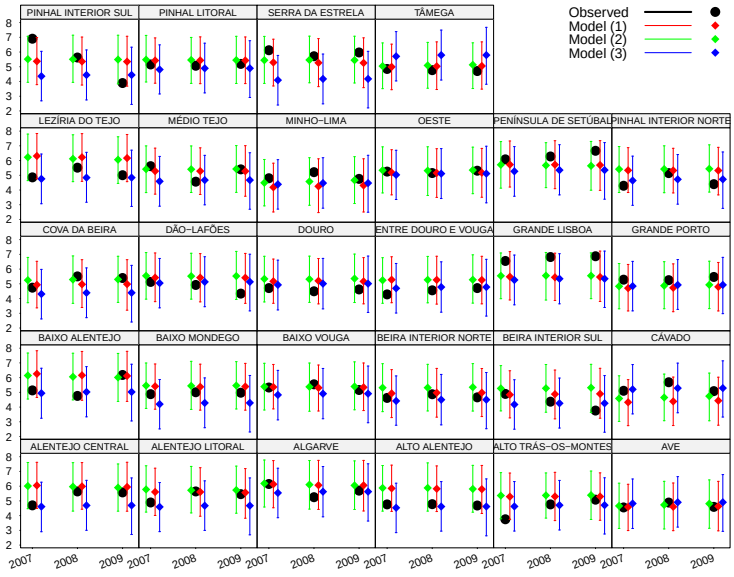

Figure 11: Estimated transformed fertility rates versus observed values for $j = 6$.

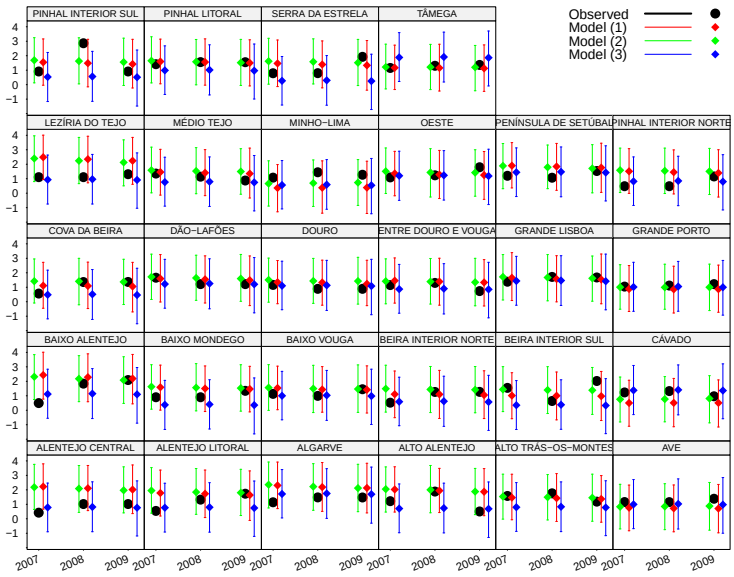

Figure 12: Estimated transformed fertility rates versus observed values for $j = 7$.