Spatio-temporal models for georeferenced unemployment data

Soraia Pereira
FCUL, Universidade de Lisboa, Lisboa, Portugal.
E-mail: soraia.gpereira@gmail.com
K F Turkman
FCUL, Universidade de Lisboa, Lisboa, Portugal.
Luís Correia
Instituto Nacional de Estatística, Lisboa, Portugal.
Håvard Rue
King Abdullah University of Science and Technology, Saudi Arabia

Abstract. The Portuguese National Statistical Institute is responsible for estimating and publishing quarterly labour market figures at national level for both NUTS I and NUTS II regions. Over recent years it has become increasingly important to identify these figures at more disaggregated levels. However, based on the established direct estimation method, it is not possible to produce satisfactorily precise estimates at higher spatial resolutions. From the 4th quarter of 2014 onwards, all the sampling units, namely the residential buildings, of the Portuguese Labour Force Survey (PLFS) were georeferenced. To take full advantage of this information, Pereira *et al.* (2019) proposed applying a spatial marked point process approach to unemployment estimation, in which the estimation of the unemployment intensity becomes the focal point. There, the sampling units were assumed to be a realization of a spatial point process, specifically a log Gaussian Cox process, with the number of unemployed in each unit being their respective marks. Recently, further information on the geo-referenced locations of all units of the population, namely all residential buildings in the national territory, became available. Consequently, it is no longer necessary to model the spatial configuration of the units of the population. Thus, we propose a new point referenced model for the marks based on the sampled units and extrapolate this to all units of the population. As expected, this extra information, and as a consequence the new model itself, produce estimates with higher precision.

Keywords: point-referenced data models ; geostatistics; Bayesian inference; spatio-temporal analysis; unemployment estimation; small area estimation; INLA; SPDE

1. Introduction

In Portugal, the National Statistical Institute (NSI) is responsible for performing, on a quarterly basis, the Labour Force Surveys (LFS) covering the entire national territory and for supplying the national and European entities with the conclusions taken from these sample surveys. Consequently, the NSI publishes official quarterly labour market statistics, including the estimated unemployment figures at different spatial resolutions, typically for NUTS I and NUTS II regions. NUTS is the classification of territorial units for statistics, created by the Eurostat and the National Statistical Institutes of the European Union, and includes three hierarchical levels: NUTS I, NUTS II and NUTS III (see Figure 1).

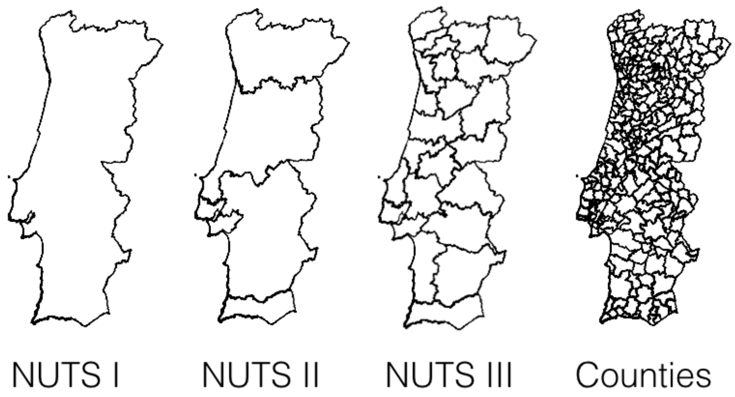


Figure 1. NUTS (version 2013) and counties in mainland Portugal

Together with the increase in demand for ever more detailed information at higher spatial resolutions, the demand for more reliable estimates without increasing the inherent costs of larger samples also increases. Typically, the NSI produces unemployment estimates derived from direct estimation methods based on the Horvitz-Thompson estimator (Horvitz and Thompson, 1952). However, these direct estimation methods do not perform well in small areas, increasing the demand either for larger samples or for small area estimation methods (Rao, 2003) that borrow strength from neighboring observations.

There have been considerable methodological developments to solve small area estimation problems in an unemployment context. The majority of small area methods are based on generalized linear models applied to areal data by modelling an appropriate counting process (Pereira *et al*, 2018, 2019).

From 2014 onwards, all of the sampling units in the LFS were georeferenced, namely the dwellings in which the observation units (i.e. individuals) are interviewed. This new data structure permits using point referenced models (Banerjee *et al.*, 2004). Pereira *et al.* (2019), based on the assumption that spatial distribution of the dwellings in the population is not known, proposed a marked point process approach, where the households chosen as the sampling units together with the number of unemployed individuals observed in each of these units are a realization of a spatial marked point process. This marked point process is then modelled by a log gaussian Cox process together with Poisson marks. Study of such marked spatial point process models focuses on the spatial intensity functions of the points and the marks. Once these quantities are estimated, the number of unemployed can be estimated by integrating the product intensity function over any desired region. Comparative studies show that point process methods not only produce estimates with much better precision compared to direct estimation methods but also produce reliable estimators over smaller regions, such as counties, in which it is not possible to obtain estimates using direct estimation methods due to reduced samples. Comparison of this point-level model with the standard small area estimation (SAE) areal models are given in (Pereira *et al.*, 2019).

Recently, the NSI provided new information on the geo-referenced locations of all the units of the population, namely the geo-referenced positions of all residential buildings in the national territory. With this new detailed, geo-referenced information, the spatial distribution of all dwellings is known and fixed, and therefore new spatial models, without the need to model randomness of points should, in principle, produce more precise estimates with reduced sampling variation. Hence, the objective now becomes to model the spatial variation of the number of unemployed people in each of the sampled dwellings. Thus, adjusting an appropriate point referenced model for the unemployed individuals in each sampling unit, and then extrapolating in space to all geo-referenced dwellings using spatial smoothing, should in principle produce estimates with higher precision than the point process approach, as the uncertainty regarding the spatial configuration of residential units is no longer present. The number of unemployed people in any areal unit A can then be calculated as the sum of the unemployed in all of the dwellings in that areal unit. Thus, in the presence of this new information, the modelling strategy we propose, based on 14,000 dwellings sampled in each quarterly sample survey, is based upon fitting a Poisson generalized linear model with a latent spatio-temporal structured random effect for the number of unemployed people observed in residential units, and, by spatial smoothing, extending these unemployment figures to all dwellings in the population whose geo-referenced positions are now known.

The proposed analysis is based on 9 sequentially observed quarterly sampling surveys (from the 4th quarter of 2014 to the 4th quarter of 2016).

The structure of the manuscript is as follows: In sections 2 and 3, the sampling scheme of the LFS and the data are described. We formulate the statistical model and discuss its implementation within the integrated nested Laplace approximation (INLA) platform in sections 4 and 5. In section 6, we give the results of inference, and we make a comparative study to assess the importance of the different georeferencing methods employed in the sampling surveys. We also make a sensitivity analysis of the covariates effects in section 6. Finally, the results and conclusions are discussed in section 7.

2. Sampling design of the Labour Force Survey

The portuguese LFS is a continuous survey, with the indicators published quarterly, and it is directed to the individuals living in dwellings of main residence in the national territory. The survey provides an understanding of the socioeconomic situation of these individuals during the week prior to the interview (reference week). The sampling units are the private dwellings and the observation units are the inhabitants living in these dwellings.

From the 4th quarter of 2014 onwards the sampling frame is selected from the National Dwellings Register (NDR). The NDR includes the geo-referenced positions of all residential buildings across the national territory, based on the 2011 Census data. The georeferencing process is made using a WEB application of Geographic Information

Systems.

The LFS follows a stratified multi-stage sampling design. First, the sampling frame (National Dwellings Register, built from the 2011 census) was stratified into 25 regions (NUTS III or groups of NUTS III). Then, in each strata, a multi-stage sampling was conducted, where the primary sampling units are areas consisting of one or more contiguous cells of the km^2 INSPIRE grid, and the secondary units are private dwellings. All the inhabitants living in the selected dwellings are surveyed.

From one quarter to another, the sample changes through a rotation system comprising six waves. Dwellings are kept in the sample for six consecutive quarters before being replaced by an identical number of dwellings in the same statistical section. One sixth of the sample is replaced each quarter. Consequently, each individual in the sample is surveyed over 6 consecutive quarters, inducing strong temporal dependence between the quarterly surveys.

The official estimates of the unemployment figures are calculated using a direct method, based on the Horvitz-Thompson estimator.

3. Data

We use the Portuguese LFS data from the 4th quarter of 2014 to the 4th quarter of 2016 in the mainland territory. We did not include the autonomous regions because it would increase complexity in the spatial modelling process and, moreover, they coincide with the NUTS II regions for which official estimates are already available with acceptable accuracy.

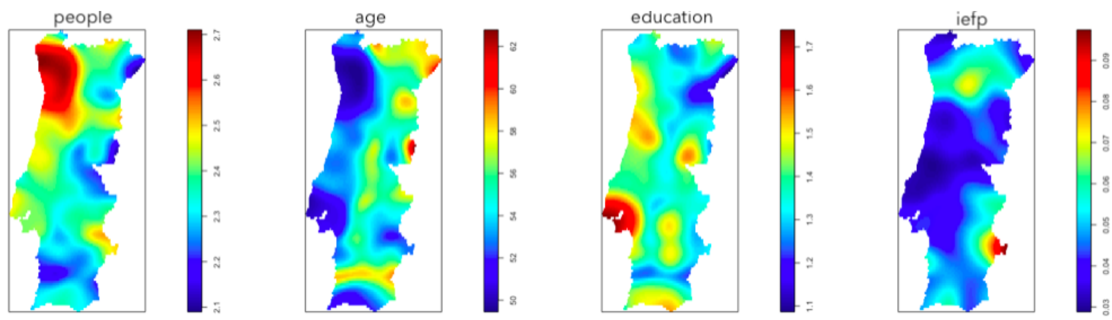
In each quarter, the sample has around 35000 observations, distributed in about 14000 dwellings, located in about 13800 residential buildings. Thus, in the majority of the sampled residential buildings, only one dwelling is selected. Each individual in the sample is questioned about their state in the labour market (employed, unemployed, inactive), gender, age, education level (primary level, secondary level, higher level), etc.

The georeferencing of all residential buildings of the population as well as of the sample are now available. Although a residential building can have multiple dwellings, the coordinates are available only for the buildings themselves. Since there may be more than one dwelling in each residential unit, particularly in areas of high population density, multiple dwellings in the survey have the same spatial location. To avoid an overlap in the locations within the modelling process, the observation units we consider are the residential buildings. In the following sections, we will denote the average number of unemployed people per dwelling in the residential building at location s_j and quarter t by $y(s_j, t)$ (rounding to the nearest integer). Here, we intend to extrapolate the values observed in the sampled locations to all residential buildings (around 2300000) by spatial smoothing based on the proposed model. The exact number of dwellings per residential building is known. Note that 5/6 of the sampled dwellings are the same from one quarter to another due the rotative sampling design explained in the previous section. In general, each individual is surveyed in 6 consecutive quarters, which causes an highly induced temporal correlation in our data.

In the modelling process, we use some covariates at residential building level for each quarter, namely the mean age and the median of the education level. Although the education level does not constitute a quantitative variable, it was treated as such due to its ordinal meaning (1-primary level, 2-secondary level, 3-higher level). The average number of people per dwelling in each residential building was considered as an offset (in log-scale). We also use information about the proportion of unemployed people registered in the employment centers (IEFP), available by counties. A spatial extrapolation of the covariates for the whole domain and study period is required (Figure 2 shows the covariates for the 4th quarter of 2016). For that, we used a simple non-parametric kernel spatial smoothing method (Nadaraya, 1964, 1989; Watson, 1964), giving us a reasonably accurate auxiliary information for out-of-sample residential units. The idea behind this method, also known as the Nadaraya-Watson smoother is the following: if the observed values are $y(s_1), ..., y(s_n)$ at locations $s_1, ..., s_n$ respectively, then the smoothed value at a location u is given by

$$g(u) = \frac{\sum k(u - s_i) y(s_i)}{\sum k(u - s_i)} \tag{1}$$

where k is a probability density. Here, we considered an isotropic Gaussian probability density which is known as Gaussian kernel. A preliminary analysis of the covariates tells us that the north of the country is the region with the most people living in the same residential building, the coast is the area with the youngest population, Lisbon is the region with the highest education level, and the north and Alentejo are the regions with the highest proportion of registered unemployed people in the employment centers.



a) b) c) d)

Figure 2. Kernel estimates regarding to the 4th quarter of 2016 for: a) average number of people per dwelling in each residential building; b) mean age per residential building; c) median of the education level per residential building; d) proportion of registered unemployed people in centers of employment

4. Point referenced data models for unemployment estimation

We will assume a Poisson distribution for $y(s, t)$, the average number of unemployed people per dwelling observed at residential building with spatial location at s and in quarterly survey t:

$$y(s,t)|\lambda(s,t) \sim \text{Poisson}(\lambda(s,t)) \tag{2}$$

with

$$\log(\lambda(s,t)) = \alpha + \text{offset}(s,t) + \sum_{m=1}^{M} \theta_m z_m(s,t) + W(s,t), \tag{3}$$

where $W(s,t)$ is a latent spatio-temporal process, $\theta = c(\alpha, \{\theta_m, m = 1, ..., M\})$ are the model parameters, $\text{offset}(s,t)$ is the offset term described in the data section and $\{z_m(s,t), m = 1, ..., M\}$ are the spatio-temporal covariates.

The total unemployed in any given area A and quarter t is given by

$$N(A,t) = \sum_{s_j \in A} y(s_j,t)N(s_j) \tag{4}$$

$N(s_j)$ is the number of dwellings in the residential building at s_j. Here, the number of s_j in A and $N(s_j)$ are known, fixed values.

We denote by $\mathbf{x}(t) = (s_j, t, y(s_j,t), z(s_j,t))$ the observed data obtained from the sampling survey in quarter t, and $\mathbf{x} = (\mathbf{x}(1), ..., \mathbf{x}(9))$ the collected data in the 9 quarters of study.

Our specific target quantities are the posterior predictive mean and variance of the random variable $N(A,t)$ given by respectively

$$\begin{aligned} \mathrm{E}(N(A,t)|\mathbf{x}) &= \mathrm{E}_{(W(s,t),\theta|\mathbf{x})}[\mathrm{E}(N(A,t)|\mathbf{x}, W(s,t), \theta)] \\ &= \int_{W(s,t),\theta} \mathrm{E}(N(A,t)|\mathbf{x}, W(s,t), \theta) p(W(s,t), \theta|\mathbf{x}) ds dt d\theta, \end{aligned} \tag{5}$$

and

$$\begin{aligned} \mathrm{Var}(N(A,t)|\mathbf{x}) &= \mathrm{Var}_{(W(s,t),\theta|\mathbf{x})}\left[\mathrm{E}(N(A,t)|\mathbf{x}, W(s,t), \theta)\right] \\ &+ \mathrm{E}_{(W(s,t),\theta|\mathbf{x})}\left[\mathrm{Var}(N(A,t)|\mathbf{x}, W(s,t), \theta)\right]. \end{aligned} \tag{6}$$

Here, the quantities of interest are simpler versions of the similar quantities given in Pereira *et al.* (2019), as the exact position of all residential buildings as well as the number of dwellings in each residential units are now known, so that counts based on residential units are no longer treated as random.

The mean and the variance of the predictive distribution given in (5) and (6) can be calculated numerically, as we can sample from the joint posterior density of the parameters as well as from the predictive distribution of $N(A)$ within the INLA platform. We used 1000 samples from an approximated posterior of the fitted model, calculated in the INLA platform. The technical details of such calculations are given in Pereira *et al.* (2019) and will largely be omitted in this manuscript.

4.1. Model fitting

As we introduce in the last section, we will assume the following hierarchical model

(a) Data|Parameter

$$y(s,t) \sim \text{Poisson}(\lambda(s,t)) \tag{7}$$

(b) Parameter|Hyperparameters

$$\log(\lambda(s,t)) = \alpha + \text{offset}(s,t) + \sum_{m=1}^{M} \theta_m Z_m(s,t) + W(s,t), \tag{8}$$

where $W(s,t)$ is the latent spatio-temporal process, as described in Blangiardo *et al.* (2015):

$$W(s,t) = aW(s,t-1) + \xi(s,t) \tag{9}$$

with $t = 1, ..., 9$, $|a| < 1$, and $W(s,1) = W_0 + \xi(s,1)$, where $W_0 \sim Normal(0, \sigma^2/(1-a^2))$. The term $\xi(s,t)$ is a Gaussian field with mean zero, temporally independent and with the following covariance function

$$cov(\xi(s,t), \xi(j,u)) = \begin{cases} 0, & if\ t \neq u, \\ cov(\xi(s), \xi(j)), & if\ t = u. \end{cases}$$

(c) Hyperparameters

$$\alpha \sim N(0, 1000) \tag{10}$$

$$\theta_m \sim N(0, 1000),\ m = 1, ..., M \tag{11}$$

We assume that the latent field ξ belongs to the Matern class with $\nu = 1$.
The Gaussian field ξ can be approximated by a Gaussian Markov random field, which is a discretized representation. That approximation is based on the stochastic partial differential equation (SPDE) approach (see Lindgren *et al.*, 2011), and depends on a triangulation, called a mesh, of the spatial domain. Figure 3 shows the mesh we considered in this study. For details about this approximation, the reader is referred to Pereira *et al.* (2019).

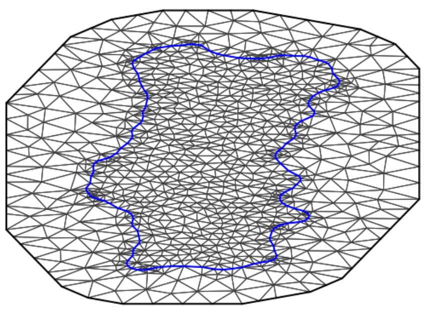


Figure 3. mesh with 913 vertices

Table 1. DIC, WAIC and the effective number of parameters

model	DIC	$WAIC$	p_{DIC}	p_{WAIC}
α	105324.06	105324.17	1.16	1.28
$\alpha + \text{offset}$	105120.28	105120.39	1.16	1.28
$\alpha + \text{offset} + age$	104799.76	104800.01	2.16	2.42
$\alpha + \text{offset} + age + edu$	104801.57	104801.89	3.16	3.49
$\alpha + \text{offset} + age + IEFP$	104459.93	104460.23	3.16	3.46
$\alpha + \text{offset} + age + IEFP + W$	103550.23	103569.02	306.41	324.02

Since it was necessary to know the covariates and offset in the locations of the observations and in those of the mesh nodes, we predicted them using a kernel estimation method, as we explained in section 3 (see Figure 2).

For an introduction to computational Bayesian statistics, the reader is referred to Amaral Turkman *et al.* (2019). Here, the inference was made using the INLA approach. See Rue *et al.*, 2009 and Rue *et al.*, 2017 for a better understanding of this approach.

For the model selection we used the deviance information criterion (DIC) and Watanabe-Akaike information criterion (WAIC), proposed by Spiegehalter *et al.* (2002) and Watanabe (2010) respectively. Table 1 shows that the best model is the one that considers the offset, the age, the proportion of unemployed people registered in the employment centers and the spatio-temporal random effects.

5. Results

To perform the spatial prediction, we created a regular grid of $1km^2$ in the domain. A projection from the mesh to the grid was performed and the resultant map of the posterior mean of the average number of unemployed people per dwelling at location s and quarter t, $\lambda(s,t)$, is shown in Figure 4. We can see that the average number of unemployed people per dwelling is higher in the Porto, Peninsula de Setubal and Alentejo regions. We also see a slight decrease of this indicator across time.

We generate 1000 samples from an approximated posterior of the fitted model, using the INLA function *inla.posterior.sample*, to estimate the target quantities, $E(N(A,t)|\mathbf{x})$, through Monte Carlo sampling. The logarithmic transformation of these quantities are given in Figure 5. As we might expect, the highest values are in Lisbon and Porto, where the population dimension is higher.

The aggregation of these estimates by NUTS III regions are shown in Figure 6. See Figure 7 for a better understanding of the temporal evolution in the study period. Here, we can see that there was a slight decreasing tendency across time during the study period for the majority of regions.

In addition to the spatial predictions, this methodology also allows us to make temporal predictions on unemployment. Figure 8 gives the estimates of unemployed people by NUTS III regions for the 4th quarter of 2016, and the respective temporal prediction. For this temporal prediction, we considered the covariates evaluated for the 4th quarter of 2016. As we can see, the estimates and the temporal predictions are similar in all regions. However, the coefficients of variation of the temporal predictions are higher than the coefficients of the estimates (see figure 9), as we would expect. In any case, it should be noted that the CVs are lower than 15% for all regions, even for the temporal predictions.

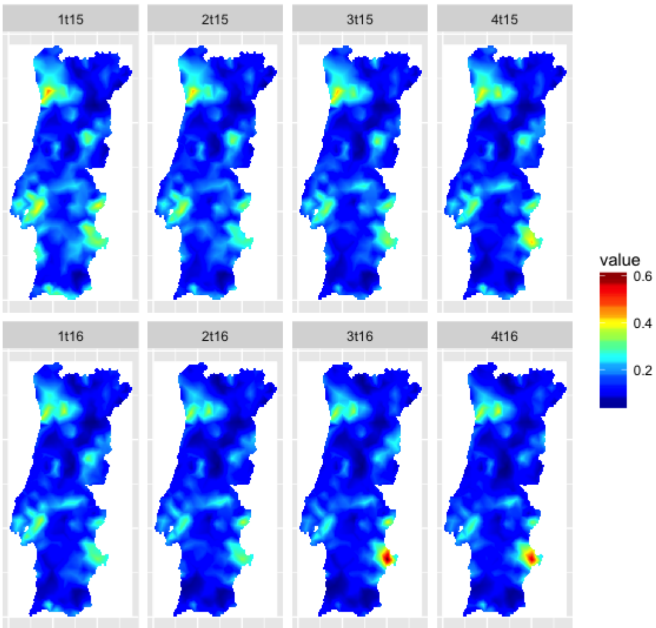


Figure 4. Posterior mean of the average number of unemployed individuals per dwelling by grid cell

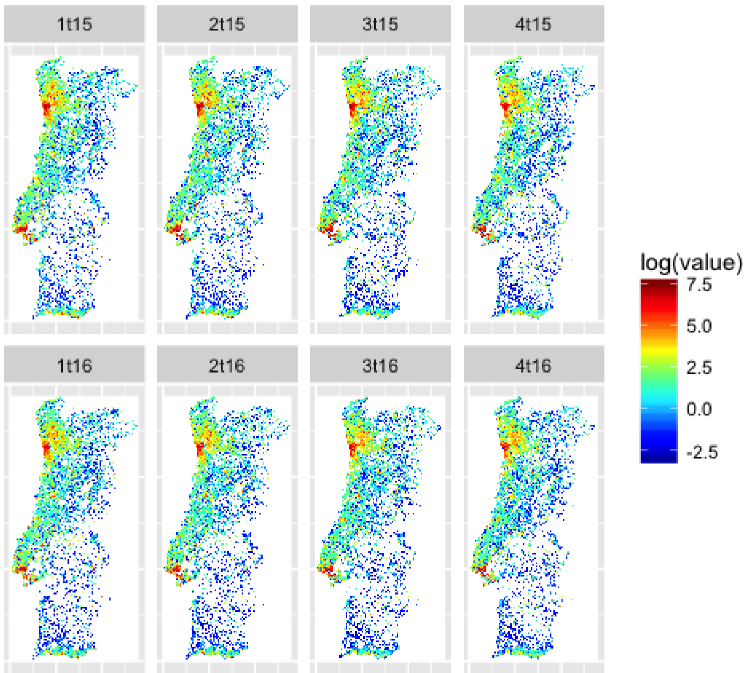


Figure 5. Logarithmic transformation of the posterior predictive mean of total unemployed individuals by grid cell. The white cells in the figure correspond to low populated cells for which the log-value of expected unemployment is so low it would shift the whole scale.

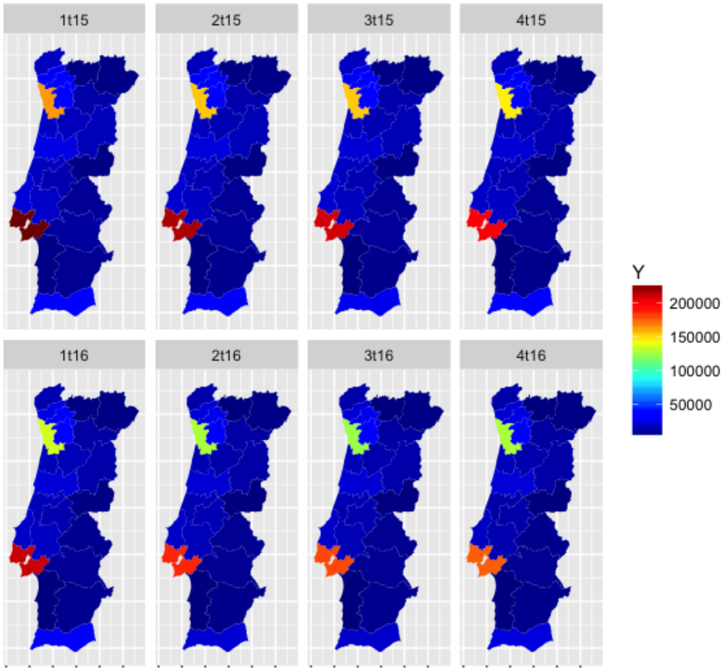


Figure 6. Posterior predictive mean of the total unemployed individuals by NUTS III regions

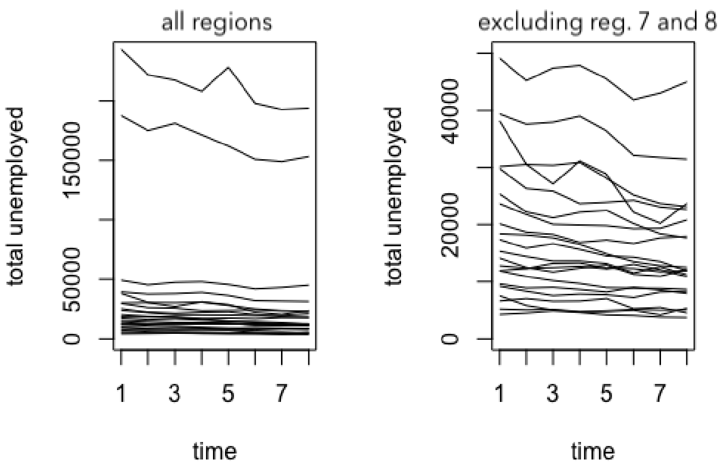


Figure 7. Temporal evolution of the total unemployed for the 23 NUTS III regions (left) and a zoom on the 21 regions with lower values, i.e., excluding Lisbon and Porto regions (right).

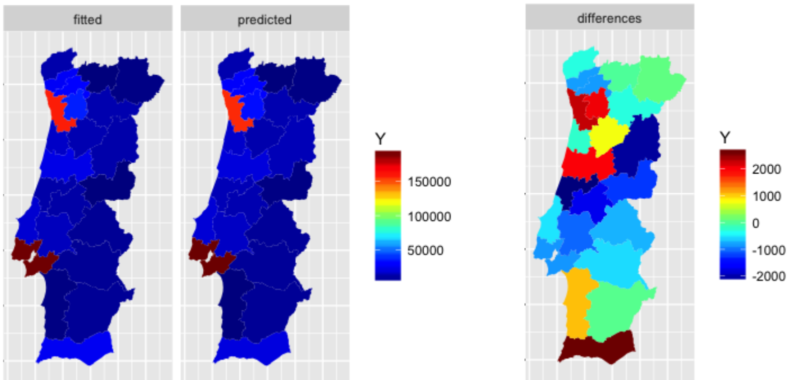


Figure 8. Spatial and spatio-temporal predictions for the 4th quarter of 2016 (left), and the differences (right). The first map was obtained using data from the 4th quarter of 2014 to the 4th quarter of 2016 and the second one was obtained excluding the last quarter in the data

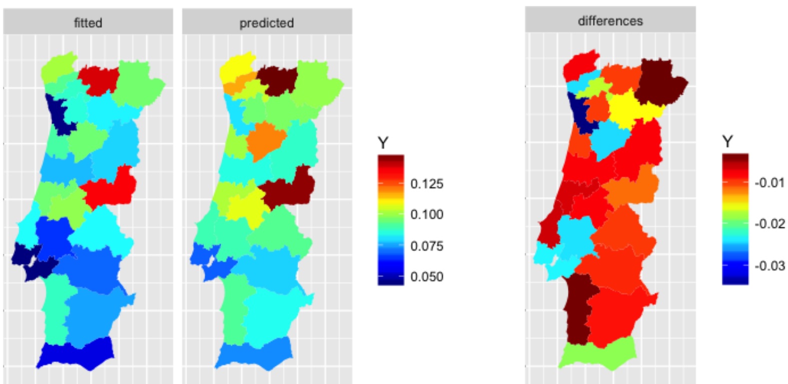


Figure 9. Coefficients of variation of the estimates obtained using the spatial and spatio-temporal predictions (left), and the differences (right).

6. Comparison between the results of LGCP model (Pereira *et al.*, 2019) and geostatistical model

For a comparison between the log Gaussian Cox process (LGCP) model, based on a spatial point process approach, proposed in Pereira *et al.* (2019) and the geostatistical model here proposed, we considered the respective spatial versions (without temporal extensions), using only data from the 4th quarter of 2016. The new version of the geostatistical data model uses the same mesh as was used in the spatial point processes model to be comparable. We also considered the same set of covariates in the two approaches.

Figure 10 shows the estimates of the total number of unemployed people by NUTS III regions (NUTS-2013) for the 4th quarter of 2016, using both approaches, and the respective coefficients of variation (CV, defined as the ratio of the standard deviation to the mean). The x-scale in these graphs represents the NUTS III regions without any specific order. Although significant differences are not visible in the estimates produced by each of the models, the coefficients of variation are distinguishable. Usually the portuguese National Statistical Office requires CVs lower than 20% for the estimates to be published as official figures. The two methods proposed respect and adhere to this requirement. Since the point-referenced data model does not require the modelling of the points, we expect less variability in comparison with the spatial point processes model.

Figures 11 and 12 permits a better analysis of the results in space. Although the spatial distribution of the estimates is similar in both methods, the CVs have a different behaviour. The regions with the highest CVs using the LGCP model, Beira Baixa and Baixo Alentejo, are not highlighted in the geostatistical CVs map.

In Pereira *et al.* (2019) we gave a comparison between the LGCP model and the traditional SAE methods. The LGCP model was highlighted as the model with better precision in high spatial resolutions. Thus, the method we suggest now seems to be the best way of estimating unemployment in small areas.

In addition to these results, it is important to note that the LGCP model and the geostatistical model bring many advantages in comparison with the direct method and areal data models (Pereira *et al.*, 2019). A summary of some of these advantages are the possibility of providing estimates for all counties or in even more detailed geographical regions, the coherence between different geographical levels, and the provision of information about the number of unemployed people per residential building, while taking into account specific information about the families.

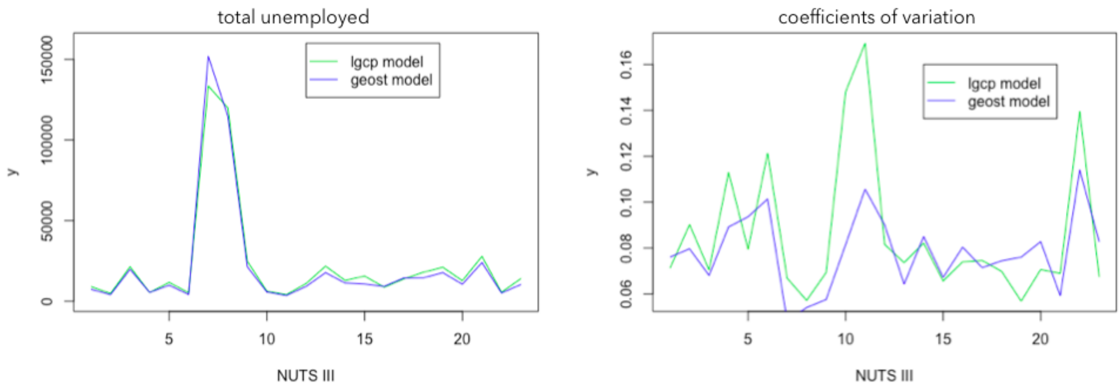


Figure 10. Estimates of total unemployed by NUTS III for the 4th quarter of 2016 (left), and the respective coefficients of variation (right)

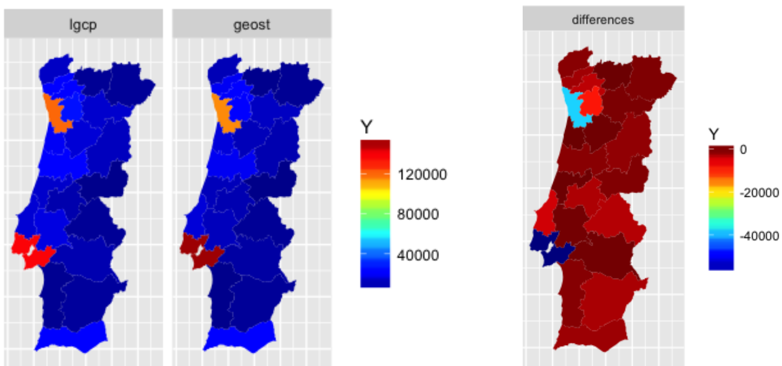


Figure 11. Estimates of total unemployed by NUTS III for the 4th quarter of 2016 using the LGCP model and the geostatistical model (left), and the differences (right)

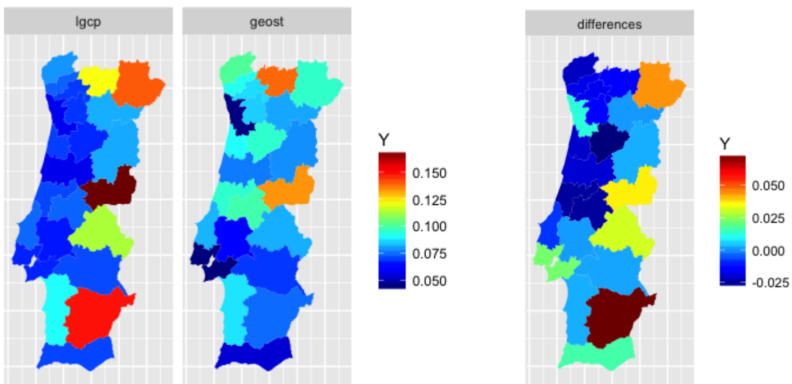


Figure 12. Coefficients of variation of the estimates obtained by the LGCP model and geostatistical model (left), and the differences (right)

6.1. Sensitivity analysis

The choice of mesh, the smoothing bandwidth (the standard deviation of the Gaussian kernel density) used for the prediction of covariates, and the priors used for the SPDE effects, can all affect the results significantly. Therefore, a sensitivity analysis is recommended. Figure 13 shows the 95% credible intervals by NUTS III region, using two different meshes (one with 3923 nodes and another with 913 nodes), two different smoothing bandwidths ($k = 5$ and $k = 20$), penalising complexity (PC) priors for the SPDE effects and using the default priors in the package INLA. As we can see, the model estimates seem to be sensitive to the mesh used. For the majority of regions, the credible intervals are lower for the mesh with 3923 nodes. Indeed, the estimates obtained with the most detailed mesh, present higher precision, as we would expect. In this study, the PC priors used for the SPDE parameters did not produce significant changes in the estimates, as we can see when comparing Figures 13 a) and c). In Lisbon region (index 7 on Figure 12), the population dimension is high, therefore we expect that the direct method should perform well. Notice that in this region the model using $k = 5$ is closer to the direct method. For this reason, we believe that the choice $k = 5$ is most appropriate in this study.

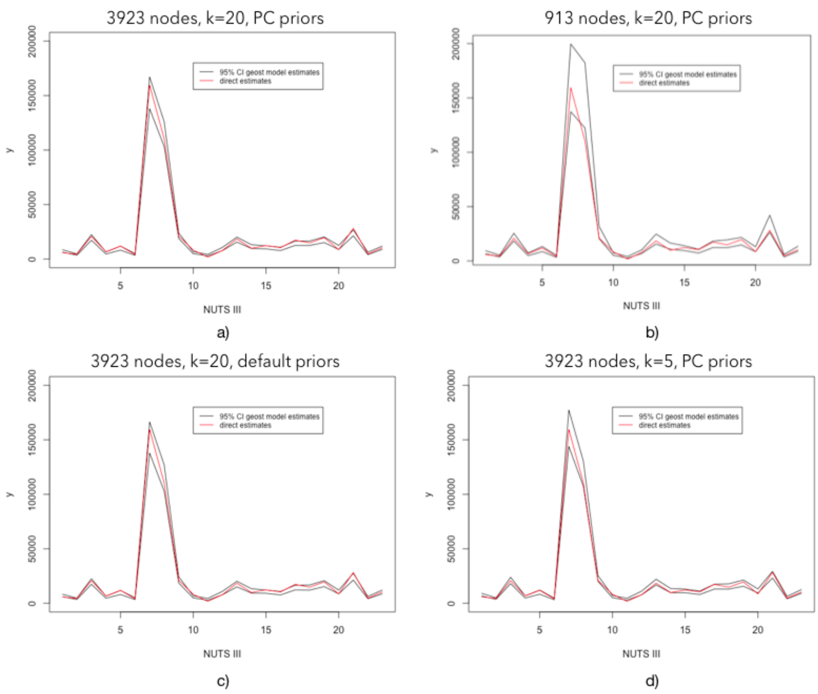


Figure 13. 95% Credible intervals for the total unemployed estimates and the direct estimates by NUTS III regions, using: a) mesh with 3923 nodes, $k = 20$ in the covariates prediction method, PC priors for SPDE effects; b) mesh with 913 nodes, $k = 20$ in the covariates prediction method, PC priors for SPDE effects; c) mesh with 3923 nodes, $k = 20$ in the covariates prediction method, default priors for SPDE effects; d) mesh with 3923 nodes, $k = 5$ in the covariates prediction method, PC priors for SPDE effects

7. Discussion, conclusions and further work

In this study we looked at unemployment data from a new perspective. In most National Statistical Institutes, the unemployment estimation is made using a direct method. However, although some of these institutes are starting to use areal models to produce accurate estimates for small areas, these models do not take into account specific information about the dwellings, and the geographical information is not sufficiently detailed.

During this last year, we proposed to look at unemployment through a marked spatial point process (Pereira *et al.*, 2019), where the points are the locations of the residential buildings and the marks attached are the total unemployed in each point.

Meanwhile however, the locations of all residential buildings in the national territory became available, so there is now no need to model its intensity, producing extra variability. Here, we proposed to look at unemployment data as geostatistical data, assuming that all locations of the residential buildings are known. Moreover, we considered a spatio-temporal extension, using data from the 4th quarter of 2014 to the 4th quarter of 2016.

This methodology not only provides unemployment estimates for every quarter in the study with good accuracy, but also for every area (counties, NUTS, etc) using the same model in a consistent way. Moreover, as we saw, the estimates obtained are more accurate than those produced by the spatial point processes model and, consequently (Pereira *et al.*, 2019), by the direct method and the traditional SAE methods.

We also concluded that the choice of the mesh used in the model is very important. Moreover, the model is sensitive to the smoothing parameter in the kernel smoothing used to perform the spatial prediction of the covariates included in the model. Therefore, we suggest that in these cases, a sensitive analysis must be conducted.

For future investigation, we think it would be interesting to conduct an elicitation of priors for the hyperparameters and compare these results with those obtained using PC priors.

Acknowledgements

This work was supported by the projects $UID/MAT/00006/2019$, $PTDC/MAT-STA/28649/2017$, and the PhD scholarship $SFRH/BD/92728/2013$ from Fundação para a Ciência e Tecnologia. Instituto Nacional de Estatística and Centro de Estatística e Aplicações da Universidade de Lisboa are the reception institutions.

Note

This study is the responsibility of the authors and does not reflect the official opinions of Instituto Nacional de Estatistica.

References

Amaral Turkman, M. A., Paulino, C. D., Muller,P. (2019) *Computational Bayesian Statistics*. Textbooks with ISBA.

Banerjee, S., Carlin, B. P., Gelfand, A. E. (2004) *Hierarchical Modeling and Analysis for Spatial Data.* Chapman and Hall/CRC.

Blangiardo, M., Cameletti, M. (2015) *Spatial and Spatio-temporal Bayesian Models with R-INLA.* Wiley.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2017) Constructing Priors that Penalize the Complexity of Gaussian Random Fields. arXiv:1503.00256

Krainski, E., Lindgren, F., Simpson, D., Rue, H. (2016) The R-INLA tutorial on SPDE models. http://www.math.ntnu.no/inla/r-inla.org/tutorials/spde/spde-tutorial.pdf

Lindgren, F., Rue, H., Lindstrom, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the SPDE approach (with discussion). *Journal of Royal Statistical Society Series B*, **73**, 423-498.

Pereira, S., Turkman, F., Correia, L. (2018) Spatio-temporal analysis of regional unemployment rates: A comparison of model based approaches. *Revstat.* **16**, 515-536.

Pereira, S., Turkman, F., Correia, L., Rue, H. (2019) Unemployment estimation: Spatial point referenced methods and models. *Spatial Statistics* (in press). https://doi.org/10.1016/j.spasta.2019.01.004

Rao, J.N.K. (2003) *Small Area Estimation.* New York: Wiley.

Rue, H., Martino, S., Chopin, N. (2009) Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). *Journal of the Royal Statistical Society Series B*, **71**, 319-392.

Rue, H., Riebler, A., Sorbye, S. H., Illian, J. B., Simpson, D. P., Lindgren, F. K. (2017) Bayesian computing with INLA: A review. *Annual Reviews of Statistics and Its Applications*, **4**, 395-421.

Simpson, D. P., Rue, H. , Riebler, A. , Martins, T. G. , and Sørbye, S. H. (2017) Penalising model component complexity: A principled, practical approach to constructing priors (with discussion). *Statistical Science*, **32**, 1-28.

Spiegelhalter, D. J., Best, N.G., Carlin, B.R., van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of Royal Statistical Society Series B*, **64**, 583-639.

Watanabe, S. (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571-3594.