A Bayesian time-varying random partition model for large spatio-temporal datasets

Giulio Beltramin[a], Andrea Cremaschi[b], Annalisa Cadonna[c], Alessandra Guglielmi[a], Fernando Andrés Quintana[d]

[a]*Department of Mathematics, Politecnico di Milano, Milan, Italy*
[b]*School of Science and Technology, IE University, Madrid, Spain*
[c]*Department of Statistics, University of Klagenfurt, Klagenfurt, Austria*
[d]*Departamento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile*

Abstract

Spatio-temporal areal data can be seen as a collection of time series which are spatially correlated, according to a specific neighbouring structure. Motivated by a dataset on mobile phone usage in the Metropolitan area of Milan, Italy, we propose a semi-parametric hierarchical Bayesian model allowing for time-varying as well as spatial model-based clustering. Our approach incorporates the notion of regimes that describe changing patterns over work and night hours as well as weekdays/weekends. Changes across regimes are considered by means of temporal changepoint components that allow for different hierarchical structures specified across time points. The changepoints might occur within fixed time windows over the day. The model features a novel random partition prior that incorporates the desired spatial features and encourages co-clustering based on areal proximity. We explore properties of the model by way of extensive simulation studies from which we collect valuable information. Finally, we discuss the application to the motivating data, where the main goal is to spatially cluster population patterns of mobile phone usage.

Keywords: Areal data, Bayesian Nonparametrics, Mobile data, Population density dynamics, Spatio-temporal clustering
62H11, 62Gxx, 62F15

1. Introduction

In recent years, there has been a rapid increase in the availability of data collected over time on areal units. Areal units can be defined by geographical boundaries (e.g., regions, counties, municipalities), or by tessellating the territory of interest. The available data are often large-to-huge collections of (possibly long) time series that are spatially correlated, according to a specific neighbourhood structure. Data of this type are extremely helpful to understand population distribution in an urban space, which is critical for urban planning and provision of municipal services; see, e.g. [13], [37], and [38]. However, traditional methods for exploring human population dynamics, such as censuses and surveys, can be very expensive. Moreover, more detailed changes over time, such as daily commute and urban transportation, are challenging to assess with traditional methods, but crucial nonetheless, to urban planning.

Mobile phone-based data have been extensively used to extract geographical knowledge in previous studies; see, for instance, [36] and [25]. Previous studies [see 40, and the references therein] suggest that mobile phone data are a valid proxy for human activities and interactions.

Motivated by the study of population density dynamics arising in the context of area-level mobile data, we specify an appropriate model and develop efficient algorithms for the statistical analysis of large-to-huge spatio-temporal areal data. The proposed hierarchical model takes into account various characteristics of the data, including (a) varying regimes corresponding to day/night and weekday/weekend times; (b) spatial dependence across areas; and (c) clustering of areal time series. We accomplish these goals by employing a novel Bayesian nonparametric prior for clustering longitudinal areal data.

The motivating dataset has been previously considered in [26] and [35], where each time series is modeled as the realization of a functional data process. One of the main goals of this work is to model the population dynamics, and in particular, to understand how this changes across different regimes, and thus our approach does not employ a functional data perspective but rather builds on a combination of harmonic regression models for the temporal component and conditionally autoregressive (CAR) models for the spatial association. CAR models are commonly used to represent spatial autocorrelation for areal data, and can be thought of as the conditional specification of a (Gaussian) Markov random field (GMRF). A comprehensive review of

GMRF models can be found in [33]. CAR models were originally introduced as spatial models in [4, 5], and they have been used since as the likelihood for the observations themselves in one-stage models, or as the distribution of spatial random effects, as part of a Bayesian hierarchical model [see 6]. The different CAR models proposed in the literature correspond to specific choices of the precision matrix for the corresponding GMRF. Thanks to the Markov property, the precision matrix of a GMRF is potentially very sparse, which enables efficient computation through linear algebra algorithms for sparse matrices. An efficient algorithm for block updating in the context of Markov Random Fields models is introduced in [22]. Different CAR prior specifications have been proposed, such as the intrinsic and Besag-York-Mollié priors [both in 6]. We adopt instead the formulation proposed in [24], which is useful to estimate spatial correlation among the random effects.

The current literature on spatial clustering does not include the partition of areal units in the random parameters of the model, but obtains cluster estimates through the application of empirical clustering methods (e.g. hierarchical clustering) to posterior summaries obtained from the data. In particular, focusing on the literature on non-Bayesian statistical and machine learning techniques, we mention spatio-temporal K-means [14], a 2-step procedure that creates clusters first using a K-means procedure that considers clusters at two consecutive time points and then classifies trajectories by assessing which cluster appears more frequently over time. The ST-DBSCAN algorithm for spatial–temporal data is employed in [7] by applying DBSCAN to spatial and temporal features in a simultaneous and independent fashion. A hierarchical clustering method with spatial constraints, `ClustGeo` is used in Chavent et al. [9], based on Ward hierarchical clustering. This is a 2-step procedure that first applies `hclust` to a feature matrix (e.g., variables of interest), and then "corrects" the clustering by taking as dissimilarity matrix a linear combination of the original dissimilarity (based on features) and the one computed with the geospatial information. Another approach based on hierarchical clustering is presented in [18], where Gaussian processes are employed to handle nonstationary time series. The mixed geographically weighted regression with spatial auto-correlation method `MGWRSAR` is implemented in the R package `mgwrsar` [15]. The authors consider SAR-type models with various combinations of constant and spatially-varying coefficients. A combination of geographically weighting with artificial neural networks, `GWANN`, is proposed by [17]. Neither `MGWRSAR` nor `GWANN` directly provide cluster estimates, but one can naively apply clustering techniques to

the estimated spatial effects or predicted values generated from these techniques; see Supplementary Section 3. Instead, the main advantage of our approach is that we directly model the partition of areas and make this a central object of posterior inference. Our prior, here denoted areal product partition model (aPPM), penalizes excessive areal disconnectedness through a spatial association parameter ξ. The prior is new in the BNP literature for random partition models, though it builds on ideas in [19]; see also [30] for an application to multivariate analysis of data from mosquito-borne diseases in Brazil that employs the Hegarty and Barry prior. Note that product partition models for estimating change points for a single time series were originally introduced in [2].

Finally, a different Bayesian nonparametric model for clustering has been proposed in [8] using the same dataset analyzed in this work. There are two main differences between their model and the one proposed in this work. First, in [8], areal units are clustered together if their behaviour is similar over the entire time series under analysis, while we allow for clustering changes across regimes. Second, [8] use an ANOVA-DDP prior which does not consider the neighbouring structure of the areal units, while the aPPM introduced in Section 3.2 explicitly takes this into account.

Our contributions to the analysis of spatio-temporal areal datasets can be summarized as follows: (i) we propose a novel spatio-temporal random partition model that incorporates well studied priors in the Bayesian parametric and nonparametric framework and that provides a satisfactory fit to the available data; (ii) we estimate spatially-driven clusters of areal regions, and assess their changes over regimes; (iii) we provide inference on change-points, where the dynamics of mobile phone usage transit between different regimes; (iv) we are able to handle missing data.

The remainder of the paper is organized as follows. Section 2 describes the data and motivations. Section 3 introduces the modelling framework and the proposed aPPM prior. Section 4 presents the application to population density dynamics in Milan, and Section 5 concludes. A Supplementary Materials file is provided, containing details on the tailored MCMC algorithm, simulation studies, additional results and figures, as well as comparisons with alternative approaches.

2. Description of the Dataset and Motivation

Motivated by the study of population density dynamics in the metropolitan area of Milan, Italy, we propose a model for spatio-temporal clustering of areal data. We consider data as in [35] (more details are reported below), though we focus on a smaller grid to understand the clustering of areas in the inner part of the city, which should be easy to interpret in terms of the city's particular features such as entertainment and restaurant areas, or neighbourhoods that are mostly residential. The municipality of Milan, located in the metropolitan area centre, is a major productivity and financial centre for the entire northern portion of Italy. About 1.3 million people live and 600 thousand people commute every working day between Milan and the metropolitan area. Using mobile phone data, we study how the dynamics of the population density evolves over time, potentially uncovering temporal and spatial patterns. Moreover, we would like to partition the areas covered by the data into subregions sharing a similar population dynamic pattern across the observed time window. This is of great interest to urban planners, city managers and network providers. For instance, the data in [35] is also analyzed in [26], who investigated relevant urban usage by proposing diversified management policies for increased efficiency of public services supply. [40] propose a Bayesian spatio-temporal model focusing on area-level mobile phone users data. Specifically, their data are the total number of mobile phone users actively recorded by cell towers in one-hour intervals in an overall time window of 24 hours in Shenzhen, China. They do not assume a random partition, and infer the clustering structure of the spatial data via a hierarchical clustering procedure on the estimated time profiles. They nevertheless conclude that their Bayesian spatio-temporal model can enhance the understanding of the space-time variability of population distribution using mobile phone data.

The dataset we consider is provided by Telecom Italia, the largest mobile company in Italy, as part of the Green Move Initiative (financed by Regione Lombardia), through a research agreement between Telecom and Politecnico di Milano. See further details in [35]. These authors segment the metropolitan area of Milan into subregions that share the same activity pattern along time in terms of population density dynamics. To this end, they integrate a Treelet analysis for dimensional reduction with a Bagging Voronoi strategy for the exploration of spatial dependence, in order to reduce the dimension of spatially dependent signals. They propose the Bagging Voronoi Treelet al-

gorithm, that decomposes the massive dataset ($10,573$ areas for $1,308$ time points, resulting in more than 13 millions records) into relevant spatial and temporal dynamics.

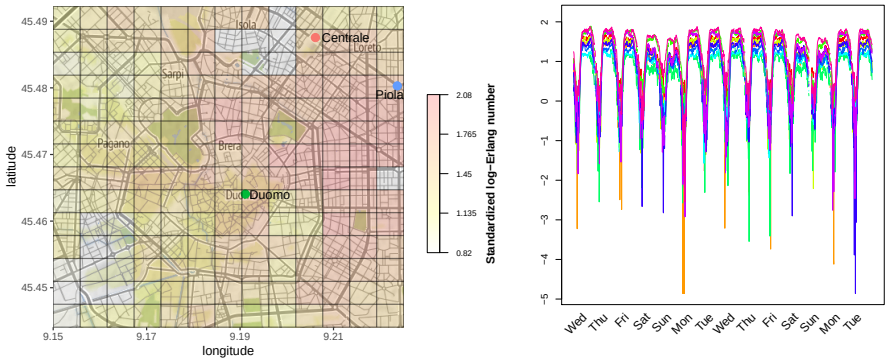
In this work, we focus on a portion of the metropolitan area of Milan between 45.444° and 45.49° in latitude and between 9.15° and 9.225° in longitude. This portion covers approximately the area inside a city belt called *circonvallazione* (i.e., circumvallation), hosting a significant flow of private and public transportation vehicles during rush hours. The districts inside this belt line can be considered the city centre. We have partitioned the central area into a grid of $I = 13 \times 14$ sites (areal units).

Each areal unit was recovered from the original data by Telecom, putting together 4 of the original sites from a uniform lattice of an area including the metropolitan area though we consider only inner areal units as stated above. See Figure 1 (a) presenting the portion of the central municipality of Milan under study. The data are recorded every 15 minutes from March 18th 2009, 00:15 to March 31st 2009, 23:45, yielding a time series of length $T = 1343$ for each of the lattice sites. In total, $244,608$ records are available of which $22,068$ are missing. We analyze the Erlang number, calculated as the sum of lengths of all calls in a given time interval, divided by the length of the interval. In other words, the Erlang number is equivalent to the average number of mobile phones simultaneously calling through the network, and can be considered proportional to the number of active users [see 35, for more details]. The Erlang number, recorded over all areal units in a region, can be used as a proxy for population density in that region, and through its changes over time, of population density dynamics.

Figure 1(a) shows the standardised log-Erlang number recorded on Wednesday, March 18th 2009 at noon, for the selected portion of the metropolitan area. The financial district in the centre of the city is identifiable, as it is the area with high mobile activities during working hours, as well as the eastern area corresponding to a busy portion of the city. We can also identify, for example, peripheral areas with less mobile traffic in the western part. Locations on the grid corresponding to missing observations are left blank, showing only the underlying map of the city. In what follows, we always report summary statistics or plots of globally standardised log-Erlang numbers.

Figure 1(b) shows the time series for ten randomly selected locations. As we can see, the mobile activity is higher during the day hours and lower at night. Moreover, we can see differences between workdays and weekends. We consider these temporal patterns in the proposed model in such a way

that the clustering structure changes over time through a simplified approach based on the notion of temporal regimes (workdays and weekends, days and nights). Another distinguished feature of our Bayesian model is that we can handle missing responses in a natural fashion unlike, for instance, [35].



(a) (b)

Figure 1: Telecom data. (a) Data recorded on Wednesday March 18th 2009, at noon. The gray areas correspond to missing observations. Points of interests as Piazza Duomo, Stazione Centrale and Piazza Piola are denoted as colored points on the map. (b) Time series for ten randomly selected locations. Data shown are standardised log-Erlang numbers for all areal units.

3. Model specification

We can frame the problem considered as the clustering of the same I items (the 182 areal units in Milano) at different times $t = 1, \dots, T$. Hence, the parameters of interest are the random partitions of the same total area at each time point $t = 1, \dots, T$. We could thus use the model to estimate one partition for each time t. By the spatial connotation of the data, it is clear that the marginal prior of any random partition has to incorporate information on the spatial structure. By this we mean that the probability that two neighbouring areas co-cluster should be larger than the probability of including in the same cluster two areas far apart. However, because $T = 1343$ is a large number and we expect to see little or no variation between the cluster estimates separated by 15 minutes (the gap time between

observations), we focus on the notion of regimes and assume that there are change points over time where the key parameter, the random partition of the same I areal units, changes, as in changepoint models for a single time series.

We now introduce the spatio-temporal hierarchical model for the motivating dataset, which has three main components, namely: (1) the likelihood part for (standardised) log-Erlang numbers, based on a mixed regression with time-varying coefficients and spatially-correlated random effects, in Section 3.1; (2) the random partition prior specification for areal units, which incorporates information on the spatial structure for each regime in Section 3.2; and (3) the prior specification for the regimes (Section 3.3). The final dynamic Bayesian model is described in Section 3.3.

3.1. Likelihood and spatial random effects

Consider I areal units, in our case given by tessellations of part of the city of Milan, as described in Section 2. Denote by Y_{it} the observation for areal unit $i = 1, \ldots, I$ at time $t = 1, \ldots, T$, yielding a dataset of $I \times T$ observations. In our application, the response variable Y_{it} is the standardised logarithm of the Erlang number in area i at time t. To avoid null Erlang numbers, before applying the logarithmic transformation to the data, we transform them by adding a quantity equal to the smallest observed non-zero Erlang number. It is important to remark that a zero Erlang number does not mean zero activity, but that this fell below a certain detection threshold. Then, the log-transformed data are standardised across all areas and timestamps to have mean 0 and standard deviation 1. We applied these transformations to ensure that the data support is not limited to the positive reals, allowing the observations to be modeled as Gaussians. For each areal unit $i = 1, \ldots, I$, and time $t = 1, \ldots, T$, we model observation Y_{it} as:

$$Y_{it} \mid \boldsymbol{x}_t, \tilde{\boldsymbol{\beta}}_{it}, \tilde{u}_{it} = \boldsymbol{x}_t' \tilde{\boldsymbol{\beta}}_{it} + \tilde{u}_{it} + \epsilon_{it}, \quad \epsilon_{it} \overset{\text{ind}}{\sim} \mathrm{N}\left(0, \sigma^2_{\epsilon_t}\right)$$

where ϵ_{it}, for $i = 1, \ldots, I$ and $t = 1, \ldots, T$, are conditionally independent spatio-temporal residuals. The zero-mean $\tilde{u}_{it}$'s are defined below. Here, $\mathrm{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2. The p-dimensional extension of it will be used later and indicated as $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ for any positive integer p. The variance of ϵ_{it} in the full Bayesian model will depend on the regime associated to t. The mean surface in this model is given by $\boldsymbol{x}_t' \tilde{\boldsymbol{\beta}}_{it}$, through which we

model, at each location, the periodicity of the data by resorting to a harmonic regression. Concretely, we let $\boldsymbol{x}_t$ denote a p-dimensional design vector, chosen from the harmonic functions $(\cos(\omega_j t), \sin(\omega_j t))$, where $\omega_j = 2\pi j/T$ and $j = 1, 2, \ldots, T/2$. If T is not an even number, we assume T to be the next even number and the corresponding observation to be missing. See Supplementary Section 1 for details on Bayesian imputation of the missing Erlang numbers. The dataset of interest has $T = 1344 = 2 \times 7 \times 24 \times 4$ time measurements, since our time span is two weeks with four observations per hour. Since our responses are standardised, we exclude the intercept term in $\boldsymbol{x}_t$. Using harmonic regression corresponds to approximating each underlying signal through a trigonometric polynomial. Motivated by the characteristics of our dataset and pragmatic knowledge of Milan vehicle traffic, we choose to select weekly, daily, semi-daily and hourly frequencies in $\boldsymbol{x}_t$. This is equivalent to assuming $p = 8$, and

$$\begin{aligned}\boldsymbol{x}_t = &(\cos(\omega_2 t), \sin(\omega_2 t), \cos(\omega_{14} t), \sin(\omega_{14} t), \\ &\cos(\omega_{28} t), \sin(\omega_{28} t), \cos(\omega_{336} t), \sin(\omega_{336} t))\end{aligned} \tag{1}$$

For instance, to set the daily frequency, we assume $j = \frac{1344}{24\times 4} = 14$, so that $\omega_{14} = \frac{2\pi}{96}$. In this case $\cos(\omega_{14}(t + 96)) = \cos(\omega_{14} t)$, i.e. the function has period equal to 96, which corresponds to one day interval (4 observations in an hour for 24 hours). The vector $\tilde{\boldsymbol{\beta}}_{it} = \left(\tilde{\beta}_{it,1}, \ldots, \tilde{\beta}_{it,p}\right)'$ is the vector of harmonic coefficients and, as discussed later in Section 3.3, will be used to cluster the areal units.

Having controlled seasonality through $\boldsymbol{x}_t$ in our model, we now consider spatial autocorrelation. This is done via a spatial random effects vector $\tilde{\boldsymbol{u}}_t = (\tilde{u}_{1t}, \ldots, \tilde{u}_{It})'$ on which we put a spatial CAR prior, that is:

$$\tilde{\boldsymbol{u}}_t \mid \tau^2, Q\left(\zeta_t, W\right), \zeta_t \sim \mathrm{N}_I\left(\mathbf{0}, \tau^2 Q(\tilde{\zeta}_t, W)^{-1}\right) \tag{2}$$

where W is the $I \times I$ matrix encoding the contiguity structure of the I areal units specified as $W_{i,j} = 1$ if areal units i and j are neighbours and $W_{i,j} = 0$ otherwise. Here, we specifically define the neighbours of a site i as the 8 cells surrounding i in a grid layout. To help the description of W, we list its minor diagonals from the main diagonal to the bottom left corner. The only minor diagonals whose elements differ from zero are the first, 12th, 13th and 14th minor diagonals, while those from the 2nd to 11th contain only zero values, as do those from the 15th to 182th. Moreover, the matrix

W is a block-tridiagonal matrix with 14 blocks of dimension 13×13, with each block being tridiagonal itself. We next specify Q in (2) following the construction discussed in [24], that is, we set $Q\left(\tilde{\zeta}_t, W\right) = \tilde{\zeta}_t(\text{diag}(W\mathbf{1}) - W) + (1 - \tilde{\zeta}_t)\mathbb{I}_I$, where $\mathbb{I}_I$ is the I-dimensional identity matrix and $\mathbf{1}$ is an I-dimensional vector of ones. Let d_i be the number of neighbours of site i. The matrix $(\text{diag}(W\mathbf{1}) - W)$ has elements equal to d_i if $i = j$, equal to -1 if i and j are neighbours ($i \sim j$), and equal to 0 otherwise. Here, $\text{diag}(\boldsymbol{a})$ denotes a matrix with diagonal given by $\boldsymbol{a}$ and that is zero otherwise. In addition, parameter $\tilde{\zeta}_t$ controls the spatial autocorrelation structure: $\tilde{\zeta}_t = 1$ corresponds to the intrinsic CAR prior [6], where the conditional expectation is the mean of the random effects in geographically adjacent areal units. On the other hand, $\tilde{\zeta}_t = 0$ corresponds to independent random effects. The class of CAR models is large; see further details in [4], [10], [21], [11], [32], and references therein.

We introduce next the concept of *regime*. Given the nature of the data, we do not actually expect the parameters of the model to vary at each time t. Instead we expect to observe different states (called regimes), each with a specific set of parameters. A similar notion has been employed in [3], who consider switching weather regimes when modelling Tropical Pacific sea surface temperatures. The choice of number of regimes in our context is based on information about the evolving dynamics of the system. In our specific application, we assume the number of regimes to be based on the days of the week (weekday/weekend), and on the period of the day (night/day). The regime indicators are denoted by $r_t \in \{1, \ldots, n_R\}$, for $t = 1, \ldots, T$, where n_R is the number of regimes allowed in the model. In light of the previous discussion, we assume some of the model parameters to be regime-specific, namely the harmonic regression coefficients: at each location i, we have $\boldsymbol{\beta}_{i1}, \ldots \boldsymbol{\beta}_{in_R}$. In addition, we consider n_R I-dimensional vectors of spatial random effects, $\boldsymbol{u}_1, \ldots \boldsymbol{u}_{n_R}$, as well as regime-specific scaling parameters $\tau_1^2, \ldots, \tau_{n_R}^2$ and observation variances $\sigma_{\epsilon_1}^2, \ldots, \sigma_{\epsilon_{n_R}}^2$. Finally, the parameters of the spatial precision matrix Q are also regarded as regime-specific, $\zeta_1, \ldots, \zeta_{n_R}$. Using the correspondence between time point t and the regime present at that time r_t, we have that, at each t, $\tilde{\boldsymbol{\beta}}_{it} = \boldsymbol{\beta}_{ir_t}$, $\tilde{\boldsymbol{u}}_t = \boldsymbol{u}_{r_t}$ and $\tilde{\zeta}_t = \zeta_{r_t}$. This implies a substantial reduction in the number of model parameters. In other words, we distinguish between $\{\tilde{\zeta}_t\}$, the time-specific spatial effects, and ζ_{r_t}, the regime-specific spatial association parameter, while establishing a simple mapping among these. A similar argument applies to the time-dependent regression

parameters and the random effect parameters. Conditionally on regimes, we can rewrite the model as follows:

$$
\begin{aligned}
&Y_{it} \mid \boldsymbol{x}_t, \boldsymbol{\beta}_{ir_t}, u_{ir_t} = \boldsymbol{x}_t' \boldsymbol{\beta}_{ir_t} + u_{ir_t} + \epsilon_{it}, \quad \epsilon_{it} \overset{\text{ind}}{\sim} \mathrm{N}\left(0, \sigma^2_{\epsilon_{r_t}}\right) \\
&\boldsymbol{u}_{r_t} \mid \tau^2_{r_t}, Q\left(\zeta_{r_t}, W\right), \zeta_{r_t} \sim \mathrm{N}_I\left(\mathbf{0}, \tau^2_{r_t} Q\left(\zeta_{r_t}, W\right)^{-1}\right)
\end{aligned}
$$

where each $\boldsymbol{\beta}_{ir_t}$ is p-dimensional, while $\boldsymbol{u}_{r_t}$ is an I-dimensional parameter vector. The marginal priors for the parameters, as well as the whole model specification, are given in Section 3.3.

3.2. Areal Product Partition Model (aPPM)

We next discuss the time-varying spatial clustering structure component of the proposed model for the I time series of responses across regions. As commonly done in Bayesian nonparametrics, we define a clustering model by considering a prior distribution for the random partition parameter $\rho_r = \{C_1^r, C_2^r, \ldots, C_{K_r}^r\}$ that denotes the partition of areas $\{1, 2, \ldots, I\}$ in the sample at regime r. We define a prior distribution by building on product partition models [PPM, see, e.g., 31], and by modifying the spatially-oriented PPM proposed in [19]. To explain the proposal, we recall these models here; for notational convenience, we drop the regime subscript r until the end of this section. Under a PPM, the distribution on a partition ρ of a set of indices $[I] = \{1, \ldots, I\}$ into K subsets $\rho = \{C_1, C_2, \ldots, C_K\}$ is constructed in terms of a *cohesion* function for any subset $C_j \subset [I]$, which measures the strength of prior belief that the elements of C_j are to be grouped together. The PPM prior is thus expressed as

$$
p(\rho = \{C_1, \ldots, C_K\}) = \mathcal{K} \prod_{j=1}^{K} c\left(C_j\right), \qquad \rho \in \mathcal{P}\left([I]\right)
$$

where $\mathcal{P}([I])$ denotes the set of all partitions of $[I]$ and $\mathcal{K}$ is an appropriate normalizing constant depending on the cohesion function. A typical choice of cohesion function is $c(C_j) = \kappa \times \Gamma(n_j)$, with $n_j = |C_j|$, for $j = 1, \ldots, K$ from which we recover the exchangeable partition probability function (EPPF) corresponding to the Dirichlet process (DP) with mass parameter $\kappa > 0$.

Recent developments in the study of PPMs have explored the possibility of adding additional prior information to the definition of c, such as covariates [28, 29]. In our approach, we include the information about the areal

structure of the data, and we do so by following the idea of [19], who introduce the notion of *boundary length* $\ell^j(i)$ of the i-th areal unit belonging to cluster C_j as the number of neighbours of i that do not belong to C_j, for $j = 1, \ldots, K$. The boundary length of a cluster C_j is computed as the sum of the boundary length of each areal unit in C_j. The model by [19] is thus given by

$$p(\rho = \{C_1, \ldots, C_K\} \mid \xi) = \mathcal{K}(\xi, I) \prod_{j=1}^{K} \mathrm{e}^{-\xi \sum_{i \in C_j} \ell^j(i)} \tag{3}$$

where the normalizing constant $\mathcal{K}(\xi, I)$ is a function of I and of the positive hyperparameter ξ, so that

$$\frac{1}{\mathcal{K}(\xi, I)} = \sum_{\rho \in \mathcal{P}([I])} \prod_{j=1}^{K} \mathrm{e}^{-\xi \sum_{i \in C_j} \ell^j(i)}$$

As a prior for the random partition ρ, the proposed model combines the DP-based PPM and the spatial prior by [19], with the difference that we consider as neighbours the eight areal units surrounding $i \in [I]$, and not just the ones with a common side. Concretely, we consider

$$p(\rho = \{C_1, \ldots, C_K\} \mid \kappa, \xi) = \mathcal{K}(\kappa, \xi, I)\, \kappa^K \prod_{j=1}^{K} \Gamma(n_j)\, \mathrm{e}^{-\xi \sum_{i \in C_j} \ell^j(i)} \tag{4}$$

where $\mathcal{K}(\kappa, \xi, I)$ is now a function of the positive hyperparameters κ, ξ and the number I of items to cluster. Compared to (3), prior (4) introduces an additional term, that coincides with that of the product form arising from the Dirichlet process prior when written in PPM form. In particular, following the notation in [19], the normalizing constant $\mathcal{K}(\kappa, \xi, I)$ is such that

$$\frac{1}{\mathcal{K}(\kappa, \xi, I)} = \sum_{\rho \in \mathcal{P}([I])} \prod_{j=1}^{K} \kappa \Gamma(n_j)\, \mathrm{e}^{-\xi \sum\limits_{i \in C_j} \ell^j(i)}$$

Thus, the number of clusters a priori grows with κ. On the other hand, ξ is related to spatial association by penalizing excessive spatial disconnectedness between subsets (i.e. small values for the boundary length), in the sense that a larger ξ encourages fewer clusters. The combination of both approaches strikes a balance between a "rich gets richer" DP-based clustering

and a spatially-oriented setting proposed by [19]. As shown in Supplementary Section 2, we conduct simulation studies aimed at understanding the intertwining role of κ and ξ in controlling the prior partition structure.

To further understand the partition structure imposed by the proposed model (4), set $\eta = \exp(-\xi) \in (0, 1)$ and consider one, two or three areas, i.e., $I = 1, 2, 3$. Table 2 shows the prior probability for different configurations given by the proposed prior distribution, up to a normalizing constant. When $\eta = 1$ (equivalently, $\xi = 0$), the partition probabilities reduce to the DP prior case. This means that, a-priori, $\kappa > 1$ will assign more probability to partitions composed of singletons, while $0 < \kappa < 1$ to the partition with only one cluster. Considering a fixed value of $\kappa > 1$, similar considerations can be made for the value of η. Notice that the partition with one cluster is always favored when $\kappa \leq 1$ and that, in the case of three areas depicted in Table 2, two different areal configurations are allowed, yielding different a-priori probabilities. In particular, the partition with three clusters is penalised more when the areal units are not aligned, since in our 8-neighbours setting all units are in contact with each other. Secondly, the partitions with two clusters in the second scenario have the same prior probabilities, while in the first scenario the prior probability depends on the specific partition. The partition in which areas 1 and 3 are clustered together is less probable than the one in which 1 and 2 or 2 and 3 are clustered together. Moreover, when κ is fixed and $\eta \to +\infty$, the probability of the partition with one cluster tends to one. This simple study shows the spatial "local" effect of the parameter η on the clustering.

When the number of areal units is larger than $I = 3$, it is difficult to see analytically what this implies. We propose in this section a simulated example aimed at understanding the properties of the distribution of the partition a-priori. We show results applied to a rectangular grid of size 14x13, the same as the one used in the Telecom data application, yielding $I = 182$ areal units. The sensitivity analysis reported below presents the distribution of the number of clusters and the properties of the partition a-priori under the proposed model. For comparison, we also consider the [19] prior. MCMC samples from this model follow easily from the Gibbs sampler algorithm in Section 1 of Supplementary Material by suitably dropping the likelihood and DP parts. For the proposed model, we fix κ to 1 so that the expected number of clusters under the regular DP prior is $\mathbb{E}(K) = \sum_{j=1}^{I} \kappa/(\kappa + j - 1) \approx 5.78$. We also select a grid of values for ξ for comparison; see Supplementary Figures 13 and 14.

# areal units	Configuration	Partition	Prior probability $\propto$
1	$\boxed{1}$	$\{1\}$	1
2	$\boxed{1}\boxed{2}$	$\{1,2\}$	κ
		$\{1\},\{2\}$	$\kappa^2\eta^2$
3	$\boxed{1}\boxed{2}\boxed{3}$	$\{1,2,3\}$	2κ
		$\{1\},\{2,3\}$	$\kappa^2\eta^2$
		$\{1,3\},\{2\}$	$\kappa^2\eta^4$
		$\{1,2\},\{3\}$	$\kappa^2\eta^2$
		$\{1\},\{2\},\{3\}$	$\kappa^3\eta^4$
	$\boxed{2}\boxed{3}$ $\boxed{1}$	$\{1,2,3\}$	2κ
		$\{1\},\{2,3\}$	$\kappa^2\eta^4$
		$\{1,3\},\{2\}$	$\kappa^2\eta^4$
		$\{1,2\},\{3\}$	$\kappa^2\eta^4$
		$\{1\},\{2\},\{3\}$	$\kappa^3\eta^6$

Figure 2: Prior probabilities (up to adequate normalising constants) for different configurations of one, two or three adjacent areal units. The probabilities are expressed as function of the mass parameter κ and the boundary length parameter $\eta = \exp(-\xi)$.

In [19], the authors report that their prior distribution induces partitions with fewer large clusters for large values of ξ and vice-versa, small values of ξ induce partitions with many clusters of reduced sizes. In particular, Supplementary Figure 13 shows the prior distribution of the number of clusters for a grid of ξ values for the [19] prior setting. As we can observe, the distribution of the number of clusters is concentrated on lower values as the value of ξ increases.

The behaviour of our partition prior is summarized in Supplementary Figure 14 by reporting the prior distribution of the number of clusters. We can observe how the effect of ξ, as in the previous simulation, is to produce coarser partitions. However, this effect is incremented by the presence of the DP part in our prior specification, evident by comparing these results with Supplementary Figure 13, where the prior number of clusters is, given the same value of ξ, much higher.

3.3. Regime- switching aPPM - Time varying spatial clustering

As mentioned in Section 1, we are interested in detecting which areal units share similar temporal patterns. For this reason, the strategy we develop below builds the clustering structure on the array of coefficients $\boldsymbol{\beta} = [\boldsymbol{\beta}_{ir_t}]$, for $i \in [I]$, $t \in [T]$, and associated regime $r_t \in [n_R]$, in the spirit of Bayesian

nonparametric priors of discrete nature. Thanks to the introduction of the concept of model regimes, the clustering induced on the observations via $\boldsymbol{\beta}$ is able to reflect changes appearing through time, allowing for a regime-specific prior distribution for the partition of areas.

Firstly, we elaborate on the concept of regime and its relationship with the clustering of observations. A regime r_t at time $t \in [T]$ is associated to a partition of the areas indexed in $[I]$ indicated as $\rho_{r_t} = \{C_1^{r_t}, \ldots, C_{K_{r_t}}^{r_t}\}$. The number of clusters for each regime-specific partition is denoted by K_{r_t}. Because we allow for the same regime to exist at multiple time points over the time period under study, we must permit the same partition ρ_{r_t} to exist at these time points, effectively exploiting the division of the set $[T]$ of time points into $M \geq n_R$ non-overlapping sets. For instance, in our application, we distinguish between night and day regimes, as well as weekday/weekend, therefore considering $n_R = 4$ distinct regimes. However, we consider $M = 15$ non-overlapping intervals, since we have two weeks of observations, and there are two changes within each day to separate day by night (see Figure 3). Of course, this choice is suggested by the specific application considered here, but either simpler or more general choices could be considered. Changes in the regime status are identified by the time points $\bar{t}_0, \ldots, \bar{t}_M$, dividing the time set into non-overlapping intervals, such that:

$$[T] = \{\bar{t}_0, 2, \ldots, \bar{t}_1\} \cup \{\bar{t}_1 + 1, \ldots, \bar{t}_2\} \cup \cdots \cup \{\bar{t}_{M-1} + 1, \ldots, \bar{t}_M\} \quad (5)$$

where $\bar{t}_0 = 1$, $\bar{t}_M = T$ and $\bar{t}_1 < \bar{t}_2 < \cdots < \bar{t}_{M-1} < \bar{t}_M$. Depending on the application under study, one might possess prior information on when regimes change. However, in many situations such as the one presented here, the time points $\bar{t}_1, \ldots, \bar{t}_{M-1}$ are not deterministically known. Hence, we impose a prior distribution over a range of possible time points within each day of the week. Recalling that $\bar{t}_0 = 1$ and $\bar{t}_M = T$, the centre of the regime change intervals are denoted by λ_m, for $m = 1, \ldots M-1$. Specifically, we assume a discrete uniform prior distribution over $2n_\lambda + 1$ discrete time points. The regime change interval centres can be fixed according to the prior belief or estimated from the data themselves. For example, if λ_1 has been fixed or estimated to be 7:00am, we could set $n_\lambda = 4$ and assume $\bar{t}_1$, the first point of regime changes, to be uniformly distributed over the time points corresponding to 6am, 6:15am, 6:30am, 6:45am, 7am, 7:15am, 7:30am, 7:45am, and 8am, representing the prior belief of the city awakening period. Of course, we always assume (5) to be a partition of $[T]$. This is imposed by selecting an integer n_λ such that the support of each variable $\bar{t}_m$

is $\{\lambda_m - n_\lambda, \ldots, \lambda_m + n_\lambda\}$. Notice that, given the values $\bar{t}_m$, for $m = 0, \ldots, M$, the regime status r_t at each time point is known deterministically.

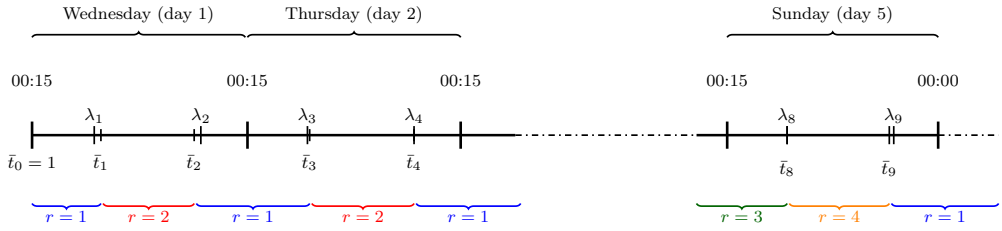


Figure 3: Illustration of the regime scheme in the case under study for the first 5 days. Throughout the time interval, the regimes are depicted in different colors, here indicating night/day separation. Notice how the regime interval is spanned by the points $\bar{t}_r$, which are not necessarily coinciding with the central time points λ_r. The total number of regimes is 4.

Finally, we assume that, conditionally on all the change-of-regimes points $\bar{\boldsymbol{t}} = (\bar{t}_0, \ldots, \bar{t}_M)$, the n_R regime-specific partitions $\rho_1, \ldots, \rho_{n_R}$ are independent, each distributed according to (4). Each regime-specific partition $\rho_r = \{C_1^r, \ldots, C_{K_r}^r\}$, for $r \in [n_R]$ can be equivalently represented through the introduction of a regime-specific vector of allocation variables $\boldsymbol{s}^r = (s_1^r, ..., s_I^r)$. Furthermore, a vector of unique values $\boldsymbol{\beta}_r^* = (\boldsymbol{\beta}_{1r}^*, \ldots, \boldsymbol{\beta}_{K_r r}^*)$ for each regime can be linked to $\tilde{\boldsymbol{\beta}}$, and in such a way to the observations, so that $s_i^r = j \iff \tilde{\boldsymbol{\beta}}_{it} = \boldsymbol{\beta}_{jr}^* \iff i \in C_j^r$ and $r = r_t$.

After introducing the allocation vector $\boldsymbol{s}^r = (s_1^r, ..., s_I^r)$, and the set of unique parameter values $\boldsymbol{\beta}_r^*$, we can describe the predictive distribution of $\boldsymbol{s}^r$. For notational convenience, we temporarily omit the index r indicating the regime. Consider the partition $\rho^i = \{C_1^i, \ldots, C_{K^i}^i\}$ obtained by clustering the first i elements into K^i clusters. The predictive law of a new element $i+1$ can be computed as follows:

$$P\left(s_{i+1} = j \mid s_1, \ldots, s_i\right) = \tag{6}$$
$$\begin{cases} \frac{p(\rho^{i+1}=\{C_1^i,\ldots,C_j^i \cup \{i+1\},\ldots,C_{K^i}^i\}|\kappa,\xi)}{p(\rho^i=\{C_1^i,\ldots,C_{K^i}^i\}|\kappa,\xi)} \propto n_j^i \mathrm{e}^{-\xi \ell^j(\{i+1\})}, & j = 1, \ldots, K^i \\ \frac{p(\rho^{i+1}=\{C_1^i,\ldots,C_{K^i}^i,\{i+1\}\}|\kappa,\xi)}{p(\rho^i=\{C_1^i,\ldots,C_{K^i}^i\}|\kappa,\xi)} \propto \kappa \mathrm{e}^{-\xi \ell^j(\{i+1\})}, & j = K^i + 1 \end{cases}$$

where $n_j^i = |C_j^i|$ is the j-th cluster size before we assign the $(i+1)$-th observation, and K^i is the number of clusters identified by $(s_1, \ldots, s_i)$ so that

$K^i + 1$ identifies the new cluster label. To recover the first line of formula (6) above, note that the ratio there is equal to

$$\frac{\mathcal{K}^r(\kappa,\xi,i+1)}{\mathcal{K}^r(\kappa,\xi,i)}\frac{\Gamma\left(n_j^i+1\right)}{\Gamma\left(n_j^i\right)}\frac{\mathrm{e}^{-\xi\sum_{m\in C_j^r\cup\{i+1\}}\ell^j(m)}}{\mathrm{e}^{-\xi\sum_{m\in C_j^r}\ell^j(m)}}$$
$$=\frac{\mathcal{K}^r(\kappa,\xi,i+1)}{\mathcal{K}^r(\kappa,\xi,i)}n_j^i\frac{\mathrm{e}^{-\xi\sum_{m\in C_j^r}\ell^j(m)-\xi l^j(\{i+1\})}}{\mathrm{e}^{-\xi\sum_{m\in C_j^r}\ell^j(m)}}$$

so that the ratio $\mathcal{K}^r(\kappa,\xi,i+1)/\mathcal{K}^r(\kappa,\xi,i)$ does not depend on j. Similar calculations hold for the second line in (6).

Finally, we impose the following prior distribution for the vector of unique values $\boldsymbol{\beta}_r^*$, for $r = 1,\ldots,n_R$:

$$\boldsymbol{\beta}_{1r}^*,\ldots,\boldsymbol{\beta}_{K_r r}^* \mid \rho_r,\boldsymbol{\mu}_{\boldsymbol{\beta}_r},\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r} \overset{\text{iid}}{\sim} \mathrm{N}_p(\boldsymbol{\mu}_{\boldsymbol{\beta}_r},\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r})$$
$$\boldsymbol{\mu}_{\boldsymbol{\beta}_r},\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r} \sim \mathrm{N}_p(\boldsymbol{\mu}_{\boldsymbol{\beta}_r} \mid \boldsymbol{m}_{\boldsymbol{\beta}_r},\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r})\prod_{j=1}^{p}\text{inv-Gamma}(\sigma^2_{\boldsymbol{\beta}_r j} \mid a_{\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r}},b_{\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r}})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r} = \mathrm{diag}\left(\sigma^2_{\boldsymbol{\beta}_r 1},\ldots,\sigma^2_{\boldsymbol{\beta}_r p}\right)$, $r = 1,\ldots,n_R$. Here $\text{inv-Gamma}(\cdot \mid a,b)$ denotes the inverse gamma density with mean $b/(a-1)$. We note that the unique values of the array $\boldsymbol{\beta}_r^*$, associated with the r-th regime, are shared by all those coefficients $\tilde{\boldsymbol{\beta}}_{it}$ for which regime r is active, hence for all $t \in \{\bar{t}_m+1,\ldots,\bar{t}_m\}$. We also assume independence among the $\boldsymbol{\beta}^*$s parameters across different regimes.

The final model, which we refer to as regime-switching areal PPM (RS-aPPM), can be described as follows:

$$Y_{1t},\ldots,Y_{It} \mid \boldsymbol{x}_t,\boldsymbol{\beta}_{1r}^*,\ldots,\boldsymbol{\beta}_{K_r r}^*,\rho_r=\{C_1^r,\ldots,C_{K_r}^r\},\boldsymbol{s}^r,\boldsymbol{u}_r,\sigma^2_{\epsilon_r},r=r_t$$
$$\overset{\text{ind}}{\sim}\prod_{j=1}^{K_r}\prod_{i\in C_j^r}\mathrm{N}(y_{it} \mid \boldsymbol{x}_t'\boldsymbol{\beta}^*_{s_i^r r}+u_{ir},\sigma^2_{\epsilon_r}),\ \text{for all } t: r_t = r \quad (7)$$

for $r = 1, \ldots, n_R$:

$$\boldsymbol{u}_r|\tau_r^2, Q(\zeta_r, W) \sim \mathrm{N}_I(\boldsymbol{u}_r|\boldsymbol{0}, \tau_r^2 Q(\zeta_r, W)^{-1}) \tag{8}$$

$$\boldsymbol{\beta}^*_{1r}, \ldots, \boldsymbol{\beta}^*_{K_r r}|\rho_r, \boldsymbol{\mu}_{\boldsymbol{\beta}_r}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_r} \overset{\text{iid}}{\sim} \mathrm{N}_p(\boldsymbol{\mu}_{\boldsymbol{\beta}_r}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_r}) \tag{9}$$

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_r}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_r} \overset{\text{ind}}{\sim} \mathrm{N}_p(\boldsymbol{\mu}_{\boldsymbol{\beta}_r}|\boldsymbol{m}_{\boldsymbol{\beta}_r}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_r}) \prod_{j=1}^{p} \text{inv-Gamma}(\sigma^2_{\boldsymbol{\beta}_r j}|a_{\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r}}, b_{\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r}}) \tag{10}$$

$$\text{with } \boldsymbol{\Sigma}_{\boldsymbol{\beta}_r} = \mathrm{diag}(\sigma^2_{\boldsymbol{\beta}_r 1}, \ldots, \sigma^2_{\boldsymbol{\beta}_r p}) \tag{11}$$

$$\rho_r \overset{\text{ind}}{\sim} p(\rho_r|\kappa, \xi, \bar{\boldsymbol{t}}) = \mathcal{K}^r(\kappa, \xi, \boldsymbol{n}^r)\kappa^{K_r} \prod_{j=1}^{K_r} \Gamma(n_j^r) \mathrm{e}^{-\xi \sum\limits_{i \in C_j} \ell^j(i)} \tag{12}$$

and prior for the regime changes:

$$\bar{t}_m \overset{\text{ind}}{\sim} Unif\{\lambda_m - n_\lambda, \ldots, \lambda_m + n_\lambda\}, \quad m = 1, \ldots, M-1 \tag{13}$$

with $\{1, \ldots, T\} = \{\bar{t}_0, 2, \ldots, \bar{t}_1\} \cup \{\bar{t}_1 + 1, \ldots, \bar{t}_2\} \cup \cdots \cup \{\bar{t}_{M-1} + 1, \ldots, \bar{t}_M\}$. We complete the prior specification with, for $r = 1, \ldots, n_R$:

$$\tau_r^2 \overset{\text{iid}}{\sim} \text{inv-Gamma}(a_{\tau_r^2}, b_{\tau_r^2}) \tag{14}$$

$$\sigma^2_{\epsilon_r} \overset{\text{iid}}{\sim} \text{inv-Gamma}(a_{\sigma^2_{\epsilon_r}}, b_{\sigma^2_{\epsilon_r}}) \tag{15}$$

Furthermore, prior independence is assumed among the parameters in the different equations above. Recall also that $Q(\zeta_r, W) = \zeta_r(\mathrm{diag}(W\boldsymbol{1}) - W) + (1 - \zeta_r)\mathbb{I}_I$, where W is the proximity matrix, as previously introduced in Section 3.1. We could assume ζ_rs random, e.g. beta-distributed. However, it is well-known (see, for instance, [1], Section 6.4.3.3, or [16]) that this leads to non-identifiability issues. For this reason, in the data application we have fixed $\zeta_r = 0.95$ for each regime r, encouraging spatial association, and have assumed informative marginal priors for τ_r^2 and $\sigma^2_{\epsilon_r}$. See Section 4 for the specific choice.

Let us denote by $\boldsymbol{\phi}$ the vector containing all the model parameters, that is $\boldsymbol{\phi} = \left(\boldsymbol{y}^{mis}, \boldsymbol{\beta}, \boldsymbol{s}, \bar{t}, \boldsymbol{u}, (\tau_1^2, \ldots, \tau_{n_R}), \left(\sigma^2_{\epsilon_1}, \ldots, \sigma^2_{\epsilon_{n_R}}\right), \boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right)$, where $\boldsymbol{s} = (\boldsymbol{s}^1, \ldots, \boldsymbol{s}^{n_R})$ and $\bar{t} = (\bar{t}_0, \ldots, \bar{t}_M)$. To obtain posterior samples from the full joint posterior distribution $\pi\left(\boldsymbol{\phi}|\boldsymbol{y}^{obs}\right)$, we implement a sequence of Metropolis-within-Gibbs steps. See Section 1 of Supplementary Material for its full description. Note that since there are missing values in the log-Erlang numbers, denoted by $\boldsymbol{y}^{mis}$, we incorporate them in the parameters to be simulated from the full conditionals, i.e. $\boldsymbol{\phi}$ contains $\boldsymbol{y}^{mis}$. See step 1. of the MCMC algorithm in Section 1 of Supplementary Material.

At first sight, model (7)-(15) might seem complicated and overly parameterised. However, it is in reality a *sparse* model since it considers only n_r random partition parameters $\rho_1, \ldots, \rho_{n_R}$, instead of one per each of the T time points. We will see in Section 1 of Supplementary Material that updating the n_R partitions is computationally expensive, but still reasonable. However, using $T = 1344$ random partitions would have been computationally prohibitive.

3.4. Summary of the simulation study

We perform extensive simulation studies investigating the effect of prior elicitation on the clustering estimation and posterior distributions of the parameters of interest. We provide in this section a summary of these simulations, while more details are reported in Section 2 of the Supplementary Material.

We simulate data from (7)-(8), that is, we simulate from $\boldsymbol{x}_t'\boldsymbol{\beta}^*_{s_i^r r} + u_{ir}$ plus independent and identically distributed errors, varying the regression parameter values and the error distribution. The aim of this simulation study is to evaluate the proposed model through a comprehensive set of simulation experiments designed to assess clustering performance under varying conditions, including model misspecification, missing data, and multiple regimes. For some of the examples, we also compare our cluster estimates with the ones from [19], which forgets the "rich gets richer" aspect of the cohesion function (4).

We illustrate some characteristics of our model via simulated data, with one ($n_R = 1$) or two ($n_R = 2$) regimes. In all scenarios, a grid of size 12×10 is used ($I = 120$), with varying underlying true regime-specific clustering structure, as well as the error distributions (Gaussian, t, or skew normal), allowing for misspecification. We also simulate the time-varying covariates, the *true* values of the coefficients $\boldsymbol{\beta}^*$ and of the random effects $\boldsymbol{u}$. Then, for each areal unit, a time series $\{y_{it}, t = 1, \ldots, T\}$ of length $T = 100$ is generated. In all the simulations, we run the MCMC algorithm described in Supplementary Section 1 for 15,000 iterations, of which the first 13,000 are then discarded as burn-in, and the last 2,000 are thinned to obtain a sample of 1,000 to be used in the posterior inference. In all the experiments, the cluster estimate for each regime is computed by minimizing the posterior expectation of the variation of information loss function [27]. We report the Adjusted Rand Index (ARI) between the true clustering structure and the

one estimated by our model for all simulated scenarios; see Supplementary Tables 1-2.

Across all scenarios, posterior chains for most parameters show stable behaviour, although the variance parameter τ^2 exhibits identifiability issues. Missing data affect clustering accuracy primarily when entire time series are unobserved. In total, we consider eight simulation settings, and their results are summarized next:

Examples 1-2 (single regime): assess recovery of spatial clusters under contamination and missing data. The model achieves near-perfect clustering (ARI approximately equal to 1), with slight degradation when data are missing.

Examples 3–4 (two regimes): evaluate changepoint detection and clustering under correct specification, absence of changepoints, and misspecified prior support. The model accurately recovers changepoints when supported by the prior, but performance deteriorates when changepoints are absent or outside the prior range.

Examples 5–6 (misspecification): investigate robustness to non-Gaussian errors (t and skew-normal). The model remains effective, though clustering accuracy decreases in heavy-tailed settings.

Examples 7–8: examine challenging scenarios with weak spatial structure and non-Gaussian noise; the model still successfully recovers clustering and changepoints.

Overall, results show that the proposed approach reliably identifies clustering structures and regime changes, demonstrating robustness to misspecification and missing data, while highlighting sensitivity to prior assumptions on changepoints.

Finally, Section 3.1 of the Supplementary Material reports cluster estimates for two of the simulated datasets, specifically the ones used in Examples 1 and 6, using some of the distance-based techniques discussed in Section 1. All three alternative methods tested resulted in cluster estimates that do not exactly match the simulation truth. In particular, two methods, MGWR and GWANN, fail to correctly recover the areal structure of the data. The third method, `ClustGeo`, provides slightly better estimates as it includes spatial information in the distance-based technique. However, these

methods inherit drawbacks typical of distance-based clustering algorithms: the number of clusters needs to be fixed in advance; the choice of dissimilarity matrices and of the linkage distance can strongly affect the results and computational efficiency of the method; absence of uncertainty quantification; and difficulty in handling missing data.

4. Application to Telecom data

In this section we fit our model to the Telecom data described in Section 2. Recall the log-Erlang values were previously standardised. We assume (7) - (8) with $\boldsymbol{x}_t$ given in (1) and we fix $n_R = 4$ regimes. Prior specification is as defined in (9) - (12) and (14) - (15). We found that the changepoint support centres have a strong influence on the overall inference. Therefore, we adopted an empirical Bayes approach to estimate these centres λ_ms for $m = 1, \ldots, M-1$, via the `R` package `ecp` [20]; see Section 5 in Supplementary Material for further details. The estimated values of λ_ms are in agreement with our prior belief on the changepoint location: for each day, there is typically a first changepoint in the morning due to commuting work/home hours and school or offices opening times, and a similar changepoint in the evening. However, our model allows for random changepoints to be estimated from the data. The diagonal elements of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_r}$ are a-priori inverse Gamma distributed with mean and variance equal to 1 and 0.01, respectively. The hyperparameters $a_{\tau_r^2}$, $b_{\tau_r^2}$, $a_{\sigma^2_{\epsilon_r}}$, $b_{\sigma^2_{\epsilon_r}}$ in (14) and (15) are fixed in order to get informative priors with means and variances equal to 1 and 0.1, respectively. This was done to mitigate possible unidentifiability issues in the posterior estimates of $\tau^2, \sigma^2_\epsilon$ (see, for instance, [1], Section 6.4.3.3, or [16]). The hyperparameters in the marginal prior for ρ_r in (12) are ξ and κ. From the analytic expression of this prior, and the prior simulation study, we know that the number of clusters increases with κ, while it decreases with ξ. Given the amount of data allocated to each regime, we found that the corresponding likelihood terms have predominant weight in the marginal posterior distributions of ρ_r. Therefore, in an attempt to induce parsimony in the resulting clustering, we conduct the analysis by fixing $\xi = 2$ and $\kappa = 1$, inducing a strongly informative prior on K_r as seen in Supplementary Figure 14. See the end of this section where we discuss sensitivity with respect to ξ and κ and how to fix them.

We run the MCMC for model (7)-(15) with fixed change-of-regimes times for a total of 50,000 iterations, with a final sample size of 2,500 iterations,

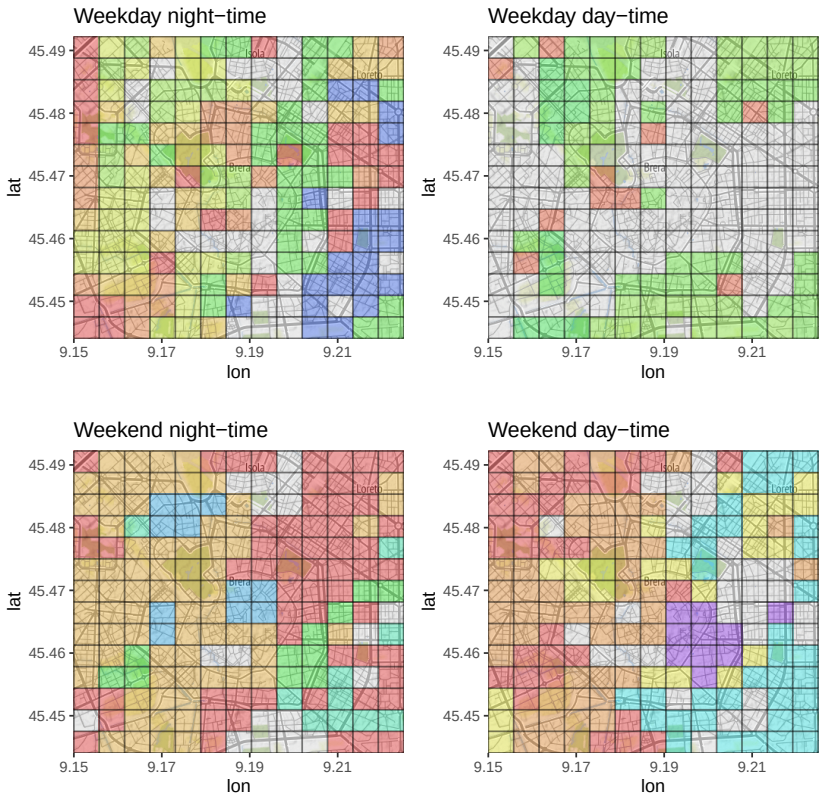


Figure 4: Telecom data. Posterior estimate of the random partition ρ_r, for $r = 1, \ldots, n_R = 4$, given by minimizing the VI's loss function. Areal units in clusters of size smaller than ten are left uncoloured. Each colour corresponds to a cluster, with the colour scale not reflecting the intensity of a parameter. A map of the total area is superimposed.

after a burn-in of 45,000 and thinning of 2. For each regime, we report a point estimate of the random partition for areal units, minimizing the posterior expectation of the variation of information (VI) loss function with equal misclassification cost parameters. Since the cardinality of the visited partitions is quite large, we employ a hierarchical clustering algorithm with distance equal to the complement of the posterior co-clustering probability and average linkage [39]. Calculations are performed via the R package `salso` [12]; see also `https://CRAN.R-project.org/package=salso`.

We show the estimated regime-specific partitions in Figure 4, where each cluster is identified by a different colour. For visualization purposes, areal units in clusters of size smaller than ten are left uncoloured. We can observe

how clusters of large sizes identify regions of the metropolitan area of Milan corresponding to the centre, external rings, and specific hotspots on the map. The grouping of the areal units changes between regimes, reflecting the different types of trajectories observed. Supplementary Figure 15 displays the same plot as in Figure 4, without the map of the whole area underneath the coloured clustering assignments.

Figure 4 and Supplementary Figure 15 show some interesting spatio-temporal characteristics, involving major public/travelling spots in the city of Milan. In the regime corresponding to weekend day-time, big clusters are made by contiguous areal units overlapping the more external rings: for instance, big clusters are made by contiguous areal units overlapping the more external rings: for instance, we see a (blue) cluster around Porta Romana and Piazzale Lodi, a portion of the city with many restaurants and bars. There is a large cluster around the area of Porta Venezia, including Piazzale Loreto, which is a shopping area. The west part of the city is split into two clusters, the largest of them overlapping the more external rings. For weekend night-time, the city seems to be divided into two larger clusters (the red cluster in the west part and the orange cluster in the east part). During the weekday day-time, it is clear that there is a large part of the city that is split into small clusters (the gray area), but there are two (light and darker green) clusters that split the external rings. Finally, in Figure 4 at weekday night-time, the cluster estimate we get is less homogeneous. For instance, there is a blue cluster at the bottom right corner, overlapping the external ring (viale Isonzo, piazzale Lodi, viale Umbria), an orange cluster at the top right including Piazzale Loreto. Note, however, that the total number of estimated clusters is 16 for weekday night-time, 9 for weekend night-time, 12 for weekend day-time and 31 for weekday day-time. For this last regime, comparison to Figure 4 shows that there are many clusters of size smaller than ten (left uncoloured).

In order to understand the differences between the random partitions in the regimes, Table 1 reports the posterior summary statistics of the Adjusted Rand Index (ARI) between ρ_i and ρ_j, $i \neq j$, a similarity measure for partitions, where values closer to 1 indicate partitions that are more similar. It is clear from the table that the cluster estimates are very different across regimes: in particular, weekday day-time seems the one less similar to all the others. Table 1 indicates that incorporating a regime-specific random partition appears to be the most suitable modelling option.

We display in Figure 5 the posterior mean of $\boldsymbol{x}_l'\boldsymbol{\beta}_{ir_t}$ as a function of

Regimes	1 Vs 2	1 Vs 3	1 Vs 4	2 Vs 3	2 Vs 4	3 Vs 4
Quantile 2.5%	0.100	0.162	0.134	0.026	0.063	0.363
Median	0.104	0.169	0.137	0.029	0.064	0.369
Quantile 97.5%	0.108	0.176	0.140	0.030	0.067	0.375

Table 1: Telecom data. Quantiles (2.5%, 50% and 97.5%) of the posterior distribution of the Adjusted Rand Index between ρ_i and ρ_j, $i \neq j$. Here $i = 1$ is weekday night-time, $i = 2$ weekday day-time, $i = 3$ weekend night-time and $i = 4$ weekend day-time.

time t, for three different areas corresponding to Stazione Centrale (Central Station), Piazza Duomo and Piazzale Piola, previously shown in Figure 1(a). These plots suggest that each of these locations exhibit high or low activity, depending on specific periods of the day/week. Duomo is very active on weekdays but less so during weekends, while Piola oscillates between high activity on weekends or low activity late at night.

Section 3.2 of Supplementary Material shows comparison with the distance-based methods discussed in Section 1 for the Telecom data. The Telecom data presents several missing observations which cannot be handled by the alternative methods. Therefore, we impute the missing values via linear interpolation using the R package `longitudinalData`; see Supplementary Section 3.2 for details on the procedure. However, to overcome issues relative to the amount of memory required to implement these methods, 100 time points were selected uniformly at random from the imputed dataset for further analysis. For one of the methods, GWANN, we only selected ten time points due to memory issues arising with a bigger subset of the data. For one of the methods, `ClustGeo`, which includes the spatial information in the distance-based technique, the estimated clusters show some of the structures characterizing the geography of the metropolitan area of Milan; see Supplementary Figure 11(c). The other two methods are able to weakly identify some interpretable features of the partition of areal units, namely the concentric regions of the map corresponding to the circular structure of Milan's topology; see Supplementary Figure 12. As in the case of simulated data, beyond memory issues when using the associated package, these methods need the number of clusters to be fixed in advance and have no uncertainty quantification.

Section 4 of Supplementary Material shows a performance comparison between the proposed model and some of its variations as well as Bayesian

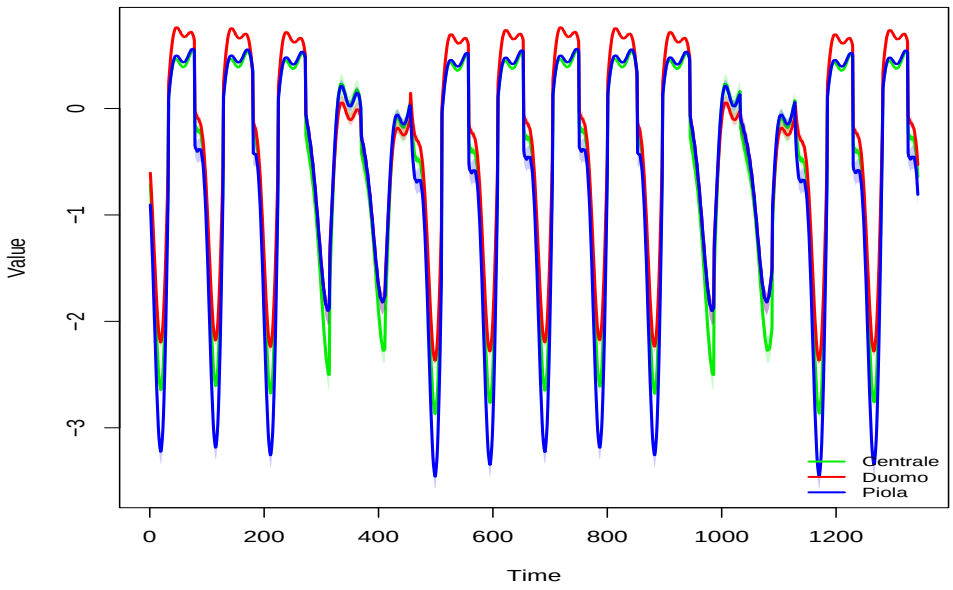


Figure 5: Telecom data. Posterior mean and 95% credible interval for $\boldsymbol{x}_t'\boldsymbol{\beta}_{ir_t}$, as a function of t, when i denotes the areal unit including Stazione Centrale, Piazza Duomo and Piazzale Piola.

competitors available in the literature. Beyond the full model as described at the beginning of this section, the competing models include the RS-aPPM model equipped with either a Dirichlet Process prior for the partition (full model with $\xi = 0$) or the same prior used in [19] ($\kappa = 1$ and dropping the $\Gamma(n_j^r)$ terms in (12)), and the spatio-temporal conditionally auto-regressive models of [23] which are available for implementation through the R package `CARBayesST`. In particular, for the latter we focus on the `ST.CARar` model with $\rho = 0, 0.95$. All the models are evaluated on the same Telecom dataset. See Supplementary Table 3 where we report the posterior mode of the number of clusters within each regime, the LPML and the WAIC for each scenario. In terms of these two indicators, the best model is ST.CARar with $\rho = 0.95$. It is interesting to observe that the ST.CARar model detects the presence of spatial correlation, as reflected by comparing the cases $\rho = 0$ and $\rho = 0.95$. The ST.CARar model includes a more sophisticated and saturated specification of spatio-temporal random effects than our CAR structure, which may

explain the LPML and WAIC values here reported. In terms of goodness of fit, the ranking of the three specifications of the RS-aPPM, after ST.CARar ($\rho = 0.95$), is DP version as first, then HB version, followed by the full version (our proposed model) with the lower number of estimated clusters in all regimes, confirming the well-known trade-off between clustering and density estimation. Note that all the models from the R package `CARBayesST`, included ST.CARar, do not allow for any clustering structure estimation, neither do they provide modelling of regimes.

Recall that, a priori, κ and ξ control the clustering behaviour of our model. This can be seen from the full conditional allocation probabilities of the latent variables $\mathbf{s}^r$ ($r = 1, \ldots, n_R$), reported in Supplementary Section 1, which are the only quantities whose full conditional involves κ and ξ. In particular, the probability of creating a new cluster when allocating area i during regime r is proportional to $\kappa e^{-\xi \ell_j(\{i\})}$, where $\ell_j(\{i\})$ is the boundary length of the i-th areal unit belonging to the cluster C_j, that is the number of neighbours of i that do not belong to C_j, for $j = 1, \ldots, K_r$. Under our neighbourhood definition, which includes the eight surrounding areas, this probability is bounded below by $\kappa e^{-8\xi}$. Fixing $\xi = 2$ therefore keeps the prior probability of creating new clusters small, discouraging excessive fragmentation and preventing the likelihood from dominating the posterior, given the large sample size (about $60,000$ observations per regime). For existing clusters, the allocation probability is proportional to $n_j^{-i} e^{-\xi \ell_j(\{i\})}$, which balances cluster size and spatial cohesion. Our extensive simulations suggest that fixing κ at the conventional value of 1, as often done by default in the BNP literature, the user can obtain better control by moving ξ over its range. We have opted for a conservative prior that favors a smaller number of clusters (as achieved with a relatively large choice of $\xi = 2$); see Figure 14 in the SM. Similar results have been obtained with a modest change of these values ($\xi = 1$, results not shown). Partitions are similar, though $\xi = 1$ gives a larger number of estimated clusters. In the application, with such a large amount of data (approx $60,000$ observations per regime) we had to choose hyperparameters corresponding to a prior concentrated on very small values for each K_r. In other applications, if data size were smaller, we could allow a larger prior variability for K_r. Summing up, we suggest that the procedure to set κ of the same order of magnitude as 1 and fix ξ based on the size of the problem and the distribution a-priori of the induced number of clusters.

5. Summary and Conclusions

Motivated by the analysis of mobile phone usage in part of the city of Milan, Italy, we propose a semi-parametric random partition model for the analysis of large spatio-temporal data. The model features a random partition prior distribution that combines the well-known DP with the HB specifications, providing a balance among spatially cohesive areal grouping and number and sizes of clusters. Due to the nature of the data, consisting of series of measurements every 15 minutes over the entire span of two weeks, the model also incorporates the notion of switching regimes, to reflect differences over weekday/night and weekend patterns of mobile phone usage.

Through extensive simulation studies we find that the model fares well when compared to other alternative Bayesian models, including the random partition models corresponding to either the DP or the HB priors (i.e., not combined) and the models included in `CARBayesST` package available in `R`. We have also compared our model with some non-Bayesian statistical and machine learning approaches which could accommodate or be extended to be used in tandem with spatio-temporal clustering procedures. However, none of these methods/packages are able to handle missing data, different temporal regimes, and cluster estimates simultaneously. Moreover, we encountered severe memory issues and we had to drastically reduce the time points (only 100 or even 10 in one case) for the Telecom application. The overall conclusion is that our approach, being tailored to the particular application at hand, is an all-inclusive approach. Specifically, it is able to account for missing responses, varying regimes corresponding to different times, time evolving responses recorded throughout a number of days, cluster estimates of the areal units, and last but not least, it is capable to quantify uncertainty of the estimates.

We would like to point out that an alternative to doing inference based on MCMC simulation is to use a Integrated Nested Laplace Approximations (INLA, [34]). INLA has become increasingly popular in the analysis of spatio-temporal data, also thanks to the R-INLA package. However, while R-INLA can fit many different spatio-temporal models, it does not provide a way to implement a model equipped with a random partition.

A limitation of our model, in fact shared with many related random partition models is the computational burden required to implement posterior simulation via MCMC. However, it has to be noted that the main bottleneck of the code is the number of timestamps, which largely impacts the compu-

tational time. We were also able to run our algorithm for about 3 times the number of areas without any dramatic change in the computational time. For example, with 182 areas per 1344 time points as in Section 4, we ran the code on a university server and we recorded an approximate number of 34 iterations/minute. On the other hand, we ran all the simulated examples, consisting of 120 observation per 100 time points, on our personal laptops, measuring an approximate 2419 iterations/minute for the two-regimes case. A possible topic for future research is the adoption of divide and conquer techniques that split the data in smaller segments that can be dealt with in parallel, to be later suitably combined to produce a reasonable approximation to the actual posterior distribution. This approach would yield a faster and more computationally feasible strategy that can be applied to larger datasets, as well as used to fit models featuring a higher number of parameters with the aim of capturing area-specific effects.

6. Acknowledgments

Andrea Cremaschi acknowledges the Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore and the Singapore Institute for Clinical Sciences (SICS, A*STAR) for the support. This research was (partially) completed while three authors (Cremaschi, Guglielmi, Quintana) were visiting the Institute for Mathematical Sciences, National University of Singapore, 2024.

7. Funding

Andrea Cremaschi acknowledges partial support by PID2024-155187OB-I00 granted by MCIU, and by RYC2024-050330-I, funded by MICIU/AEI/10.13039/501100011033 and FSE+. Fernando Quintana acknowledges partial support by the grant ANID Fondecyt Regular 1220017. Alessandra Guglielmi has been partially supported by MUR, PRIN project 2022CLTYP4. Giulio Beltramin and Alessandra Guglielmi acknowledge the support by MUR, Grant Dipartimento di Eccellenza 2023–2027.

References

[1] S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2014.

[2] Daniel Barry and John A Hartigan. Product partition models for change point problems. *The annals of Statistics*, pages 260–279, 1992.

[3] L. Mark Berliner, Christopher K. Wikle, and Noel Cressie. Long-lead prediction of pacific ssts via bayesian dynamic modeling. *Journal of Climate*, 13(22):3953–3968, 2000. ISSN 08948755, 15200442. URL http://www.jstor.org/stable/26247700.

[4] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.

[5] Julian Besag. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 24(3):pp. 179–195, 1975. ISSN 00390526. URL http://www.jstor.org/stable/2987782.

[6] Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, Mar 1991. ISSN 1572-9052. URL https://doi.org/10.1007/BF00116466.

[7] Derya Birant and Alp Kut. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & knowledge engineering*, 60(1):208–221, 2007.

[8] Annalisa Cadonna, Andrea Cremaschi, and Alessandra Guglielmi. Bayesian modeling for large spatio-temporal data: an application to mobile networks. In *SIS 2019-Smart Statistics for Smart Applications*, pages 691–696. Pearson, 2019.

[9] Marie Chavent, Vanessa Kuentz-Simonet, Amaury Labenne, and Jérôme Saracco. Clustgeo: an R package for hierarchical clustering with spatial constraints. *Computational Statistics*, 33(4):1799–1822, 2018.

[10] N Cressie. *Statistics for Spatial Data*. Wiley, 1993.

[11] Noel Cressie and Christopher K Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.

[12] David B Dahl, Devin J Johnson, and Peter Müller. Search algorithms and loss functions for Bayesian clustering. *Journal of Computational and Graphical Statistics*, 31(4):1189–1201, 2022.

[13] Pierre Deville, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R Stevens, Andrea E Gaughan, Vincent D Blondel, and Andrew J Tatem. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45):15888–15893, 2014.

[14] Olga Dorabiala, Jennifer Webster, Nathan Kutz, and Aleksandr Aravkin. Spatiotemporal k-means. *arXiv preprint arXiv:2211.05337*, 2022.

[15] Ghislain Geniaux and Davide Martinetti. A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Regional Science and Urban Economics*, 2017.

[16] T Goicoa, A Adin, MD Ugarte, and JS Hodges. In spatio-temporal disease mapping models, identifiability constraints affect pql and inla results. *Stochastic Environmental Research and Risk Assessment*, 32: 749–770, 2018.

[17] Julian Hagenauer and Marco Helbich. A geographically weighted artificial neural network. *International Journal of Geographical Information Science*, 36(2):215–235, 2022.

[18] Matthew J Heaton, William F Christensen, and Maria A Terres. Nonstationary gaussian process models using spatial hierarchical clustering from finite differences. *Technometrics*, 59(1):93–101, 2017.

[19] Avril Hegarty and Daniel Barry. Bayesian disease mapping using product partition models. *Statistics in medicine*, 27(19):3868–3893, 2008.

[20] Nicholas A James and David S Matteson. ecp: An r package for nonparametric multiple change point analysis of multivariate data. *Journal of Statistical Software*, 62:1–25, 2015.

[21] Mark S Kaiser and Noel Cressie. The construction of multivariate distributions from markov random fields. *Journal of Multivariate Analysis*, 73(2):199–220, 2000.

[22] Leonhard Knorr-Held and Håvard Rue. On block updating in markov random field models for disease mapping. *Scandinavian Journal of Statistics*, 29(4):597–614, 2002. ISSN 03036898, 14679469. URL `http://www.jstor.org/stable/4616737`.

[23] Duncan Lee, Alastair Rushworth, and Gary Napier. Carbayesst: An r package for spatio-temporal areal unit modelling with conditional autoregressive priors. *R Package Version*, 2, 2015.

[24] Brian G. Leroux, Xingye Lei, and Norman Breslow. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In M. Elizabeth Halloran and Donald Berry, editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191, New York, NY, 2000. Springer New York. ISBN 978-1-4612-1284-3.

[25] Yu Liu, Xi Liu, Song Gao, Li Gong, Chaogui Kang, Ye Zhi, Guanghua Chi, and Li Shi. Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3):512–530, 2015.

[26] Fabio Manfredini, Paola Pucci, Piercesare Secchi, Paolo Tagliolato, Simone Vantini, and Valeria Vitelli. Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the milan urban region. In *Advances in complex data modeling and computational methods in statistics*, pages 133–147. Springer, 2015.

[27] Marina Meilă. Comparing clusterings by the variation of information. In *Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003. Proceedings*, pages 173–187. Springer, 2003.

[28] Peter Müller, Fernando Quintana, and Gary L Rosner. A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics*, 20(1):260–278, 2011.

[29] Garritt L Page and Fernando A Quintana. Spatial product partition models. *Bayesian Analysis*, 11(1):265–298, 2016.

[30] Jessica Pavani and Fernando Andrés Quintana. A Bayesian Multivariate Model With Temporal Dependence on Random Partition of Areal Data for Mosquito-Borne Diseases. *Statistics in Medicine*, 44(3-4):e10325, 2025. doi: https://doi.org/10.1002/sim.10325.

[31] Fernando A Quintana and Pilar L Iglesias. Bayesian clustering and product partition models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):557–574, 2003.

[32] Andrea Riebler, Sigrunn H Sørbye, Daniel Simpson, and Håvard Rue. An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165, 2016. doi: 10.1177/0962280216660421. URL `https://doi.org/10.1177/0962280216660421`. PMID: 27566770.

[33] Havard Rue and Leonhard Held. *Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC, 2005. ISBN 1584884320.

[34] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society Series B*, 71: 319–392, 04 2009.

[35] Piercesare Secchi, Simone Vantini, and Valeria Vitelli. Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of Milan. *Statistical Methods & Applications*, 24(2):279–300, jul 2015. ISSN 1618-2510. URL `http://link.springer.com/10.1007/s10260-014-0294-3`.

[36] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968): 1018–1021, 2010.

[37] Patrizia Sulis, Ed Manley, Chen Zhong, and Michael Batty. Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science*, 16:137–162, 2018.

[38] Wei Tu, Tingting Zhu, Jizhe Xia, Yulun Zhou, Yani Lai, Jincheng Jiang, and Qingquan Li. Portraying the spatial dynamics of urban vibrancy using multisource urban big data. *Computers, Environment and Urban Systems*, 80:101428, 2020.

[39] S. Wade and Z. Ghahramani. Bayesian Cluster Analysis: Point Estimation and Credible Balls (with Discussion). *Bayesian Anal.*, 13(2):559 – 626, 2018.

[40] Zhensheng Wang, Yang Yue, Biao He, Ke Nie, Wei Tu, Qingyun Du, and Qingquan Li. A bayesian spatio-temporal model to analyzing the stability of patterns of population distribution in an urban space using mobile phone data. *International Journal of Geographical Information Science*, 35(1):116–134, 2021.