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

Zhen Zhang[1], Arnab Bhattacharjee[2], João Marques[3], and Tapabrata Maiti[4]

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 that is typically lacking in classical analyses of large data sets in which 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 has key demographic interpretations. Importantly, the exercise aids identification of peripheral regions with poor demographic prospects and development of regional policy for such places.

Key words: Spatio-temporal modeling; conditional autoregressive model; spatial clustering; bayesian wavelet smoothing; bayesian hierarchical model; age-specific fertility rates.

1. Introduction

[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).

The above quote highlights the central idea behind this article – to emphasize spatial and temporal dynamics as key elements of demographic study. The complexity of spatio-

[1] Eli Lilly and Company, Indianapolis Indiana, U.S.A. Email: zhangquake@gmail.com
[2] Heriot-Watt University and National Institute of Economic and Social Research, Spatial Economics and Econometrics Centre. (SEEC), Mary Burton Building, Edinburgh EH14 4AS, Scotland, United Kingdom. Email: a.bhattacharjee@hw.ac.uk
[3] University of Aveiro, Department of Social, Political and Territorial Sciences, Aveiro, Portugal. Email: jjmarques@ua.pt
[4] Michigan State University, Department of Statistics and Probability, East Lansing Michigan, U.S.A. Email: maiti@stt.msu.edu

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 useful for prediction, for understanding spatial diffusion and for regional demographic and economic policies in the peripheral regions of Portugal.

The article 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; 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 2017), 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 strictly structural explanations of fertility decline, and further, changes in important structural factors such as literacy and industrialization are found to be only weakly related to the timing of fertility decline.

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 Miron 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 fertility has steadily declined as these societies progressed through the third phase of demographic transition, and these declining trends are significantly spatially clustered. In this article, 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 Alkema et al. (2011). 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 developed countries, particularly at a fine spatial level (Bongaarts and Bulatao 1999; Potter et al. 2002). Assunção et al. (2005) developed empirical Bayes methods for small area estimation, where the data from spatially neighboring 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). Since fertility decline is spatially heterogeneous and exhibits 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 article, 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 typically 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 article, 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 article 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 2 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 (Subsection 4.1). In Subsection 4.2, we develop a spatial functional clustering method 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 Subsection 4.3. Section 5 presents evaluation of forecast performance, and Section 6 collects conclusions.

2. Smoothing and Spatial Methods in Demography

The methods for spatial clustering developed in this article 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. 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 1999; Festy et al. 2002; Alkema et al. 2011, 2012). 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 et al. 2010; Zhao and Guo 2012), 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 article. A good example of the first case is the rapid change in fertility rates, with heterogeneous 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. 2016).

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. Moreover, θ in turn can be generated by specific distributions described typically by a lower dimensional vector of 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 Bayes approaches, applied by Cavenaghi et al. (2016) 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 (2010) for a full description of these approaches. Jointly considering multiple factors and dependence structures increases the complexity of statistical inference, which in turn can create challenges for efficiently estimating and describing the uncertainty of an increasing number of modeling parameters. 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). 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. 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 literature 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. In this context, spatial spillovers imply that fertility behavior in one region depends not only on the characteristics of this index region, but also on spillovers of behavior from other regions that are in close proximity. Typically, proximity is measured by a spatial weights matrix that reflects how much each region affects other regions within a connected system. For areal data, the spatial contiguity matrix with binary elements (1 if two regions are contiguous and 0 otherwise) is typically used to measure spatial weights. 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; see Bhattacharjee and Jensen-Butler 2013; Bhattacharjee and Holly 2013; Bailey et al. 2016) for a representative selection, and Bhattacharjee et al. (2014, 2016) 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 article 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 article 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 (2003) and Morris and Carroll (2006). Moreover, one can regionally group such functions (curves) that arise spatially and share common patterns. In 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 here 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 a 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 (rjMCMC) algorithm (Green 1995) for variable-dimensional exploration of the candidate model space; see, for example, Knorr-Held and Raβer (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 clusterspecific 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 Raβer (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 multiple clusters, particularly when the data are high dimensional.

When probability distributions are related to a spatial structure, as expected 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). Here we assume 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 wavelet-based methodology for empirical analysis. In this sense, our work extends wavelet-based functional models in Morris and Carroll (2006), Morris et al. (2006) and Ciarleglio and Ogden (2016).

3. Data and the Portuguese Context

3.1. Portuguese Regional Demographic Context

The methodology developed in this article 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 patterns and trends in fertility. Though Portugal follows

the general trend of the developed countries in 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 per woman) has been consistently breached since 1982.

Even if these changes occurred across 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. Over the period 1991 to 2012, Algarve had the fastest rate of population growth at 1.5 %, followed by Lisbon (0.5 %), the north, Madeira and Acores in the middle (around 0.2 %), while Alentejo had a shrinking population and central Portugal was stagnant.

The spatial heterogeneity described above is combined with 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) studied 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) examined the short term dynamics of fertility rates driven by spatially complex regional contexts of economic growth and decline, using spatial clustering for Bayesian small-area estimation of the fertility rates across the Portuguese regions in 2009. The CAR model was routinely adopted for their region-specific random effect, yet with the adjacency matrix of spatial diffusion estimated via spatial clustering which improved prediction accuracy. Such flexible modeling of spatial diffusion allows identification of regions and clusters where policy initiatives need to be placed. In this article, 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.

3.2. *Portuguese Regional Fertility Data*

The data were provided by the National Statistical Institute of Portugal (INE) (INE 2011), under a protocol between the University of Aveiro (UA) and the Foundation for Science and Technology (FCT). The period for analysis in this work (from 1991 to 2009) captures,

on the one hand, the beginning of the period of fertility stability rates (low values at the level of the national territory), and on the other hand the end of the period of significant changes of regional patterns. Fertility rates have been structurally low in Portugal since 1990, with very different spatial behaviors at NUTS III level (Silva et al. 2011). However, significant changes at this spatial level occurred between 1990 and 2006, with the stabilization of regional fertility patterns since then (Gomes et al. 2016). In 1991, there was a strong concentration of high fertility rates in the north, contrasting with low rates in the south of the country. By 2010, this pattern was reversed, that is, the highest fertility rates were now in the south and the lowest in the north. Between 2006 and 2009, the differences are not substantial and from 2009, the differences are small. Therefore, the above period of analysis considers, on the one hand, the beginning of the period of fertility stability rates (low values at the level of the national territory), and on the other hand the end of the period of significant changes of regional patterns.

It was also decided to consider only the Mainland Portuguese regions, excluding the islands, since the spatial autocorrelation analysis would be questionable if these two non-contiguous autonomous regions were included in the analysis. Note that the spatial weight matrix involved in the CAR model (either assumed as unknown or estimated from the data), considers the territorial influences that fertility rate behavior of a set of neighboring regions have in a specific region. This logic breaks down for islands, which have no land borders.

The Mainland Portuguese NUTS III geographical regions ($N = 28$) regions are shown in Figure 1, with boundary (bold line) based on two clusters identified by spatial clustering in Castro et al. (2015) using data for the year of 2009. 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 very simplistic, because the highest fertility rates have moved from the rural and northern regions to the urban areas and the 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 form the basis for the current article.

Our data comprises the number of births B_{ijt} and the number of women E_{ijt} for $i = 1$, 2,. . ., $N = 28$ regions, $j = 1, 2,. . ., J = 7$ quinquennial age groups from 15 to 49 years old, and $t = 1, 2,. . ., T = 19$ 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 effciency in particular for high-dimensional data. In this article, we model transformed fertility rates Y_{ijt} as a Gaussian mixed-effects model with Bayesian estimation. 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 in Equation (3), and avoid additional computational burden introduced by Poisson regression. Note that when a posterior prediction of the transformed variable
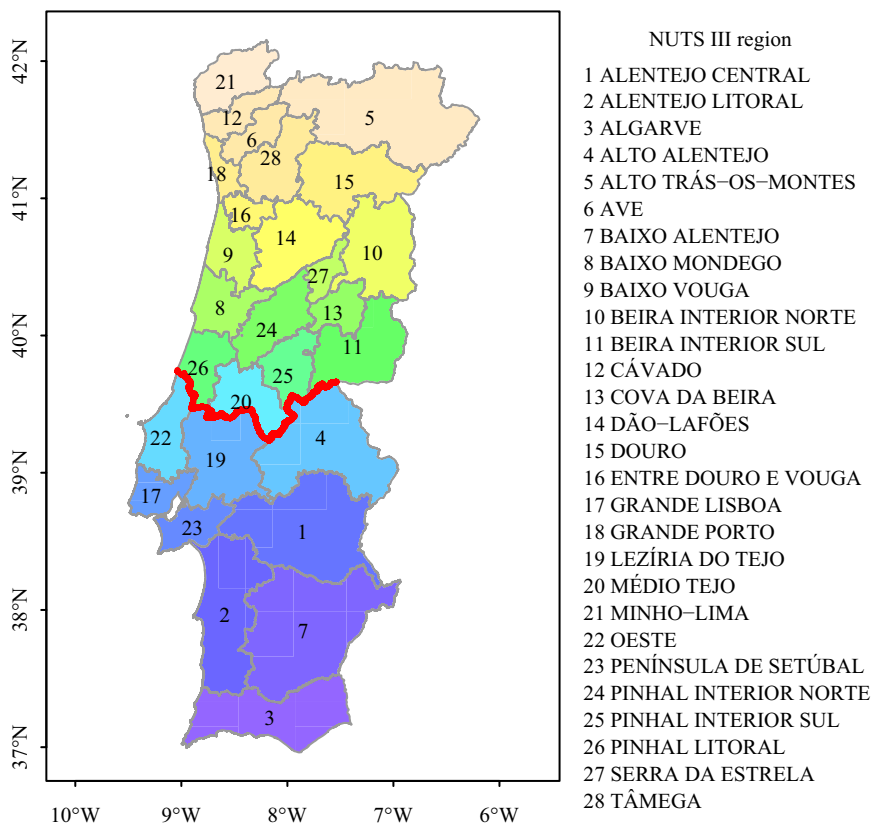
Fig. 1.   Map of the N = 28 Portuguese NUTS III regions.
Note: The Nomenclature of Units for Territorial Statistics (NUTS) is the geocode standard for referencing the subdivisions of countries for statistical purposes, developed and regulated by the European Union (EU). It covers the member states of 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.

$\hat{Y}_{ijt}$ is obtained, one can obtain the corresponding fertility rate $\hat{F}_{ijt}$ by back-transforming $\hat{Y}_{ijt}$ via

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

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

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 the capital city Lisbon, 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 the 20–24 years old age group the region Tamega (yellow) in the north has fertility rates above other regions in the years before 2005.
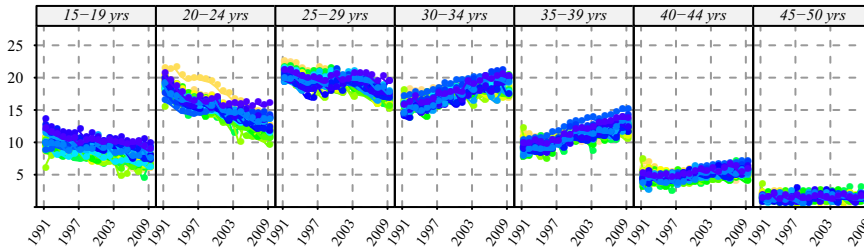
Fig. 2.   *Transformed fertility rates for the J = 7 quinquennial age over T = 19 years.*
Note: Seven quinquennial age-groups of women in their reproductive ages: 15–19, 20–24, 25–29, 30–34, 35–39, 40–44 and 45–49 years old. Childbirth is a rare event, and the highest fertility rates occur in age-groups 25–29 and 30–34 years olds, followed by 20–24 years old). Hence these three age-groups are the most crucial in determining the number of children born.

Generally, there is no visible heteroscedasticity over time and between age groups after the variance stabilization in Equation (1); however, the mean levels and time trends are clearly varying across age groups. The age groups with ages below 29 years generally exhibit a decreasing trend over the 19 years, while the age groups above 30 have an increasing trend, except the final age group (45–49 years old) in which 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, and second, estimation of the diffusion patterns by functional spatial clustering. Table 5 in the Appendix (Section 7), provides some explanation on the notations involved in the model statement.

4.1.   *STM: Spatio-Temporal Mixed-Effects 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)

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

We assume the fixed age-group effects in both the intercept and time trend, and the spatio-temporal random effect $\alpha = \left(\alpha'_1, \alpha'_2, \cdots, \alpha'_T\right)'$, where $\alpha_t = \left(\alpha_{1t}, \cdots, \alpha_{Nt}\right)'$ is the vector of random effects at time t. The age-group effect does not interact with spatial and temporal dependences; while this assumption can be strong in some other applications, it can also be relaxed. 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 (1) Markov property; and (2) Stationarity for both spatial and temporal dependence by specifying

$$\alpha_t = \Phi\alpha_{t-1} + \epsilon_t \tag{4}$$

where $\Phi = diag_{1 \leq s \leq N}\{\phi_s\}$ with $\phi_s \in (-1, 1)$ measures the temporal dependence for each region and $\epsilon_t \sim_{ind} \mathcal{N}(\mathbf{0}_N, \tau_t^2 D(\psi_t))$ with ψ_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 $\psi_t \equiv \psi$; then Equation (4) becomes

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

where "$\otimes$" denotes the Kronecker product of two matrices; We assume autoregressive (AR) structure of order 1 for the temporal dependence with covariance matrix $A(\phi) = [\phi^{n-m}]_{1 \leq m,n \leq T}$, and Conditional Autoregressive (CAR) structure for the spatial dependence with covariance matrix $D(\psi) = (M^{-1} - \psi W)^{-1}$, where both matrices have inverses in closed form, which bring efficiency to computations. 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 $1 \leq i, j \leq N$. 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. 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 $\psi \in (e_N^{-1}, e_1^{-1})$ measures spatial dependence, where e_1 and e_N are the largest and smallest eigenvalues of MW, which ensures that $D(\psi)$ is a valid covariance matrix.

Let I_M denote the identity matrix of dimension M, and 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

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

where the fixed effects $\theta = (\mu_1, \ldots, \mu_J, \beta_1, \ldots, \beta_J)'$ have the design matrix $X = [X_0|X_1]$ with $X_0 = I_J \otimes 1_{NT}$ and $X_1 = I_J \otimes ((1, 2, \ldots, T)' \otimes 1_N)$. The design matrix for the spatio-temporal random effects is $Z = 1_J \otimes I_{NT}$.

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

$$\pi(\theta_k) \propto 1, \quad \pi(\psi) = \text{Uniform}(e_N^{-1}, e_1^{-1}), \quad \pi(\phi) = \text{Uniform}(-1, 1),$$

$$\pi(\delta^2) = \texttt{igamma}(a_\delta, b_\delta), \quad \pi(\tau^2) = \texttt{igamma}(a_\tau, b_\tau). \tag{7}$$

For the prior densities that involve hyper-parameters, 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.

The STM implementation via Gibbs sampling for Bayesian estimation of $\{\theta, \alpha, \delta^2, \tau^2, \psi, \phi\}$ is detailed in Appendix (Subsection 7.1). The model can be implemented efficiently due to conjugate prior for $\{\theta, \alpha, \delta^2, \tau^2\}$ and the convenient form of likelihood for $\{\psi, \phi\}$ under AR(1) and CAR structure. Using the Griddy-Gibbs sampler (Ritter and Tanner 1992) for spatio-temporal dependence ϕ and ψ, multiple terms for the likelihood can be pre-calculated to further facilitate computation.

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

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

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

$$\text{DIC}_4 = -4\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\alpha}}[\log f(\boldsymbol{Y}, \boldsymbol{\alpha}|\boldsymbol{\theta})|\boldsymbol{Y}] + 2\mathbb{E}_{\boldsymbol{\alpha}}[\log f(\boldsymbol{Y}, \boldsymbol{\alpha}|\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{\alpha}])|\boldsymbol{Y}]$$

$$\triangleq -4\mathbb{E}_1 + 2\mathbb{E}_2$$

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 performance.

4.2. SFC: Spatial Functional Clustering

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

We specify a spatio-functional clustering (SFC) model as follows. 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 regression model $Y_i(t) = \sum_{k=1}^p \tilde{x}_{ik}(t)\tilde{\beta}_{rk}(t) + \epsilon_i(t)$ for region i in cluster $C_r, r = 1, 2, \ldots, d$, with p covariates. The kth covariate $\tilde{x}_{ik}(t)$ has effect $\tilde{\beta}_{rk}(t)$ at time point t. The same model at region level is stated as:

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

where Y_i is the $T \times 1$ functional response for i-th region, and ϵ_i is the $T \times 1$ error vector which follows i.i.d. $\mathcal{N}(0, \sigma^2)$; $\tilde{\beta}_r = (\tilde{\beta}_{r1}(t_1), \ldots, \tilde{\beta}_{r1}(t_T), \tilde{\beta}_{r2}(t_1), \ldots, \tilde{\beta}_{rp}(t_T))'$ 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 $\tilde{X}_i = (\tilde{X}_{i1}, \ldots, \tilde{X}_{ip})$ and each $\tilde{X}_{ik}$ is a $T \times T$ diagonal matrix with $\tilde{x}_{ik}(t_1), \ldots, \tilde{x}_{ik}(t_T)$ as diagonal entries. Finally, the model pooling data over all regions can be written as

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

by stacking $\tilde{\beta} = \left(\tilde{\beta}_1', \cdots, \tilde{\beta}_d'\right)'$ for the d clusters.

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 $\beta = Q\tilde{\beta}$ with the block-diagonal matrix $Q = \otimes_{i=1}^{dp} Q$, where each block matrix Q is the Discrete Wavelet Transform (DWT) matrix; that is, we reparametrize $\tilde{\beta}$ by Wavelet transforming each of its T-functional components. Then, the slope vector reduces to $\beta = (\beta_{11}', \ldots, \beta_{1p}', \ldots, \beta_{d1}', \ldots, \beta_{dp}')'$ with $T \times 1$ wavelet coefficients β_{rk} for fixed-effects function $\tilde{\beta}_{rk}(t)$. Furthermore, 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 Equation (9) can now be written as $Y = X\beta + \epsilon$ with $X = \tilde{X}Q'$. Under the clustering structure, the model can be also written at cluster level as

$$Y_r = X_r \beta_r + \epsilon_r, \quad r = 1, 2, \ldots, d, \tag{10}$$

where 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 the product of cluster-specific likelihoods:

$$f(Y|\beta, \sigma^2) = \prod_{r=1}^d \mathcal{N}(X_r \beta_r, \sigma^2 I_{n_r T}) \tag{11}$$

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 $\mathbf{X}_r = (Q, Q, \ldots, Q)' = 1_{nr} \otimes Q'$ and $\mathbf{X}'_r \mathbf{X}_r = n_r I_T$ by the orthonormality of Q.

The spatial contiguity of clusters under SFC is essential for estimating the spatial adjacency $\mathbf{W}$. Developing inferences for $\mathbf{W}$ is a major methodological innovation in the current article. 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 $\varpi = (d, z)$, where d is the number of clusters and $z = (z_1, z_2, \ldots, z_N)$ is the vector containing labels for each region, $z_i \in \{1, 2, \ldots, d\}$. For spatial clustering, we can reduce to $\varpi = (d, G_d)$ where $G_d = (g_1, \ldots, g_d)$ are cluster centers and $g_r \in S$; see, for example, Knorr-Held and Raßer (2000), Zhang et al. (2014) and Feng et al. (2016). The membership z is determined by the minimal distance criterion, that is, regions with the minimal distance from center g_r constitute the cluster C_r.

A clustering prior model is defined as $\pi(\varpi) = \pi(d)\pi(G_d \mid d)$, where $\pi(d)$ is a prior for the number of clusters, d, and $\pi(G_d \mid d)$ 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 $d \leq N_0 \leq N$. 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 exibility to choose any other minimal size that might be relevant to a specific application. Knorr-Held and Raßer (2000) considered a power penalty on the number of clusters d by eliciting $\pi(d \mid \kappa) \propto (1 - \kappa)^d$ so that $\pi(d + 1 \mid \kappa)/\pi(d \mid \kappa) = 1 - \kappa$, 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 $\pi(d \mid \kappa) = C(\kappa)(1 - \kappa)^{d-1}$ for $d = 1, \ldots, N_0$ with $C(\kappa) = \kappa/(1 - (1 - \kappa)^{N0})$ and $\pi(\kappa) \sim \text{Unif}[0, 1]$. When κ is larger, d receives higher penalty for increasing, while when κ is fixed at a tiny value, say $\kappa = 0$, 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 $\pi(G_d \mid d) = (N - d)!/N!$ 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 $\mathcal{L}_2$ penalty through a normal prior, or $\mathcal{L}_1$ penalty through a logistic density prior on $\boldsymbol{\beta}$ which is equivalent to LASSO estimates (Park and Casella 2008). More specifically, for the fixed-effects function $\boldsymbol{\beta}$, we assume each element $\beta_{rk}(lm)$ of $\boldsymbol{\beta}$ is i.i.d.

$$\beta_{rk}(lm)|\varpi \sim \mathcal{N}(0, \sigma^2 \lambda_{rkl} \gamma_{rk}(lm)), \quad \gamma_{rk}(lm) \sim \text{Bernoulli}(p_{rkl}). \tag{12}$$

Note when $\gamma_{rk}(lm) = 0$, $\pi(\beta_{rk}(lm) \mid \varpi)$ 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 (12) introduces a set of hyper-parameters: 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 and George (2000). However, the empirical Bayes estimation procedures become intractable when the random clustering structure is introduced. Furthermore, the clustering results can be sensitive to the selection of hyper-parameters. Therefore, we update these hyper-parameters during the posterior sampling procedure.

Accordingly, we further consider the conjugate Beta prior $p_{rkl} \sim$ Beta $(a_{0,kl}, b_{0,kl})$, that is, $\pi(p_{kl}) \propto p_{rkl}^{a0,kl-1} (1 - p_{rkl})^{b0,kl-1}$, where the case $a_{0,kl} = b_{0,k=1}$ corresponds the non-informative 01031uniform prior on [0, 1], and one can assume Inverse Gamma prior for $\lambda_{rkl} \sim$ 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 ϵ, we consider the inverse-Gamma prior $\sigma^2 \sim$ igamma$(^a{}_\sigma, b_\sigma)$, where we choose $a_\sigma = b_\sigma = 0$ for non-informative Jeffreys prior to allow data-driven model search. The hyper-parameters $a_{0,kl}, b_{0,kl}, a_{1,kl}$ and $b_{1,kl}$ are also re-evaluated from the posterior samples via fitting SFC with fix-cluster runs by picking some reasonable partition (e.g., K-means). This gives reasonable prior information of the scaling and shrinkage levels of the wavelet coefficients. Our simulation studies suggest this data-driven model search works very well under various scenarios.

The Bayesian estimation of the spatial clusters ϖ involves variable model dimensions, that is, changing number of parameters associated with a new cluster introduced or an existing cluster deleted during a stochastic search. For example, a new cluster r can potentially introduce a new set of λ_{rkl}'s, p_{rkl}'s, and β_r. To handle the variable-dimensional model exploration, we adopt an embedded rjMCMC step for $\in$ within the conventional Gibbs sampler which then updates the remaining parameters. The full implementation of SFC is detailed in Appendix (Subsection 7.2). The key to the success of rjMCMC for spatial clustering to alleviate the known local-trap issue (Zhang et al. 2014) is to propose a new spatial cluster configuration without substantially deviating from the current $\in$ and avoid extreme values of likelihood ratio used for acceptance/denial. Our construction of different move types for stochastic search of spatial clustering configurations under rjMCMC is given in Appendix (Subsection 7.3).

4.2.3. SFC Posterior Inference

The prediction for the above spatial clustering model is based on averaging over all models α'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 **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 sameclustering membership in the b-th sample, and 0 otherwise. Let $w_{i1i2} = \sum_{b=1}^{B} w_b(i_1, i_2)/B \in [0, 1]$ and $w_{i_1 i_1} \equiv 0$ be the estimated adjacency between regions i_1 and i_2. We can then fit the STM with our estimate $\hat{W} = ((w_{i1i2}))_{1 \leq i_1, i_2 \leq N}$. To further obtain a central clustering configuration ϖ^* from the posterior samples $\varpi^1, \ldots, \varpi^B$, 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 $\hat{d}$ 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 $\hat{\beta}_r(t) = \sum_{b=1}^{B} \sum_{s \in Cr} \beta_s^{(b)}(t)/(n_r B)$ 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 $\hat{Y}_{ijt}$ based on $\beta_s^{(b)}(t)$'s for each age group j, and transform it back using Equation (2) to obtain the posterior estimates of the fertility rates $\hat{F}_{ijt}$ together with 95% confidence bands.

The performance of SFC in both curve clustering and estimation is assessed via simulation studies in Appendix (Subsection 7.4). It is compared with other popular clustering methods, such as K-means with the number of clusters selected using GAP statistic (Tibshirani et al. 2001) and a popular software SaTScan. The simulated data set mimics the real Portuguese fertility data (age-group 20–24), obtained by perturbing the cluster-mean (obtained by averaging the real response curves as cluster representative) for the specified true clustering configurations. As shown in Figure 3, the simulated mean curves for clusters 1–3 (from the real data) are similar to each other in means and trends, but there are local differences in spikes, fluctuations and slopes at specific time points.

Indeed, all three methods – SFC, Kmen + GAP, and SaTScan – are able to detect 4 clusters in more than half of the 500 such simulated data sets. One interesting finding is that if we use Kmeans + Silhouette criterion (predates GAP), it frequently picked 2 clusters only with poor simulation performance, without being able to distinguish between cluster 1, 2 and 3. The results show that competing methods, especially SFC, can well capture the true number of clusters under various scenarios, even when started from different initial values and hyper-parameter re-estimation (we used initial $d = 8$). We also tried a true $d = 1$ (no cluster), and SFC/K-means can pick the no-cluster pattern well.

The simulation results demonstrate the superiority of SFC serving both ends. We also compared SFC with the tool SaTScan for spatial clustering, and noticed several differences in the goals despite of consistent clustering results and comparable performance. One important difference is that SaTScan scans for spatial clusters in a certain time frame (subinterval of the study period) that exhibit extremely high or low values. Therefore, some regions may not receive any cluster label due to their absence at all detected hot-spots, and this happened in our simulated data. Whereas SFC or other functional-clustering techniques, such as K-means clustering, would consider the complete time course to cluster all regions. However, SFC does share similar features as scan-statistic-based approach, such as spatial contiguity that aids in the interpretation of clustering results. Our model-based SFC approach also provides parameter estimation (including Bayesian uncertainty quantification about the number of clusters) and allows covariates, while scan-statistic-based approach has less constraints on the cluster shapes such as voronoi tessellation, and is robust against violation of model assumptions. We
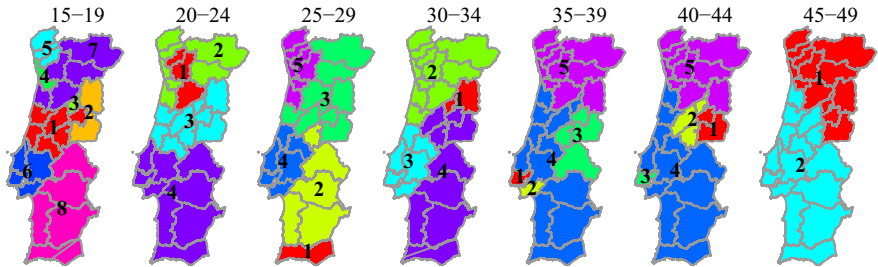

Fig. 3. *Estimated central clustering configurations by quinquennial age group j (j = 1,..., 7: ages 15 to 49 years old), with cluster labels.*

have added one graph to elaborate the difference and similarity between both approaches, in addition to the comparison with machine learning algorithms like K-means clustering in the simulation section. Finally, we also provide a code for easy implementation. See Appendix (Subsection 7.4) for further details.

4.3. Analysis of Portuguese Regional Fertility Rate Evolution

Using Portuguese data for $N = 28$ (NUTS III) regions, $T = 19$ years from 1991 to 2009, and $J = 7$ groups, we fit the spatio-temporal mixed-effects model (3) by proceeding with 3 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 SFC 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 two, 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 three years data are retained for evaluation of out of sample forecasts. Then, we conduct five MCMC runs with a total of 60, 000 iterations per chain for each age group $j = 1, 2,. . ., J = 7$. The five 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 a total of 60,000 iterations takes around nine minutes to complete on a regular 32MB RAM machine. The convergence of all five 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 five chains, and sample at every ten iterations to obtain a total of $B = 5,000$ samples for posterior inference.

4.3.1. 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

Table 1. *Posterior probabilities of the number of clusters (d) for SFC in each age group.*

d	15−19	20−24	25−29	30−34	35−39	40−44	45−49
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	

the $J = 7$ age groups in terms of both the posterior mode and spread of d. Figure 1 shows the central clustering configuration $\hat{\varpi}_j$ based on the estimated adjacency matrix $\hat{W}_j$ or 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 $\hat{F}_{ijt}$ 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 1. For each cluster in each age group in Figures 4 and 5, we also report the estimated Variation Rate (VR)

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

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 four to eight clusters
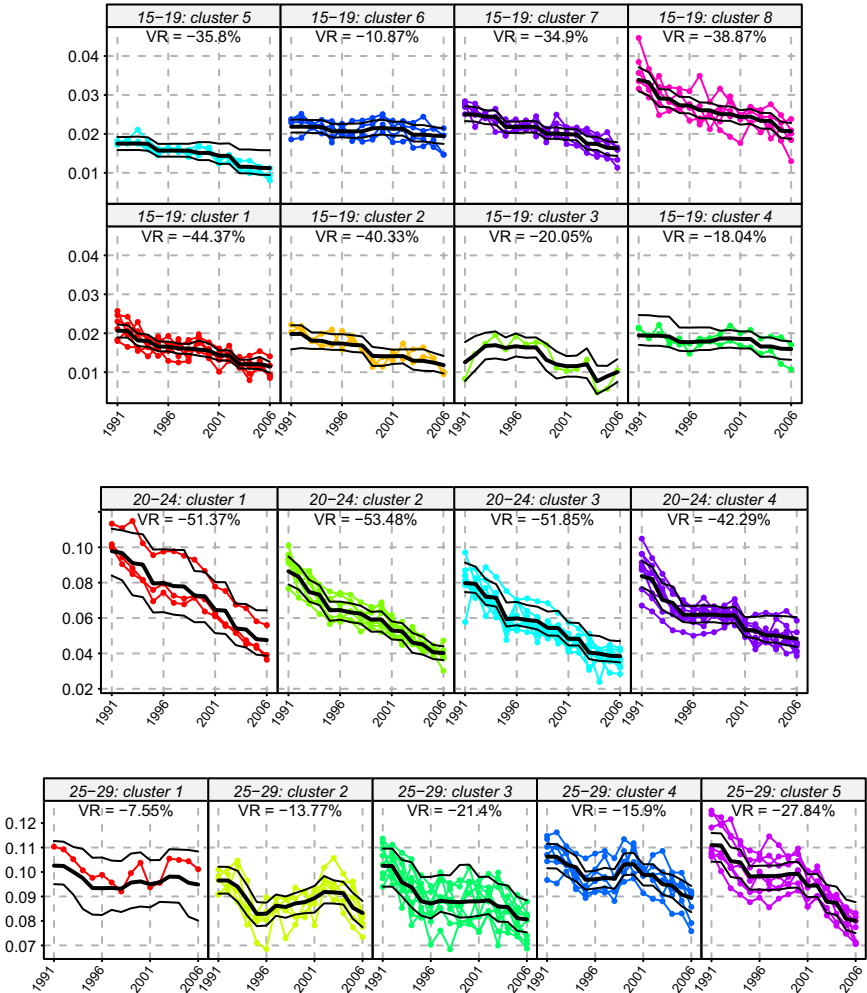

Fig. 4. *Functional clustering groups of fertility rates by quinquennial age group $j = 1, 2, 3$: 15 to 29 years old, together with variation rate (VR).*
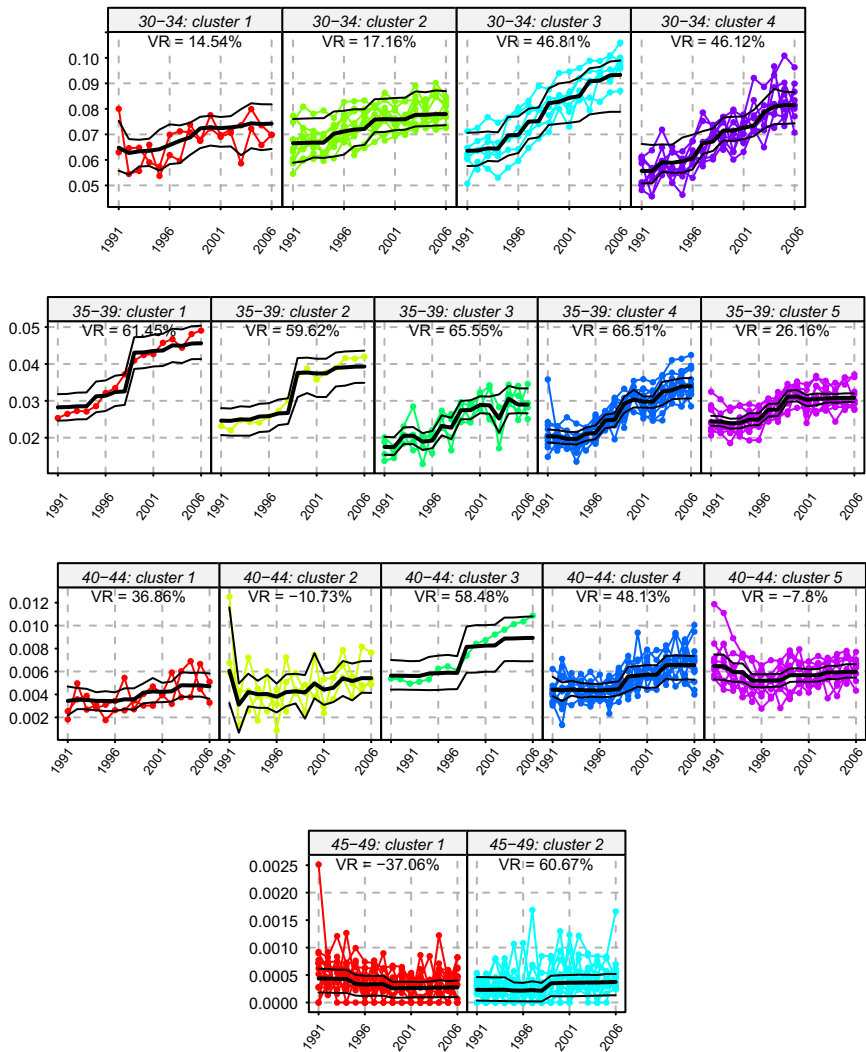
Fig. 5.   Functional clustering groups of fertility rates by quinquennial age group j = 4, 5, 6, 7: 30 to 49 years old, together with variation rate (VR).

in most age-groups, except the final age-group, $j = 7$: 45 to 49 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 two to eight. 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 (and the estimates, in Tables 2 and 3, of intercepts $\mu_1, \mu_2, . . ., \mu_7$ and slopes $\beta_1, \beta_2, . . ., \beta_7$ for the different age-groups), there are two features of temporal dynamics: (1) 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 (2) 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 ($j = 7$, 45–49 years), is an exception, presenting two clusters with opposite trends; however, this is connected to the low frequencies of births at these ages. Survey findings from Mendes et al. (2006) reveal four related factors that lead to the postponement of having children: prolongation of studies, subsequent entry into the labour market, the instability or lack of a marital relationship and the delayed leave from parental home. Indeed, Portugal now has one of the highest rates of women attending higher education in the world.

Beyond temporal dynamics, our work reveals interesting regional variations and spatial dynamics. Gomes et al. (2016) also discuss socio-economic conditions that explain the variation of regional fertility rates, highlighting the importance of female employment rates, income per capita and proportion of the unmarried couples. Even if the general territorial pattern of temporal variation in fertility rates is similar to the national level, the intensity with which these dynamics are expressed locally are different, creating clear regional variations. This leads us into spatial diffusion, where the clusters identified by our spatio-temporal modelling are very useful.

Several points may be highlighted. First, our approach of modelling spatial dynamics by clustering is clearly a substantial improvement over the common practice of assuming an *a priori* known contiguity based spatial weights matrix. Second, 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 Lisbon 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 the South around the capital are more persistent. 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, in which Lisbon (Grande Lisboa) (cluster 1) and Península de Setubal (cluster 2) form clusters distinct from the far south 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 other regions.

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 STM in Equation (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 $\hat{W}_j$ 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),

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

	No dynamics	Spatial-only	Temporal-only	Spatio-temporal	STM($\hat{W}_1$)
μ_1	9.92	9.94	10.04	10.01	10.00
	(9.76,10.09)	(9.67,10.21)	(9.82,10.25)	(9.50,10.57)	(9.41,10.53)
μ_2	18.51	18.54	18.64	18.61	18.60
	(18.35,18.68)	(18.27,18.80)	(18.43,18.85)	(18.11,19.17)	(18.02,19.14)
μ_3	20.64	20.66	20.76	20.73	20.72
	(20.47,20.80)	(20.38,20.93)	(20.55,20.97)	(20.21,21.29)	(20.14,21.24)
μ_4	15.47	15.49	15.59	15.56	15.56
	(15.31,15.63)	(15.22,15.77)	(15.38,15.79)	(15.05,16.11)	(14.96,16.09)
μ_5	8.89	8.90	9.01	8.98	8.97
	(8.72,9.05)	(8.63,9.18)	(8.80,9.23)	(8.47,9.53)	(8.37,9.50)
μ_6	4.20	4.21	4.32	4.29	4.28
	(4.04,4.36)	(3.94,4.49)	(4.12,4.53)	(3.77,4.83)	(3.69,4.84)
μ_7	1.25	1.27	1.37	1.34	1.33
	(1.08,1.42)	(1.00,1.54)	(1.17,1.59)	(0.83,1.90)	(0.73,1.86)
β_1	−0.14	−0.15	−0.16	−0.17	−0.15
	(−0.16,−0.13)	(−0.17,−0.13)	(−0.17,−0.14)	(−0.22,−0.12)	(−0.19,−0.11)
β_2	−0.35	−0.36	−0.37	−0.38	−0.36
	(−0.37,−0.34)	(−0.38,−0.34)	(−0.38,−0.35)	(−0.43,−0.33)	(−0.40,−0.32)
β_3	−0.17	−0.17	−0.18	−0.19	−0.18
	(−0.18,−0.15)	(−0.20,−0.15)	(−0.20,−0.16)	(−0.24,−0.15)	(−0.22,−0.13)
β_4	0.17	0.17	0.16	0.15	0.17
	(0.16,0.19)	(0.14,0.19)	(0.14,0.18)	(0.10,0.19)	(0.12,0.21)
β_5	0.18	0.17	0.17	0.15	0.17
	(0.16,0.19)	(0.15,0.20)	(0.15,0.18)	(0.10,0.20)	(0.13,0.22)
β_6	0.05	0.05	0.04	0.03	0.04
	(0.04,0.07)	(0.02,0.07)	(0.02,0.06)	(−0.02,0.07)	(0.00,0.09)
β_7	−0.00	−0.01	−0.01	−0.02	−0.01
	(−0.01,0.01)	(−0.03,0.02)	(−0.03,0.00)	(−0.07,0.02)	(−0.05,0.03)
δ^2	0.86	0.61	0.58	0.57	0.57
	(0.82,0.89)	(0.58,0.65)	(0.55,0.61)	(0.55,0.60)	(0.55,0.60)
τ^2	0.00	0.42	0.20	0.34	0.92
	(−0.08,0.08)	(0.33,0.51)	(0.15,0.26)	(0.24,0.44)	(0.66,1.18)
ϕ	0.00	0.00	0.92	0.93	0.94
	(−0.08,0.08)	(−0.08,0.08)	(0.89,0.94)	(0.91,0.95)	(0.91,0.95)
γ	0.00	0.94	0.00	0.97	0.95
	(−0.08,0.08)	(0.90,0.97)	(−0.08,0.08)	(0.95,0.99)	(0.91,0.98)
$\overline{D(\theta)}$	9990.89	9192.32	8279.49	7786.06	7942.31
PD_4	14.63	16.41	16.19	16.50	16.49
DIC_4	10005.53	9208.73	8295.69	7802.56	7958.80

spatial-only case, temporal-only case, full model with the natural adjacency matrix and estimated adjacency $\hat{W}_j$ matrices from each sub age-group $j = 1, 2,. . ., J = 7$ using the functional clustering techniques. The estimated $\hat{W}_2$ based on SFC for $j = 2$: 20–24 offers the best model fit, that is, the STM with the lowest DIC_4.

Finally, the clustering configurations have clear connections with regional migration patterns. Young people migrate to regional towns and cities for education, and then for employment, and this is evident in clusters that spread from major coastal cities – Lisboa, Porto, Aveiro and Faro – inward to interior areas in their hinterland, with some variations

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

	STM($\hat{W}_2$)	STM($\hat{W}_3$)	STM($\hat{W}_4$)	STM($\hat{W}_5$)	STM($\hat{W}_6$)	STM($\hat{W}_7$)
μ_1	10.20	10.16	9.98	10.00	10.08	10.13
	(9.70,10.78)	(9.66,10.62)	(9.45,10.55)	(9.41,10.53)	(9.53,10.60)	(9.61,10.70)
μ_2	18.80	18.76	18.58	18.60	18.68	18.73
	(18.30,19.37)	(18.25,19.23)	(18.05,19.14)	(18.02,19.14)	(18.13,19.19)	(18.20,19.28)
μ_1	20.92	20.88	20.70	20.72	20.80	20.85
	(20.41,21.50)	(20.38,21.35)	(20.17,21.26)	(20.14,21.24)	(20.27,21.32)	(20.32,21.42)
μ_1	15.76	15.71	15.54	15.56	15.63	15.69
	(15.26,16.34)	(15.21,16.19)	(14.99,16.10)	(14.96,16.09)	(15.10,16.15)	(15.16,16.24)
μ_1	9.17	9.13	8.95	8.97	9.04	9.10
	(8.66,9.75)	(8.63,9.59)	(8.40,9.51)	(8.37,9.50)	(8.49,9.56)	(8.57,9.65)
μ_1	4.48	4.44	4.26	4.28	4.36	4.41
	(3.99,5.05)	(3.93,4.90)	(3.71,4.82)	(3.69,4.84)	(3.83,4.88)	(3.89,4.96)
μ_1	1.53	1.49	1.31	1.33	1.41	1.46
	(1.02,2.10)	(1.00,1.95)	(0.78,1.88)	(0.73,1.86)	(0.87,1.93)	(0.95,2.02)
β_1	−0.16	−0.16	−0.17	−0.15	−0.16	−0.17
	(−0.20,−0.12)	(−0.21,−0.13)	(−0.22,−0.12)	(−0.19,−0.11)	(−0.21,−0.12)	(−0.21,−0.14)
β_2	−0.37	−0.38	−0.38	−0.36	−0.37	−0.38
	(−0.42,−0.33)	(−0.42,−0.34)	(−0.43,−0.33)	(−0.40,−0.32)	(−0.42,−0.33)	(−0.42,−0.35)
β_3	−0.19	−0.19	−0.19	−0.18	−0.18	−0.20
	(−0.23,−0.14)	(−0.23,−0.15)	(−0.25,−0.14)	(−0.22,−0.13)	(−0.23,−0.14)	(−0.23,−0.16)
β_4	0.15	0.15	0.15	0.17	0.16	0.14
	(0.11,0.20)	(0.11,0.19)	(0.10,0.20)	(0.12,0.21)	(0.11,0.20)	(0.11,0.18)
β_5	0.16	0.16	0.15	0.17	0.16	0.15
	(0.12,0.20)	(0.11,0.20)	(0.10,0.21)	(0.13,0.22)	(0.12,0.21)	(0.11,0.18)
β_6	0.03	0.03	0.03	0.04	0.04	0.02
	(−0.01,0.08)	(−0.01,0.07)	(−0.02,0.08)	(0.00,0.09)	(−0.01,0.08)	(−0.01,0.06)
β_7	−0.02	−0.02	−0.02	−0.01	−0.01	−0.03
	(−0.06,0.02)	(−0.07,0.02)	(−0.08,0.03)	(−0.05,0.03)	(−0.06,0.03)	(−0.06,0.00)
δ^2	0.57	0.57	0.57	0.57	0.57	0.57
	(0.54,0.60)	(0.55,0.60)	(0.54,0.60)	(0.55,0.60)	(0.54,0.60)	(0.55,0.60)
τ^2	0.52	0.73	0.85	0.92	0.83	1.66
	(0.38,0.67)	(0.53,0.94)	(0.61,1.10)	(0.66,1.18)	(0.61,1.06)	(1.20,2.12)
ϕ	0.93	0.93	0.94	0.94	0.93	0.94
	(0.91,0.95)	(0.91,0.95)	(0.92,0.96)	(0.91,0.95)	(0.91,0.95)	(0.91,0.95)
ψ	0.97	0.95	0.96	0.95	0.96	0.96
	(0.95,0.98)	(0.92,0.98)	(0.93,0.99)	(0.91,0.98)	(0.93,0.98)	(0.93,0.99)
$\overline{D(\theta)}$	7773.46	7925.02	7878.33	7942.31	7901.21	7923.85
PD_4	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

in alternate local business centers and university towns like Braga. In older age groups, some localized clusters in the interior of the country also emerge, likely because of more local migration. For example, in our central (best-fitting) clustering configuration in age-group $j = 2$ (30–34 years old), the 4 clusters represent Lisbon, Porto, Aveiro (and Coimbra) and Braga, which are also major university cities and towns. A similar pattern is observed for age-group $j = 3$ (35–39 years old), where the clusters emerge from Lisbon, Porto, Aveiro, Évora and Faro.

5. Forecast and Prediction

While our focus in this article 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 the STM as described in Equations (3)–(5) with two other related models.

5.1. Models

Our main model for the transformed fertility rates Y_{ijt} incorporating the spatial, temporal and group effects is the STM model. This we denote as **Model 1**.

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

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

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

$$\boldsymbol{\alpha}_1 \sim \mathcal{N}(0, \tau_1^2 \boldsymbol{D}(\psi)), \quad \boldsymbol{\alpha}_2 \sim \mathcal{N}(0, \tau_2^2 \boldsymbol{A}(\varphi))$$

5.2. 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 $t = 1, 2, \ldots, 16$, 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 $\boldsymbol{\theta} = \{\psi, \varphi, \boldsymbol{\mu}, \boldsymbol{\beta}, \tau^2, \delta^2\}$ 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

$$p(Y_0|Y) = \int p(Y_0|Y, \theta, \alpha_1) p(\theta, \alpha_1|Y) d\theta d\alpha_1$$

$$= \int \left(\int p(Y_0|Y, \theta, \alpha_1, \alpha_0) p(\alpha 0|Y, \theta, \alpha_1) d\alpha_0 \right) p(\theta, \alpha_1|Y) d\theta d\alpha_1$$

via composite sampling: with posterior samples $\{\theta^{(b)}, \alpha_1^{(b)}\}_{b=1}^{B}$, we draw $Y_0^{(b)}$ from the inner integration term which can be identified to a $\mathcal{N}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ since $\boldsymbol{\alpha}_0 \mid \boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\Sigma}_0 = \Sigma_{00} - \Sigma_{01} \sum_{11}^{-1} \Sigma_{10}$ and $\boldsymbol{\mu}_0 = \Sigma_{01} \sum_{11}^{-1} \boldsymbol{\alpha}$, and Σ_{ij} is the partition block of the covariance assumed for $(\alpha_0'|\alpha')'$ 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 $\boldsymbol{\mu}^* = \boldsymbol{\mu}_0 + \boldsymbol{\mu} + \boldsymbol{\beta} t$.

5.3. 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 $\overline{D(\theta)} = -2\mathbb{E}_1$ as the posterior expected value of the joint deviance, and $p_{D4} = \overline{D(\theta)} = -2\mathbb{E}_2$ 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, $k = 1, 2, \ldots, K$, which are, respectively, defined as

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

5.4. 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-specific mortality rate for each region is shown graphically in Appendix Subsection 7.5. In terms of in-sample predictive densities, as reflected for example in DIC_4, the best performance is obtained for the proposed STM. The additive model (**Model 3**) comes out the 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; 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

Table 4.   *Posterior inference for the three 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)
ψ	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)
$\overline{D(\theta)}$	6491.57	6907.71	7452.12
PD_4	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

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-parameterized 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 scenarios and policy measures.

6. Conclusion

This article develops a 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 assume 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. The MATLAB code for implementing STM and SFC is provided and maintained in the Github repository https://github.com/zhangz19/PorFerAnalysis.

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. 2016). In line with other recent evidences, for example in Bhattacharjee et 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 benefits brought about by diffusion. Reduced form models may predict better, although flexible structural models are more useful for policy.

7. Appendix

Table 5 gives some major notations used in this article.

Table 5. Notation.

Abbreviation	
STM	Spatio-temporal mixed-effects model
SFC	Spatio-functional clustering
AR	Autoregressive. STM assumes AR model for temporal dependence
CAR	Conditional autoregressive. STM assumes CAR model for spatial dependence
Indices	
N	Number of regions, with $s = 1, \ldots, N$ indicating ith region
J	Number of age groups, with $j = 1, \ldots, J$ indicating jth age group

d	Number of clusters, with $r = 1, \ldots, d$ indicating rth cluster C_r with n_r regions
p	Number of covariates, with $k = 1, \ldots, p$ indicating kth covariate
T	Number of time points, with $t = 1, \ldots, T$ indicating tthe time point (year)
L	Number of wavelet resolution levels, with $l = 1, \ldots, L$ indicating lth level
m	Index of position at each wavelet level l, $m = 1, \ldots, 2^{l-1}$
Parameters	
α	Spatio-temporal random effect
τ^2	variation explained by the spatio-temporal random effect
δ^2	unexplained variation under STM
W	$N \times N$ spatial adjacency matrix, usually defined with entry $w_{s1,s2} = 1$ if region s_1 and s_2 are neighbors, 0 otherwise.
M	$N \times N$ diagonal matrix with number of neighbors as diagonal entries
$A(\phi)$	Temporal correlation matrix under AR, with temporal dependence ϕ
$D(\psi)$	Spatial correlation matrix under CAR, with spatial dependence ψ
Q	$T \times T$ Discrete Wavelet Transform (DWT) matrix
$\in$	Spatial clustering configuration.
G_d	Set of d cluster centers. $\in$ labels are obtained by minimum distance criterion.
$\beta_{rk(lm)}$	Wavelet coefficient at position m of resolution l, for kth covariate in rth cluster
$\gamma_{rk(lm)}$	Bernoulli random variable for the spike-and-slab prior: 0 means $\beta_{rk(lm)}$ is a point mass at 0, 1 means $\beta_{rk(lm)}$ follows a Normal prior $\mathcal{N}(0, \sigma^2\lambda_{rkl})$.
σ^2	Noise level, or error variance in the wavelet regression setting
λ_{rkl}	Signal-to-noise ratio of wavelet coefficients. Λr is the diagonal matrix with λ_{rkl}'s as diagonal entries for cluster r
P_{rkl}	Wavelet shrinkage level, or probability of $\gamma_{rk(lm)} = 1$
κ	hyper-parameter for penalizing larger number of clusters d
$\pi(\theta \mid \vartheta)$	Conditional prior density of θ given ϑ
$\pi(\theta \mid \vartheta, Y)$	Full conditional posterior density of θ given ϑ and the data Y

7.1. STM Implementation

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

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

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

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

$$\pi(\delta^2|\boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \text{igamma}\ (a_\delta + NJT/2, b_\delta + \boldsymbol{\epsilon}'\boldsymbol{\epsilon}/2,\ \text{with}\ \boldsymbol{\epsilon} = \boldsymbol{\eta} - X\boldsymbol{\theta} - Z\boldsymbol{\alpha}$$

$$\pi(\tau^2|\alpha, \psi, \phi) = \text{igamma}\ (a_\tau + NJT/2, b_\tau + \boldsymbol{\alpha}'(A(\phi)^{-1}{\otimes}D(\psi)^{-1})\alpha/2)$$

$$\pi(\psi|\alpha, \tau^2, \phi) \propto |D(\psi)|^{-T/2}\exp\{\psi\alpha'(A(\phi)^{-1}{\otimes}W)\alpha/(2\tau^2)\}\cdot I(\phi\,\epsilon\,(e_N^{-1},\ e_1^{-1}))$$

$$\pi(\phi|\alpha, \tau^2, \psi) \propto |A(\phi)|^{-N/2}\exp\{-\alpha'(A(\phi)^{-1}{\otimes}D(\psi) - 1)\alpha/(2\tau^2)\}\cdot I(\phi\,\epsilon\,(-1, 1))$$

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

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

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

7.2. SFC Implementation

For notational convenience, we use $(a_{kl})_{k,l}$ to denote a vector $\boldsymbol{a}$ with elements indexed by double index $k = 1, 2,\dots, K$ and $l = 1, 2,\dots, L$, and stacked by the order of the subscripts, that is, $(a_{kl})_{k,l} = (a_{1,1},\dots, a_{K,1},\dots, a_{1,L},\dots, a_{K,L})$; we use a similar notation for more than two indexes. For example, we have $\boldsymbol{\beta} = (\beta_{rk}(lm))_{m,l,k,r}$. Let $\boldsymbol{\gamma} = (\gamma_1',\dots, \gamma_d')'$ with $\boldsymbol{\gamma}_r = (\gamma_{rk}(lm))_{m,l,k}$, $\boldsymbol{\lambda} = (\lambda_1,\dots, \lambda_d)'$ with $\boldsymbol{\lambda}_r = (\lambda_{rkl})_{l,k}$ and $p = (p_1,\dots, p_d)'$ with $\boldsymbol{p}_r = (p_{rkl})_{l,k}$. Then, the prior (12) can be informally written as $\boldsymbol{\beta}_r|\boldsymbol{\varpi} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\Lambda_r)$ where Λ_r is a diagonal matrix with $\boldsymbol{\lambda}_r\boldsymbol{\gamma}_r$ as diagonal entries (which allow zero). We also employ the generic notation $\boldsymbol{a}[c]$ to denote the c-th element (column) of the vector (matrix) $\boldsymbol{a}$, and $\boldsymbol{a}[-c]$ denotes the complement of $\boldsymbol{a}[c]$.

For Bayesian estimation of SFC parameters $\{\boldsymbol{\varpi}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{p}, \kappa\}$, each iteration of Gibbs sampling involves the following steps:

(1) Update $\boldsymbol{\varpi}$: We include an rjMCMC step for the variable-dimensional model search of the spatial clustering configuration $\boldsymbol{\varpi} = (d, G_d)$. Let $(\boldsymbol{\varpi}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{p}, \sigma^2, \boldsymbol{\beta})$ be the current state, we propose a new state $(\boldsymbol{\varpi}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*, \boldsymbol{p}^*, \sigma^{2*}, \boldsymbol{\beta}^*)$. Define the auxiliary variable $U = \boldsymbol{\theta}^* = (\sigma^{2*}, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*, \boldsymbol{p}^*)$ and let $\boldsymbol{\theta} = (\sigma^2, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{p}) = U^*$, the corresponding invertible map $\boldsymbol{q}:$ $(\boldsymbol{\theta}, U)$ to $(\boldsymbol{\theta}^*, U^*)$ is one-to-one with Jacobian determinant $|\mathcal{J}| = 1$. We first propose a new clustering configuration $\boldsymbol{\varpi}^*$. The proposed $\boldsymbol{\varpi}^*$ should be close to the

current ϖ to avoid dramatic change in likelihood that incurs the local trap problem in model search. The proposed moves that utilize the spatial information are detailed in Appendix Subsection 7.3.

Next, we propose the associated parameters $U \sim h(\cdot \mid \varpi, \theta, \varpi^*)$ via a certain proposal density h. Hence the acceptance ratio for the new state is

$$\min \left\{ 1, \frac{\pi(\varpi|\varpi^*)}{\pi(\varpi^*|\varpi)} \times \frac{\pi(\varpi^*, \theta^*|Y)}{\pi(\varpi, \theta|Y)} \times \frac{h(u^*|\varpi^*, \theta^*, \varpi)}{h(u|\varpi, \theta, \varpi} \times |\mathcal{J}| \right\} \tag{13}$$

We choose $h(u \mid \varpi, \theta, \varpi^*) = \pi(\gamma^*, \lambda^*, p^* \mid \varpi^*) \times \pi(\sigma^{2*} \mid \varpi^*, \gamma^*, Y) \times \pi(\beta^* \mid \varpi^*, \gamma^*, \sigma^{2*}, Y)$. More specifically, we first generate λ^* and p^* from the prior distributions, then γ^*, or $\gamma_{rk}(lm)$ from Bernoulli(p_{rkl}) for new clusters if any involved. Note for given ϖ^* and γ^*, the effective dimension of β^* is determined. For cluster r, let $\beta^*_{r,1}$ be the non-zero portion of each β^*_r with dimension q_r, the corresponding design matrix $X_{r,1}$, and the prior covariance $\sigma^2 \Lambda_{r,1}$ determined by γ^*_r. We first sample σ^{2*} from

$$\pi(\sigma^{2*}|\varpi^*, \gamma^*, \lambda^*, Y) \propto \pi(\sigma^{2*})\prod_{r=1}^{d} \int \pi(Y_r|\beta_r, \sigma^{2*})\pi(\beta_r|\gamma^*_r)d\beta_r \tag{14}$$

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

$$\pi(\varpi, \theta|Y) = \pi(Y|\varpi, \theta)\pi(\beta|\varpi, \gamma, \lambda, \sigma^2)\pi(\sigma^2)\pi(\gamma, \lambda, p|\varpi)\pi(\varpi) \times const.$$

and cancel out the prior and proposal part $\pi(\gamma, \lambda, p \mid \varpi,)$. Then, the Metropolis-Hasting ratio in Equation (13) becomes

$$\frac{\pi(\varpi|\varpi^*)}{\pi(\varpi^*|\varpi)} \times \frac{\pi(Y|\varpi^*, \beta^*, \sigma^{2*}, \gamma^*)\pi(\beta^*, \sigma^{2*}|\varpi^*, \gamma^*, \lambda^*)\pi(\beta, \sigma^2|\varpi, \gamma, \lambda, Y)}{\pi(Y|\varpi, \beta, \sigma^2, \gamma)\pi(\beta, \sigma^2|\varpi, \gamma, \lambda)\pi(\beta^*, \sigma^{2*}|\varpi^*, \gamma^*, \lambda^*, Y)} \times \frac{\pi(\varpi^*)}{\pi(\varpi)} \tag{15}$$

Using the fact that

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

the ration in Equation (15) reduces to

$$\frac{\pi(\varpi|\varpi^*)}{\pi(\varpi^*|\varpi)} \times \frac{\pi(Y|\varpi^*, \gamma^*, \lambda^*)}{\pi(Y|\varpi, \gamma, \lambda)} \times \frac{\pi(\varpi^*)}{\pi(\varpi)}. \tag{16}$$

When the prior and proposal density of α are both diffuse (such as under fixed tiny κ and uniform proposal), the acceptance rate is mainly determined by the marginal likelihood ratio, the second ratio in Equation (16). We have the marginal likelihood ratio $r = \pi(Y \mid \varpi^*, \gamma^*, \lambda^*)/\pi(Y \mid \varpi, \gamma, \lambda)$ with

$$\pi(Y|\boldsymbol{\varpi}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \int \left(\prod_{r=1}^{d} \int \pi(Y_r|\sigma^2, \boldsymbol{\beta}_r, \boldsymbol{\varpi}, \boldsymbol{\gamma}) \pi(\boldsymbol{\beta}_r|\sigma^2, \boldsymbol{\varpi}, \boldsymbol{\gamma}, \boldsymbol{\lambda}_r) \pi(\sigma^2) d\boldsymbol{\beta}_r \right) \sigma^2 \quad (17)$$

The integrand in Equation (17) is a product of three densities: $\mathcal{N}(X_{r,1}\boldsymbol{\beta}_r, \sigma^2 I_{n_rT})$'s, $\mathcal{N}(\mathbf{0}_{qr}, \sigma^2 \boldsymbol{\Lambda}_{r,1})$'s and `igamma`$(a_\sigma, b_\sigma)$. After the first integration with respect to $\boldsymbol{\beta}_r$'s, the first two Gaussian densities merge into $\mathcal{N}(\mathbf{0}_q, \sigma^2 V_r)$ with $V_r = I_{n_rT} - X_{r,1}(X'_{r,1}X_{r,1} + \boldsymbol{\Lambda}_{r,1}^{-1})^{-1}X'_{r,1}$. By Sherman-Morrison-Woodbury formula and Sylvester's determinant theorem

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

$$|A + UCV| = |C^{-1} + VA^{-1}U||C||A|$$

and taking $A = I_{nT}$, $U = -X_{r,1}$, $C = (X'_{r,1}X_{r,1} + \boldsymbol{\Lambda}^{-1})^{-1}$ and $V = X'_{r,1}$, we have $V^{-1} = I_{nrT} + X_{r,1}\boldsymbol{\Lambda}_{r,1}X'_{r,1}$ with determinant $|I_{nrT} + X_{r,1}\boldsymbol{\Lambda}_{r,1}X'_{r,1}| = |I_{qr} + \boldsymbol{\Lambda}_{r,1}X'_{r,1}, 1X_{r,1}|$. After the second integration with respect to σ^2, which is over `igamma`$(a_\sigma + NT/2, b_\sigma + \sum_{r=1}^{d} Y_r V_r Y_r/2)$, the marginal likelihood $\pi(Y|\boldsymbol{\varpi}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ in Equation (17) is simplified to:

$$\pi(Y|\boldsymbol{\varpi}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \frac{\Gamma(NT/2)\Delta^{-(\alpha_\sigma + NT/2)}}{\Gamma(\alpha_\sigma)(2\pi b_\sigma)^{NT/2}} \prod_{r=1}^{d} |\boldsymbol{\Lambda}_{r,1}X'_{r,1}X_{r,1} + I_{qr}|^{-\frac{1}{2}}, \quad (18)$$

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

Note that Equaiton (18) is a multivariate t-desity for ordered Y by clusters with location $\mathbf{0}_{NT}$, scale matrix $(b_\sigma/a_\sigma) \otimes_{r=1}^{d} (I_{n_rT} + X_{r,1}\boldsymbol{\Lambda}_{r,1}X'_{r,1})$ and degree of freedom $2a_\sigma$. For intercept-only model, since $X'_{r,1}X_{r,1} = n_r I_{qr}$ and, we have the simple form $\log \pi(Y | \boldsymbol{\varpi}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = const. - (a_\sigma + NT/2) \log(1 + \sum_{r=1}^{d} Y_r V_r Y_r/(2b_\sigma)) - \sum_{r=1}^{d}\sum_{c=1}^{T} \log(n_r\boldsymbol{\lambda}[c]\boldsymbol{\gamma}[c] + 1)/2$.

Since the marginal likelihood can be sensitive to the choice of a_σ and b_σ, one can choose the non-informative Jeffreys 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 Equation (18), but rather

$$\pi(Y|\boldsymbol{\varpi}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \frac{\Gamma(NT/2)\Delta_0^{-(NT/2)}}{\pi^{NT/2}} \prod_{r=1}^{d} |\boldsymbol{\Lambda}_{r,1}X'_{r,1}X_{r,1} + I_{qr}|^{-\frac{1}{2}} \quad (19)$$

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

(2) Update (β, γ): The independence of posterior distribution of β_r's and γ_r's follows the independence structure in both likelihood and prior distributions. Hence, we update (β_r, γ_r) independently for each $r = 1, \ldots, d$ through $\pi(\beta_r, \gamma_r | Y_r, \boldsymbol{\varpi}, \lambda_r, p_r) = \pi(\beta_r | \gamma_r, Y_r, \lambda_r, \boldsymbol{\varpi}) \times \pi(\gamma_r | Y_r, p, \boldsymbol{\varpi})$. We update each pair $(\beta_r[c], \gamma_r[c])$ given others $(\beta_r[-c], \gamma_r[-c])$, where $c = 1, 2,\ldots, pT$. We first sample $\gamma_r[c]$ from the posterior density that marginalizes $\beta_r[c]$ out:

$$\pi(\gamma_r[c]|Y_r, p_r) = \int \pi(\gamma_r[c], \beta_r[c]|Y_r, \beta_r[-c], \varpi, \sigma^2, \lambda_r, p_r) d\beta_r[c]$$

$$= \pi(\gamma_r[c]|p_r) \int \pi(Y_r|\beta_r[c], \beta_r[-c], \sigma^2) \pi(\beta_r[c]|\gamma_r[c], \lambda_r) d\beta_r[c] \times const.$$

Let M_s denotes the likelihood with $\beta_r[c]$ marginalized out conditioning on $\gamma_r[c] = s$ for $s = 0, 1$. Then M_1 is the density of $\mathcal{N}(X_r[-c]\beta_r[-c], \sigma^2(I_{n_r T} + \lambda_r[c]X_r[c]X_r[c]'))$. Similarly, M_0 is the density of $\mathcal{N}(X_r[-c]\beta_r[-c], \sigma^2 I_{n_r T})$. Assume the index c corresponds to the k-th covariate and l-the resolution level, then the posterior odds of $\gamma_r[c]$ being 1 is $O_\gamma = p_{rkl} M_1/[(1 - p_{rkl})M_0]$ and we sample $\gamma_r[c] \sim$ Bernoulli($O_\gamma/(O_\gamma + 1)$).

Next, we sample $\beta_r[c]$ from posterior distribution given the updated $\gamma_r[c]$. If $\gamma_r[c] = 0$, we have $\beta_r[c] = 0$; otherwise, we sample $\beta_r[c]$ through

$$\pi(\beta_r[c]|Y_r, \lambda_r, \beta_r[-c], \gamma_r[c] = 1) \propto \pi(Y_r|\beta_r[c], \beta_r[-c], \sigma^2) \pi(\beta_r[c]|\gamma_r[c] = 1, \lambda_r),$$

which is a Gaussian density with variance $\nu^2 = \sigma^2(\lambda_r[c]^{-1} + X_r[c]'X_r[c])^{-1}$ and mean $\mu = (\lambda_r[c]^{-1} + X_r[c]'X_r[c])^{-1}X_r[c]'(Y_r - X_r[-c]\beta_r[-c])$. Note that the Bayes factor M_1/M_0 for calculating posterior odds O_γ has the simple expression

$$\log(M_1/M_0) = -\frac{1}{2}\log(1 + \lambda_r[c]X_r[c]'X_r[c]) + +\frac{\mu^2}{2\nu^2}.$$

(3) Update ($\sigma^2, \lambda, p, \kappa$): During this step, we update (p, σ^2, κ) under current ($\varpi = (d, G_d), \beta, \gamma$). 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

$$p_{rkl}|\gamma \sim \beta \left(a_{0,kl} + \sum_{m=0}^{2^l-1} \gamma_{rk}(ml), b_{0,kl} + \sum_{m=0}^{2^l-1}(1 - \gamma_{rk}(lm))\right) \tag{20}$$

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

The conditional distribution of residual variance σ^2 is

$$\sigma^2|\beta, \lambda \sim \texttt{igamma}(a_\sigma^*, b_\sigma^*) \tag{21}$$

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

The conditional distribution of signal-to-noise ratio λ is

$$\lambda_{rkl}|\beta, \sigma^2 \sim \texttt{igamma}(a_{1,rkl}^*, b_{1,rkl}^*) \tag{22}$$

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

7.3.   Construction of Move Types for the rjMCMC 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. Split step $(d, G_d) \rightarrow (d + 1, G^*_{d+1})$: We create a new cluster. We first draw a random variable uniformly distributed on the $n - d$ non-center regions, to determine the new cluster C^* with center g^*. Secondly, we draw another random variable r uniformly distributed on $\{1, \ldots, d + 1\}$ to determine the position of g^* in G^*_{d+1}. The n^*_r regions that have minimal distance from g^* then automatically enter C^*. In this case, the ratios of proposal and prior probability are, respectively,

$$\frac{\pi(\varpi | \varpi^*)}{\pi(\varpi^* | \varpi)} = \frac{P(Merge)(N - d)}{P(Split)}, \frac{\pi(\varpi^*)}{\pi(\varpi)} = \frac{(1 - \kappa)^{d+1}(N - d - 1)!}{(1 - \kappa)^d (N - d)!}$$

   and the acceptance ratio in Equation (13) becomes $\min\{1, r(1 - \kappa)P(Merge)/P(Split)\}$. The ratio can be dominated by the marginal likelihood ratio $r = \pi(Y | \varpi^*, \gamma^*)/\pi(Y | \varpi, \gamma)$ particularly when one specifies a tiny κ for almost equal prior weights on the number of clusters d, and P(Merge) = P(Split) for equal chance of proceeding to Split and Merge step.

2. Merge step $(d + 1, G_{d+1}) \rightarrow (d, G^*_d)$: We delete one existing cluster and merge its members into other existing clusters. First, generate a random variable uniformly r distributed on $\{1, \ldots, d + 1\}$, which determines the cluster C_r with center g_r to be removed with all its members merging into one of the remaining clusters by the minimal distance criterion. The acceptance is the reciprocal of that in Split 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 $r \sim \mathrm{Uniform}\{1, \ldots, K(G_d)\}$ to obtain one such cluster center g_r with $R(g_r)$ non-center neighbors. Secondly, draw s from $\{1, \ldots, R(g_r)\}$ uniformly. The sth non-center neighbor becomes the new cluster center g^*_r that replaces g_r in G_d. The acceptance probability is $\min\{1, rK(G_d)R(g_r)/(K(G^*_d)R(g^*_r))\}$.

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 one 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 $\min\{1, r\}$.

Each iteration for updating α proceeds with one of the four move types based on prescribed probabilities P(Split) = P(Merge) = 0.4 and P(Shift) = P(Switch) = 0.1.

7.4.   Simulation Studies

We conduct simulation studies to evaluate the performance of the proposed SFC in both curve clustering and fertility estimation, and compare it with other popular methods such as K-means clustering. The number of clusters for K-means clustering can be determined via GAP statistic. The MATLAB code for the simulation to compare SFC with K-means clustering + GAP is included in the Github repository https://github.com/zhangz19/Por-FerAnalysis.

The simulated spatio-temporal data mimic the real Portugal fertility data analyzed in this work. We specify the true spatial clustering configuration with $d = 4$ clusters for the 28 Portugal NUTS III regions, with true cluster-specific mean transformed fertility calculated based on the real data. The true cluster labels are indicated by different colors in Figure 6 with spatial cluster centers. We then simulate 16-year temporal response for individual region by adding Gaussian noise with variance σ^2 to the corresponding cluster mean. We choose the noise level $\sigma^2 = 0.6$ based on the real data, but also investigate the performance of the clustering methods under different scenarios with smaller or larger noise level ($\sigma^2 = 0.2, 1$). A total of 500 spatio-temporal data sets were generated by perturbing the fixed true mean curve under the fixed true clustering configuration with noise level σ^2, and one example data set is provided in Figure 6 for each of the 3 scenarios.

For SFC, a total of 50,000 MCMC iterations include first 10,000 fixed-cluster runs. At end of the batch, the hypeparameters for the wavelet shrinkage and signal-to-noise ratio are re-evaluated by fitting the last 5,000 samples. Then the full SFC runs with the cluster search for the remaining 40,000 iterations, with the first 15,000 as burn-in period, and the posterior samples are drawn at every 10th iteration of the last 25,000 runs. In general, we recommend using some reasonable guess of the clustering configuration as the initial value, such as K-means clustering to obtain centroid curves and identify locations that have the closest profile as centroids for the initial spatial clustering. However, in the simulation study, we intentionally use a distinct number of clusters, e.g., $d = 8$ as the
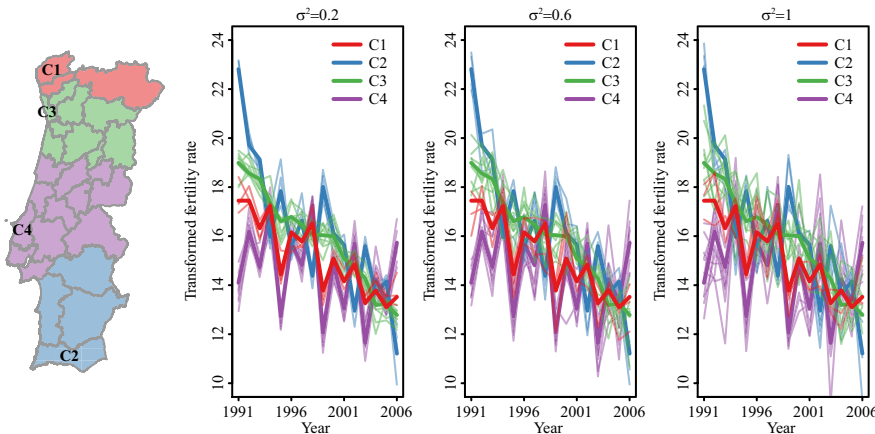

Fig. 6.   *Simulated true clusters with true mean curves (bold line) and example data set.*

initial value for SFC with hypeparameters re-evaluation, and check how well it reveals the true spatial clustering configuration with $d = 4$.

To evaluate the similarity between the clustering results under each method and the true labels, common measures for comparing two cluster labels (Hubert and Arabie 1985) are adopted, including Rand index (RI), adjusted Rand (AR) index to account for agreement by chance, Mirkin's index for disagreement, and Hubert's index to measure agreement. A value of Rand index closer to one indicates higher degree of agreement, while 0 indicates disagreement. We compute these measures between the central cluster under SFC and the true labels to evaluate the agreement. Similarly for K-means clustering and the true labels. For fertility estimation, we also provide root-mean-squared-error (RMSE) in mean curve estimation for each of the true four clusters. For SFC, the estimated mean curve is the posterior mean of $\boldsymbol{\beta}$-parameter averaged over all member regions in the true cluster. For K-means clustering, the centroid curves are used for member regions, which provide estimated mean curves for the true values by averaging over member regions in the true clusters.

The above simulation procedure is repeated 500 times to obtain the average of those performance metrics in Table 6. The results demonstrates the superiority of the proposed SFC over the frequently adopted K-mean clustering for the simulation studies in the Portugal fertility analysis context. Albeit starting from distinct initial d values (e.g., $d = 8$), SFC captures the true $d = 4$ well with an average posterior mode $d = 3.59$ over the 500 simulations, in contrast to averaged $d = 2.54$ by K-means clustering. SFC also has higher agreement with the true labels with lager agreement measures (AR, RI and HI) and lower disagreement measure (MI). For fertility estimation, the RMSE values for the four true clusters are generally much smaller under SFC when compared with K-means method, except the comparable estimation results for cluster 2 under $\sigma^2 = 0.6$ of primary interest. SFC is also robust against higher noise level in preserving agreement with the true labels as well as accuracy in mean curve estimation. This again demonstrates the merit of Bayesian model-based clustering approach comparing to non-parametric approach in this application. The above simulation procedure is also repeated with no clustering pattern under $\sigma^2 = 0.6$, both SFC and K-means + GAP were capable to capture the true $d = 1$ for all 500 simulated data sets.

In addition, we compared SFC with other spatial clustering methods such as scan-statistic-based approach implemented in popular software such as SaTScan (ver9.6). It has been adopted for detecting spatial and space-time clusters that locally exhibit high or

Table 6.   Averaged results for clustering and mean curve estimation over 500 simulations

σ^2	Method	d	AR	RI	MI	HI	$RMSE_1$	$RMSE_2$	$RMSE_3$	$RMSE_4$
1	SFC	3.98	0.75	0.90	0.10	0.79	0.78	0.55	0.42	0.30
	K-means	2.98	0.59	0.82	0.18	0.63	0.77	1.00	0.48	0.32
0.6	SFC	3.59	0.85	0.93	0.07	0.86	0.54	0.24	0.95	0.25
	K-means	2.54	0.81	0.90	0.10	0.81	1.11	0.22	1.68	0.36
0.2	SFC	4.39	0.85	0.94	0.06	0.88	0.41	0.24	0.21	0.14
	K-means	3.13	0.77	0.90	0.10	0.79	0.81	0.43	0.33	0.19

low values (e.g., risk), with statistical significance concluded using the scan statistic. In the spatial hot-spot detection context, the spatial clustering approach and scan-statistic-based approach have provided consistent results in previous studies (see, e.g., Waller 2015). However, the goal of scan-statistic-based approach is not necessarily for curve clustering with centroid curve estimation as SFC or K-means. Indeed, when applied to the simulated data above, SaTScan can detect $d = 4$ clusters with high agreement with the true labels, yet each concluded cluster has targeted time frame (subinterval in 1991-2009 period) with extreme values. Summary statistics (e.g., mean parameter) are presented inside or outside the window rather than for each time scale on a curve basis. Also, a few regions may not receive any label due to they are not at presence of any detected clusters (e.g., four regions in the example provided in Figure 6 for $\sigma^2 = 0.6$), which prohibits the further comparison using the aforementioned measures such as adjusted Rand index. Nevertheless, if the goal is to capture local clusters (regions in a specific time frame) with extreme values, the scan-statistic-based approach can be preferable due to fast implementation and less constraints on clusters. On the other hand, if the goal is to cluster regions based on their complete temporal profile, the non-parametric K-means method or the model-based SFC can be considered. In addition to spatial contiguity for clustering results that can aid in interpretation, SFC under Bayesian framework treats number of clusters d as random and can give further uncertainty quantification and insights on grouping structure. This is particularly beneficial when a single d can be misspecified, as elaborated in the simulation studies. The implementation for SFC is also publicly available, open source without call for any special library/toolbox, and carries simple input variables for easy setup.

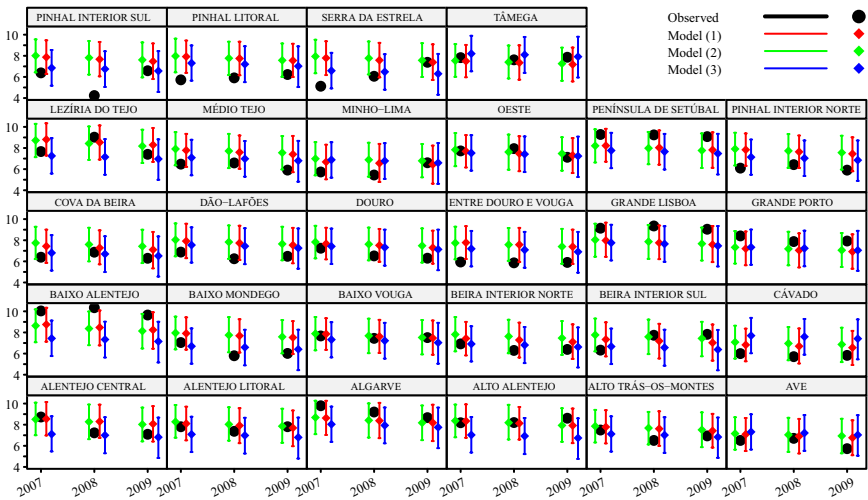### 7.5. *Forecast Performance of Models 1–3*

Fig. 7. *Estimated transformed fertility rates versus observed values for 15–19 age group.*
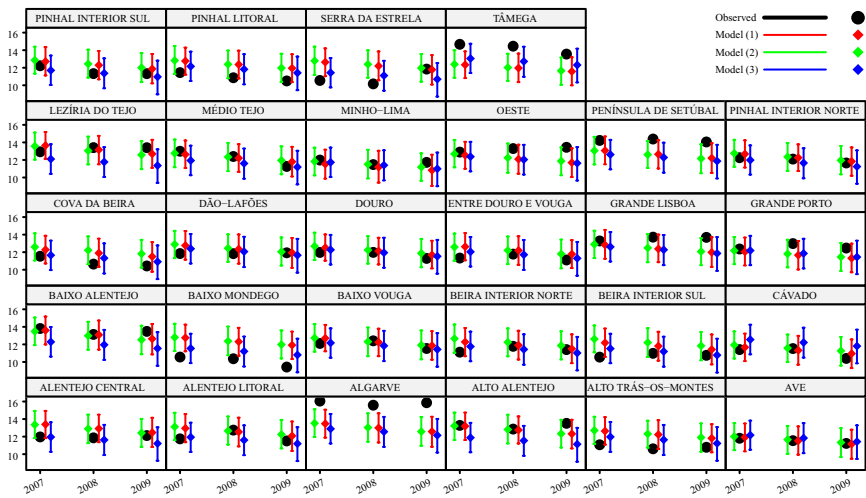
Fig. 8.   *Estimated transformed fertility rates versus observed values for 20−24 age group.*
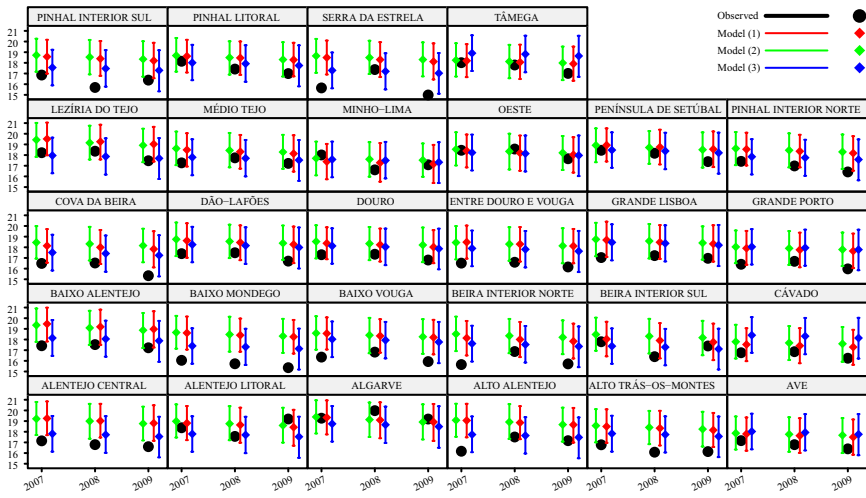

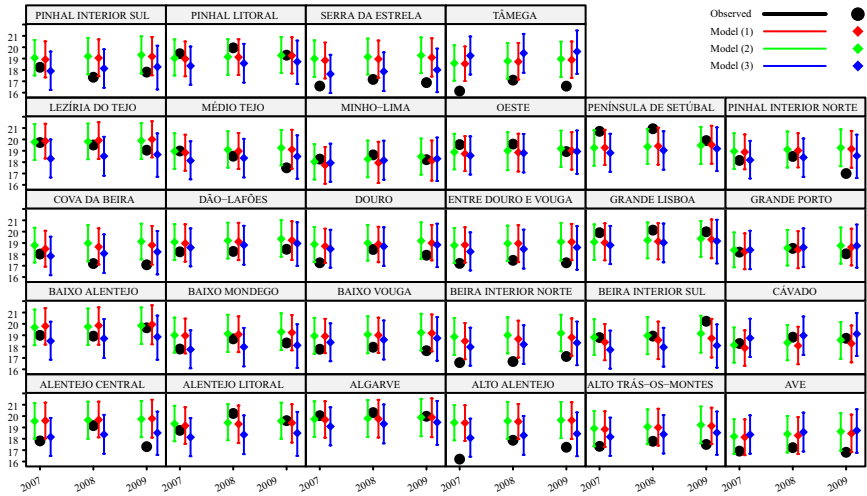Fig. 9.   *Estimated transformed fertility rates versus observed values for 25−29 age group.*

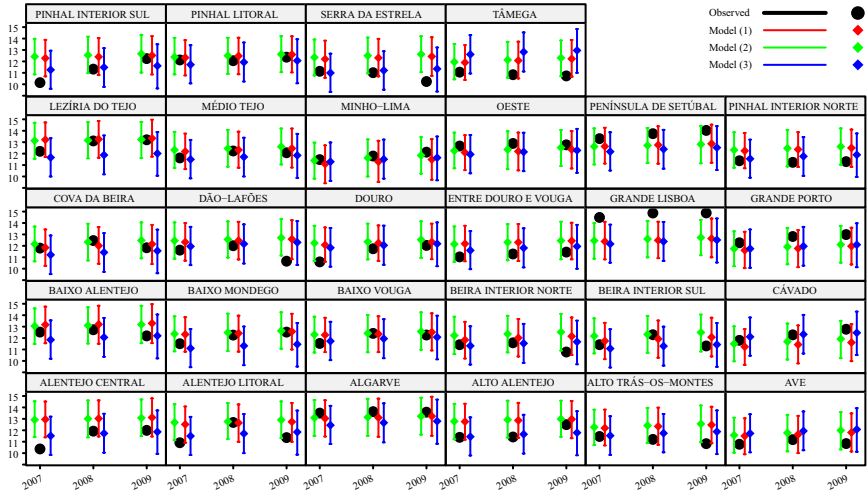Fig. 10.   *Estimated transformed fertility rates versus observed values for 30−34 age group.*

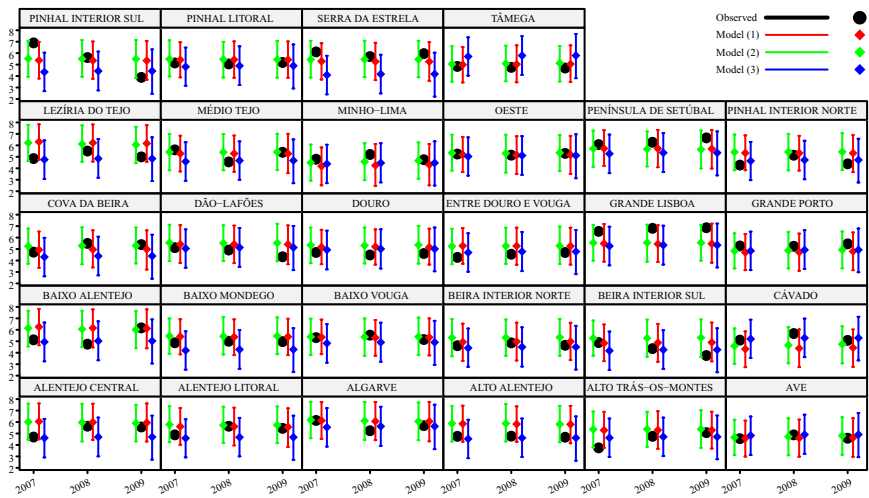Fig. 11.   *Estimated transformed fertility rates versus observed values for 35−39 age group.*

Fig. 12.   Estimated transformed fertility rates versus observed values for 40−44 age group.
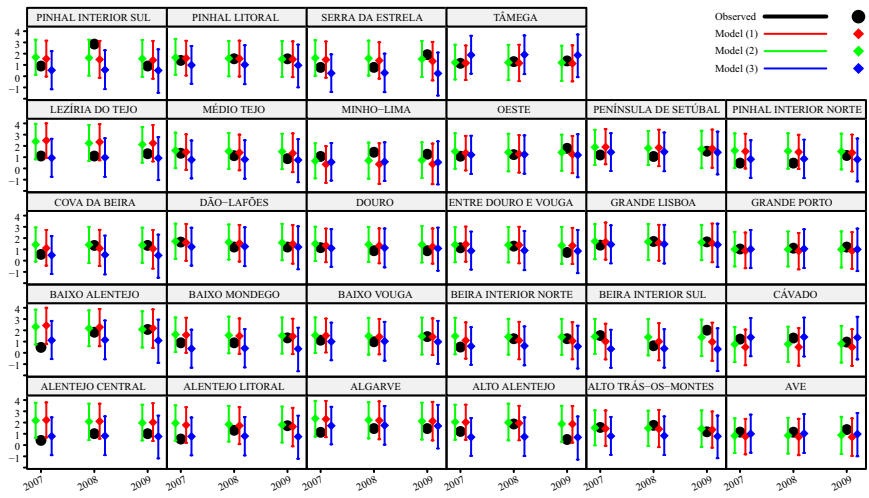

Fig. 13.   Estimated transformed fertility rates versus observed values for 45−49 age group.

8.   References

Alkema, L., A.E. Raftery, P. Gerland, S.J. Clark, and F. Pelletier. 2012. "Estimating the total fertility rate from multiple imperfect data sources and assessing its uncertainty". Demographic Research 26: 331–362. DOI: https://doi.org/10.4054/DemRes.2012.26.15.

Alkema, L., A.E. Raftery, P. Gerland, S.J. Clark, F. Pelletier, T. Buettner, and G.K. Heilig. 2011. "Probabilistic projections of the total fertility rate for all countries". *Demography* 48(3): 815–839. DOI: https://doi.org/10.1007/s13524-011-0040-5.

Assunção, R.M., C.P. Schmertmann, J.E. Potter, and S.M. Cavenaghi. 2005. "Empirical bayes estimation of demographic schedules for small areas". *Demography* 42(3): 537–558. DOI: https://doi.org/10.1353/dem.2005.0022.

Bailey, N., S. Holly, and M.H. Pesaran. 2016. "A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence". *Journal of Applied Econometrics* 31(1): 249–280. DOI: https://doi.org/10.1002/jae.2468.

Besag, J. 1974. "Spatial interaction and the statistical analysis of lattice systems". *Journal of the Royal Statistical Society: Series B (Methodological)* 36(2): 192–225. DOI: https://doi.org/10.1111/j.2517-6161.1974.tb00999.x.

Bhattacharjee, A., E. Castro, T. Maiti, and J. Marques. 2016. "Endogenous spatial regression and delineation of submarkets: A new framework with application to housing markets". *Journal of Applied Econometrics* 31(1): 32–57. DOI: https://doi.org/10.1002/jae.2478.

Bhattacharjee, A., and S. Holly. 2013. "Understanding interactions in social networks and committees". *Spatial Economic Analysis* 8(1): 23–53. DOI: https://doi.org/10.1080/17421772.2012.722669.

Bhattacharjee, A., and C. Jensen-Butler. 2013. "Estimation of the spatial weights matrix under structural constraints". *Regional Science and Urban Economics* 43(4): 617–634. DOI: https://doi.org/10.1016/j.regsciurbeco.2013.03.005.

Bhattacharjee, A., T. Maiti, and D. Petrie. 2014. "General equilibrium effects of spatial structure: Health outcomes and health behaviours in Scotland". *Regional Science and Urban Economics* 49: 286–297. DOI: https://doi.org/10.1016/j.regsciurbeco.2014.10.003.

Bigotte, J.F., A.P. Antunes, D. Krass, and O. Berman. 2014. "The relationship between population dynamics and urban hierarchy: Evidence from Portugal". *International Regional Science Review* 37(2): 149–171. DOI: https://doi.org/10.1177/0160017614524226.

Billari, F.C., R. Graziani, and E. Melilli. 2012. "Stochastic population forecasts based on conditional expert opinions". *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175(2): 491–511. DOI: https://doi.org/10.1111/j.1467-985X.2011.01015.x.

Billari, F.C., R. Graziani, and E. Melilli. 2014. "Stochastic population forecasting based on combinations of expert evaluations within the bayesian paradigm". *Demography* 51(5): 1933–1954. DOI: https://doi.org/10.1007/s13524-014-0318-5.

Blacker, C.P. 1947. "Stages in population growth". *The Eugenics Review* 39(3): 88.

Bongaarts, J., and R.A. Bulatao. 1999. "Completing the demographic transition". *Population and Development Review* 25(3): 515–529. DOI: https://doi.org/10.4054/-DemRes.2003.8.3.

Borgoni, R., and F.C. Billari. 2003. "Bayesian spatial analysis of demographic survey data: An application to contraceptive use at first sexual intercourse". *Demographic Research* 8: 61–92. DOI: https://doi.org/10.4054/DemRes.2003.8.3.

Brooks, S.P., and A. Gelman. 1998. "General methods for monitoring convergence of iterative simulations". *Journal of computational and graphical statistics* 7(4): 434–455. DOI: https://doi.org/10.1080/10618600.1998.10474787.

Carlin, B.P. and T.A. Louis. 2010. *Bayes and empirical Bayes methods for data analysis*. Chapman and Hall/CRC.

Castro, E.A., Z. Zhang, A. Bhattacharjee, J.M. Martins, and T. Maiti. 2015. "Regional fertility data analysis: a small area bayesian approach". In *Current trends in Bayesian methodology with applications*, 203–224.

Cavenaghi, S., J.E. Potter, C.P. Schmertmann, and R.M. Assunção. 2016. "Estimating total fertility rates for small areas in Brazil". *Anais*, 1–28.

Celeux, G., F. Forbes, C.P. Robert, D.M. Titterington et al. 2006. "Deviance information criteria for missing data models". *Bayesian analysis* 1(4): 651–673. DOI: https://doi.org/10.1214/06-BA122.

Ciarleglio, A., and R.T. Ogden. 2016. "Wavelet-based scalar-on-function finite mixture regression models". *Computational statistics and data analysis* 93: 86–96. DOI: https://doi.org/10.1016/j.csda.2014.11.017.

Clyde, M., and E.I. George. 2000. "Flexible empirical bayes estimation for wavelets". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62(4): 681–698. DOI: https://doi.org/10.1111/1467-9868.00257.

Clyde, M., and E.I. George. 2003. "Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis [with comment]". *Journal of the American Statistical Association* 98(463): 584–585. DOI: https://doi.org/10.1198/016214503000000422.

Coale, A.J. 2017. *The decline of fertility in Europe*. Princeton University Press.

Cressie, N., and N.H. Chan. 1989. "Spatial modeling of regional variables". *Journal of the American Statistical Association* 84(406): 393–401. DOI: https://doi.org/10.2307/2289922.

Divino, F., V. Egidi, and M.A. Salvatore. 2009. "Geographical mortality patterns in italy: A bayesian analysis". *Demographic Research* 20: 435–466. DOI: https://doi.org/10.4054/DemRes.2009.20.18.

Entwisle, B. 2007. "Putting people into place". *Demography* 44(4): 687–703. DOI: https://doi.org/10.1353/dem.2007.0045.

Feng, W., C.Y. Lim, T. Maiti, and Z. Zhang. 2016. "Spatial regression and estimation of disease risks: A clustering-based approach". *Statistical Analysis and Data Mining: The ASA Data Science Journal* 9(6): 417–434. DOI: https://doi.org/10.1002/sam.11314.

Ferguson, T.S. 1973. "A bayesian analysis of some nonparametric problems". *The annals of statistics* 1 (2): 209–230. DOI: https://doi.org/10.1214/aos/1176342360.

Festy, P., F. Prioux, N. Unies. 2020. An evaluation of the Fertility and Family Surveys project. UN.

Gomes, M.C.S., C.J. Silva, E.A.d. Castro, and J.L. Marques. 2016. "Evolucãao da fecundidade em portugal: uma perspetiva sobre a diversidade regional". *Análise Social* (218): 36–70. DOI: https://doi.org/https://www.jstor.org/stable/43755168.

Green, P.J. 1995. "Reversible jump markov chain monte carlo computation and bayesian model determination". *Biometrika* 82(4): 711–732. DOI: https://doi.org/10.1093/biomet/82.4.711.

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

Hubert, L., and P. Arabie. 1985. "Comparing partitions". *Journal of classification* 2(1): 193–218. DOI: https://doi.org/10.1007/BF01908075.

INE (2011)". *Nados-vivos por local de residência da mãae, grupo etáario da mãae, sexo e filiacãao*. Recenseamento Geral da Populacãao (Censos de 1991 a 2011), Lisboa. Available at: https://www.pordata.pt/Municipios/Nados+vivos+de+m%C3%A3es+residentes+em+Portugal+total+e+por+grupo+et%C3%A1rio+da+m%C3%A3e-104.

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

Lam, D.A., and J.A. Miron. 1991. "Seasonality of births in human populations". *Social biology* 38(1–2): 51–78. DOI: https://doi.org/10.1111/j.0006-341x.2000.00013.x.

Mendes, M.F., C. Rego, and A. Caleiro. 2006. "Educacãao e fecundidade em Portugal: As diferenças nos níveis de educacãao influenciam as taxas de fecundidade?" DOI: https://doi.org/10.13140/RG.2.1.2449.1764.

Morris, J.S., C. Arroyo, B.A. Coull, L.M. Ryan, R. Herrick, and S.L. Gortmaker. 2006. "Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study". *Journal of the American Statistical Association* 101(476): 1352–1364. DOI: https://doi.org/10.1198/016214506000000465.

Morris, J.S., and R.J. Carroll. 2006. "Wavelet-based functional mixed models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(2): 179–199. DOI: https://doi.org/10.1111/j.1467-9868.2006.00539.x.

Notestein, F.W. 1945. "Population-the long view". In *Food for the World*, edited by Theodor Schultz, 36–57. Chicago: University of Chicago Press.

Park, T., and G. Casella. 2008. "The bayesian lasso". *Journal of the American Statistical Association* 103(482): 681–686. DOI: https://doi.org/10.1198/016214508000000337.

Potter, J.E., C.P. Schmertmann, R.M. Assuncãao, and S.M. Cavenaghi. 2010. "Mapping the timing, pace, and scale of the fertility transition in Brazil". *Population and development review* 36(2): 283–307.

Potter, J.E., C.P. Schmertmann, and S.M. Cavenaghi. 2002. "Fertility and development: evidence from Brazil". *Demography* 39(4): 739–761. DOI: https://doi.org/10.1111/j.1728-4457.2010.00330.x.

Raftery, A.E. 1995. "Bayesian model selection in social research". *Sociological methodology* 25 (1995): 111–163. DOI: https://doi.org/10.2307/271063.

Raftery, A.E., D. Madigan, and J.A. Hoeting. 1997. "Bayesian model averaging for linear regression models". *Journal of the American Statistical Association* 92(437): 179–191. DOI: https://doi.org/10.1080/01621459.1997.10473615.

Ray, S., and B. Mallick. 2006. "Functional clustering by bayesian wavelet methods". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(2): 305–332. DOI: https://doi.org/10.1111/j.1467-9868.2006.00545.x.

Retherford, R., N. Ogawa, R. Matsukura, and H. Eini-Zinab. 2010. "Multivariate analysis of parity progression–based measures of the total fertility rate and its components". *Demography* 47(1): 97–124. DOI: https://doi.org/10.1353/dem.0.0087.

Ritter, C., and M.A. Tanner. 1992. "Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy-Gibbs sampler". *Journal of the American Statistical Association* 87(419): 861–868. DOI: https://doi.org/10.2307/2290225.

Sampson, R.J., J.D. Morenoff, and T. Gannon-Rowley. 2002. "Assessing "neighborhood effects": Social processes and new directions in research". *Annual review of sociology* 28(1): 443–478. DOI: https://doi.org/10.1146/annurev.soc.28.110601.141114.

Seiver, D.A. 1985. "Trend and variation in the seasonality of us fertility, 1947–1976". *Demography* 22(1): 89–100. DOI: https://doi.org/10.2307/2060988.

Silva, C., C. Gomes, M. Pinto, J. Marques and E. Castro. 2011. "Iguais mas diferentes: a importância em regionalizar os modelos de projecção da população portuguesa." XVII Congresso Nacional da APDR. 29 de Junho a 01 de Julho 2011; Bragança e Zamora-Portugal e Espanha): 275–289.

Smith, S.K., and T. Sincich. 1988. "Stability over time in the distribution of population forecast errors". *Demography* 25(3): 461–474. DOI: https://doi.org/10.2307/2061544.

Thompson, W.C. 1929. "Population". *American Journal of Sociology* 34: 959–975. DOI: https://doi.org/10.1086/214874.

Tibshirani, R., G. Walther, and T. Hastie. 2001. "Estimating the number of clusters in a data set via the gap statistic". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2): 411–423. DOI: https://doi.org/10.1111/1467-9868.00293.

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. DOI: https://doi.org/10.2307/2096388.

Voss, P.R. 2007. "Demography as a spatial social science". *Population research and policy review* 26(5–6): 457–476. DOI: https://doi.org/10.1007/s11113-007-9047-4.

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 (2): 311–324. DOI: https://doi.org/10.1016/S0378-3758(03)00111-3.

Waller, L.A. 2015. "Discussion: statistical cluster detection, epidemiologic interpretation, and public health policy". *Statistics and Public Policy* 2(1): 1–8. DOI: https://doi.org/10.1080/2330443X.2015.1026621.

Weeks, J.R. 2004. "The role of spatial analysis in demographic research". In *Spatially integrated social science*, edited by M.F. Goodchild, and D.G. Janelle, 381–399.

Weeks, J.R., M.S. Gadalla, T. Rashed, J. Stanforth, and A.G. Hill. 2000. "Spatial variability in fertility in menoufia, Egypt, assessed through the application of remote-sensing and gis technologies". *Environment and Planning A* 32(4): 695–714. DOI: https://doi.org/10.1068/a3286.

Whittle, P. 1954. "On stationary processes in the plane". *Biometrika* 41 (3/4): 434–449. DOI: https://doi.org/10.2307/2332724.

Zhang, Z., C.Y. Lim, and T. Maiti. 2014. "Analyzing 2000–2010 childhood age-adjusted cancer rates in Florida: A spatial clustering approach". *Statistics and Public Policy* 1(1): 120–128. DOI: https://doi.org/10.1080/2330443X.2014.979962.

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. DOI: https://doi.org/10.1007/s11424-012-1039-8.