SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Bayesian Statistics Project Final Report

CLUSTERING WEEKLY DATA OF ONE YEAR OF PM2.5 DATA

Laurea Magistrale in Mathematical Engineering

Authors: Borrini Elisa, Carbonara Filippo, Cefaloni Benedetta, Ettel Dina Sophie, Grignani Alessandro, Wolf Florian

Advisors: Michela Frigeri, Alessandro Carminati

Academic year: 2023-2024

Contents

1. Introduction

Air Quality Challenges in Lombardy, Italy: Air pollution, a critical environmental concern, poses significant risks to human health and the ecosystem. The Lombardy region in Italy faces significant air pollution challenges, ranking among the most polluted areas in Europe. This issue arises from factors such as limited air circulation and high emission levels.
Among the various pollutants, particulate matter with a diameter of 2.5 micrometers or smaller (PM2.5) has emerged as a key focus due to its potential for adverse health effects. PM2.5 consists of tiny particles suspended in the air, originating from diverse sources such as vehicle emissions, industrial activities, and natural processes.

Understanding the temporal patterns of PM2.5 levels is crucial for identifying trends, potential sources, and developing effective pollution control strategies. Clustering techniques will be employed to categorize weeks with similar PM2.5 concentration profiles, providing insights into the underlying patterns and contributing factors. In order to do this, our project analyzes a dataset spanning the years 2016 to 2021 [Rod+23], collecting daily values of air quality, weather conditions, emissions, livestock, and land and soil use. Pollutant data are sourced from the European Environmental Agency and the Lombardy Regional Environment Protection Agency, Weather and emissions data are obtained from the European Copernicus program, livestock data from the Italian zootechnical registry, and land and soil use data from the CORINE Land Cover project. The project focuses on analyzing and clustering weekly data of PM2.5 concentrations over the course of one year (2019), trying to assess the impact of agriculture on air quality in the selected area through statistical techniques and highlighting the relationship between the lifestock sector and the air pollutant concentrations.

2. Data and Covariates

The Agrimonia dataset integrates satellite data, model output, and in-situ measurements sourced from national and international agencies, each with varying spatial and temporal resolutions.

Source Data Overview: The dataset encompasses five key dimensions: air quality (AQ), weather and climate (WE), pollutant emissions (EM), livestock (LI), and land and soil characteristics (LA). Given the applicability of geostatistical methods in leveraging neighboring territory information for enhanced predictive capability near borders, a 0.3° buffer is applied around the Lombardy region, intersecting with several adjacent regions (fig. 1).

Causes and sources related to the emissions: Particulate matter with a diameter of 2.5 micrometers or smaller originates from various anthropogenic and natural sources. Among the main causes related to the release of significant amounts of PM2.5 into the atmosphere, we can mention intensive livestock farming, as well as combustion processes, including those from vehicles and industrial activities. Moreover, analyzing the provided dataset, it has emerged that one crucial variable influencing PM2.5 concentrations is the Boundary Layer Height (BLH) Max which represents the maximum depth of air next to the Earth's surface that is most affected by the resistance to the transfer of momentum, heat, or moisture across the surface.

Main Problems Associated with PM2.5: In order to understand the relevance of the

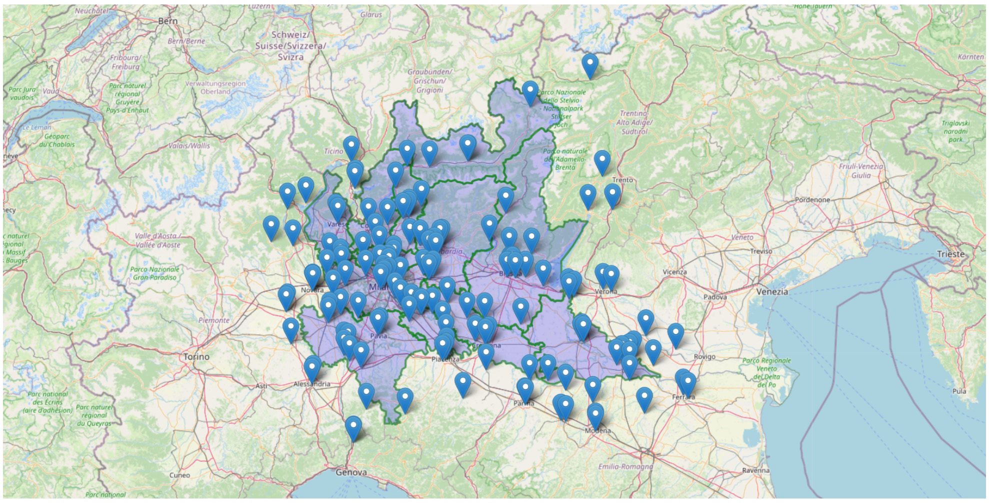


Figure 1: Buffered Area around Lombardy Region

analysis developed in this project, it is important to focus on the problems and the risks associated with high concentrations of PM2.5

- **Long Residence Time in the Atmosphere:** PM2.5 particles have an extended residence time in the atmosphere, leading to widespread dispersion and potential long-range transport. This characteristic contributes to the global distribution of PM2.5 and its diverse environmental impacts.
- **Health Impact:** Due to their small size, PM2.5 particles can penetrate deep into the human respiratory system, reaching the lungs and even entering the bloodstream. Prolonged exposure to elevated levels of PM2.5 is associated with various respiratory and cardiovascular diseases, posing a significant public health concern.
- **World Health Organization Recommendations:** The World Health Organization (WHO) recommends an **annual average** of $\leq 5\mu\frac{\text{g}}{\text{m}^3}$ for PM2.5 concentrations to safeguard public health. Exceeding these levels may lead to increased health risks, making it imperative to monitor and control PM2.5 pollution.

The outcomes of this analysis will not only enhance our understanding of PM2.5 variability but also assist policymakers and environmental scientists in formulating targeted interventions to mitigate the impact of air pollution on public health and the environment.

3. Models

Regarding the choice of models, we first focus on three basic modeling approaches: spatial-informed partitioning of the data, covariate-informed partitioning and modeling temporal dependence in partitions. Each of the modeling methods is a hierarchical model with Gaussian likelihood, a Gaussian prior for cluster-specific means and a uniform prior for cluster-specific variances. All of them allow for a number of specifications, e.g. different setting ups of priors. Later, several extensions and combinations are considered.

In the whole section we denote by n the number of measurement units, by $\rho = \{S_1, \dots, S_k\}$ a partition of the n measurement units and by c_i the index of the cluster that measurement unit i belongs to, i.e. $c_i = j$ if $i \in S_j$. Furthermore, cluster-specific values are marked with $*$. For example we consider cluster specific means $\boldsymbol{\mu}^* = \{\mu_1, \dots, \mu_k\}$ and standard deviations $\boldsymbol{\sigma}^*$.

3.1. sPPM Model: Spatial informed Clustering using location-dependent Similarity Functions

The following model is taken from [PQ16]. It is implemented as part of the R-package `ppmSuite` [Pag+23]. The overall model structure is the following

$$\begin{aligned} Y_i|\boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*}, c_i &\overset{\text{ind}}{\sim} \mathcal{N}(\mu^*_{c_i}, \sigma^{2*}_{c_i}), i = 1, \dots, n \\ (\mu^*_j, \sigma^*_j)|\mu_0, \sigma_0^2 &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_0, \sigma_0^2) \times \text{UN}(0, A) \\ (\mu_0, \sigma_0) &\sim \mathcal{N}(m, s^2) \times \text{UN}(0, B) \\ \rho &\sim \text{sPPM}(M, \boldsymbol{\theta}), \end{aligned}$$

where $m \in \mathbb{R}, s^2 \in [0, \infty)$, the bounds $A, B \in [0, \infty)$ as well as the concentration parameter $M \in (0, \infty)$ and θ that will be explained in more detail below are user-defined parameters.

The sPPM is a prior of the following form, where $\mathbf{s}$ denotes the spatial coordinates of the measurement units:

$$\mathbb{P}(\rho|\boldsymbol{s}) \propto \prod_{j=1}^{k_\rho} \left(\underbrace{M \cdot \Gamma(|S_j|)}_{=:c(S_j)} g(S_j, \boldsymbol{s}^*_j|\boldsymbol{\theta}) \right). \tag{1}$$

The so-called similarity function g is a non-negative function that measures the togetherness of the stations in the set S_j. Note that $\prod_{j=1}^{k_\rho} c(S_j)$ is proportional to the distribution of ρ in a random partition model induced by a sample from a Dirichlet process as in [Gug23, Section 8.1.3]. Therefore M takes the role of the concentration parameter in the Dirichlet process and influences the number of clusters.

For the similarity function g that incorporates the spatial information there are four options available:

1. $\theta = \alpha \in (0, \infty)$ with the distance measure

$$g_1(S_j, \boldsymbol{s}^*_j|\theta) := \begin{cases} \frac{1}{\Gamma(\alpha \mathcal{D}_h) \mathbf{1}[\mathcal{D}_h \geq 1] + \mathcal{D}_h \mathbf{1}[\mathcal{D}_h < 1]}, & \text{if } |S_h| > 1 \\ M, & \text{if } |S_h| = 1 \end{cases},$$

   with a distance function $\mathcal{D}_h := \sum_{i \in S_h} d(\boldsymbol{s}_i, \bar{\boldsymbol{s}}_h)$ and the cluster centroid $\bar{s}_{hk} = \frac{1}{n_h} \sum_{i \in S_h} s_{ik}$ for coordinates $k = 1, 2$ and $n_h := |S_h|$. Larger α favors denser clusters.
2. $\theta = a \in (0, \infty)$ with distance measure

$$g_2(S_h, \boldsymbol{s}^*_h|\theta) := \prod_{i,j \in S_h} \mathbf{1}\left[\|s_i - s_j\| \leq a\right].$$

   Larger a allows for larger neighborhoods.
3. *(Auxiliary Cohesion)* With dimension $d = 2$ and $\boldsymbol{\xi} = (\boldsymbol{m}, \boldsymbol{V}) \in \mathbb{R}^d \times \mathbb{S}^d_+ = \mathbb{R}^d \times \{X \in \mathbb{R}^{d \times d} | X \succeq 0\} =: \Xi$, we have (prior predictive conjugate model)

$$g_3(S_h, \boldsymbol{s}^*_h|\boldsymbol{\theta}) := \int_\Xi \prod_{i \in S_h} q(\boldsymbol{s}_i|\xi_h) q(\xi_h) \mathrm{d}\xi_h,$$

   with $q(\boldsymbol{s}|\boldsymbol{\xi}) = \mathcal{N}(\boldsymbol{s}|\boldsymbol{\xi})$ and $q(\boldsymbol{\xi}) = \text{NIW}(\boldsymbol{m}, \boldsymbol{V}|\boldsymbol{\mu_0}, \kappa_0, \nu_0, \Delta_0 \cdot \text{Id}_d)$ for user-defined parameters $\boldsymbol{\theta} = (\boldsymbol{\mu_0}, \kappa_0, \Delta_0, \nu_0) \in \mathbb{R}^d \times (0, \infty)^2 \times (1, \infty)$ (last part since $\nu_0 > d - 1$ has to be fulfilled).
4. *(Double Dipper)* With same structure as g_3, but with a posterior predictive conjugate model

$$g_4(S_h, \boldsymbol{s}^*_h|\boldsymbol{\theta}) := \int_\Xi \prod_{i \in S_h} q(\boldsymbol{s}_i|\xi_h) q(\xi_h|\boldsymbol{s}^*_h) \mathrm{d}\xi_h$$

   and conjugate model $q(\boldsymbol{s}_i|\xi_h) = \mathcal{N}(\boldsymbol{s}_i|\boldsymbol{m}_h, \boldsymbol{V}_h)$ and $q(\xi_h|\boldsymbol{s}^*_h) = \text{NIW}(\boldsymbol{m}_h, \boldsymbol{V}_h|\boldsymbol{s}^*_h)$ Compared to g_3 this option is more peaked and puts more weights on local partitions.

3.2. PPMx: Clustering using Covariate-dependent Similarity functions and Prior on Cluster Size

The covariate-informed partition model is taken from [PQ17] and all the introduced variants are implemented in the R-package `ppmSuite` [Pag+23]. The overall structure is the same as for the spatial-informed clustering with the difference that the similarity function now measures the homogeineity of the covariate values in a given partition set. Again, the values $m \in \mathbb{R}, s^2 \in [0, \infty)$ and bounds $A, B \in [0, \infty)$ as well as concentration parameter $M \in (0, \infty)$ and parameter(s) θ for the chosen similarity function are user-defined.

$$Y_i|\boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*}, c_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu^*_{c_i}, \sigma^{2*}_{c_i}), i = 1, \ldots, n$$

The prior for the clusters is

$$\mathbb{P}(\rho|\boldsymbol{x}) \propto \prod_{j=1}^{k_\rho} c(|S_j|)g(\boldsymbol{x}^*_j|\boldsymbol{\theta}).$$

For the so-called cohesion function c either the same function as in Section 3.1 (that is proportional to the partition probabilities in a random partition model derived from a Dirichlet process), or a uniform cohesion $c \equiv 1$ can be chosen.

In the following discussion $p = 1$ is assumed if not stated otherwise. If p dimensional covariate vectors are available,

$$\tilde{g}(\boldsymbol{x}^*_j|\boldsymbol{\theta}) = \prod_{l=1}^{p} g(\boldsymbol{x}^*_{jl}|\boldsymbol{\theta})$$

is adopted.

The PPMx has been successfully employed in a variety of settings when a relatively small number of covariates are available. However, as a large number of covariates can lead to either a large number of singleton clusters or one single large cluster, we consider two methods that cap the covariates' influence on the partitioning.

$$(1)\ \tilde{g}(\boldsymbol{x}^*_j) = \frac{g(\boldsymbol{x}^*_j)}{\sum_{i=1}^{k_j} g(\boldsymbol{x}^*_i)} \quad \text{or} \quad (2)\ \tilde{g}(\boldsymbol{x}^*_j) = g(\boldsymbol{x}^*_j)^{\frac{1}{p}}$$

For similarity functions g there are four options available:

1. With $\theta = \alpha \in (0, \infty)$
$$g_1(\boldsymbol{x}^*_j|\theta) := \exp\{-\alpha H(\boldsymbol{x}^*_j)\}.$$
For continuous covariates $H(\boldsymbol{x}^*_j) = \frac{1}{n}\sum_{l \in S_j}(x_l - \overline{x}_j)^2$ and for categorical covariates $H(\boldsymbol{x}^*_j) = \sum_{c=1}^{C} \hat{p}_{cj} \log \hat{p}_{cj}$, where C is the number of categories and $\hat{p}_{cj}$ the proportion of observations in category c in cluster j. Higher values of α lead to an increased penalty for dissimilar covariate values. Extensions to the multivariate case are possible (determinant of cluster-specific covariate matrices or multivariate entropy respectively).
2. For any number of covariates p and penalty $\theta = \alpha \in (0, \infty)$
$$g_2(\boldsymbol{x}^*_j|\theta) := \exp\left\{-\alpha \sum_{i,k \in S_j, i \neq k} d(\boldsymbol{x}_i, \boldsymbol{x}_k)\right\}$$
and
$$g_3(\boldsymbol{x}^*_j|\theta) := \exp\left\{-\frac{2\alpha}{n_j(n_j - 1)} \sum_{i,k \in S_j, i \neq k} d(\boldsymbol{x}_i, \boldsymbol{x}_k)\right\}$$

are based on the Gower Dissimilarity:

$$d(x_{il}, x_{jl}) := \begin{cases} \frac{|x_{il}-x_{jl}|}{\max_h x_{hl} - \min_h x_{hl}}, & \text{if } l\text{-th cov. continuous} \\ \delta_{x_{il}x_{jl}}, & \text{if } l\text{-th cov. categorical} \end{cases}$$

and $d(\boldsymbol{x}_i, \boldsymbol{x}_k)$ is the average of the Gower Dissimilarities in the p components.

3. *(Auxiliary Similarity Function)* With an auxiliary parameter $\boldsymbol{\xi}_j^*$, we have (prior predictive conjugate model)

$$g_4(\boldsymbol{x}_j^*|\boldsymbol{\theta}) := \int \prod_{i \in S_j} q(x_i|\boldsymbol{\xi}_j^*) q(\boldsymbol{\xi}_j^*) \mathrm{d}\boldsymbol{\xi}_j^*$$

For continuous covariates there are two options as a conjugate model. First, the *Auxiliary N-N Model* with $q(.|\xi_j^*) = \mathcal{N}(\cdot|\xi_j^*, \kappa_1 \hat{S})$ and $q(\xi_j^*) = \mathcal{N}(\xi_j^*|m_0, s_0^2)$ where $\hat{S}$ denotes the empirical variance of the covariate. The user-supplied parameters are $\boldsymbol{\theta} = (\kappa_1, m_0, s_0)$. Second, the *Auxiliary N-NIG Model* with $q(\cdot|\boldsymbol{\xi}_{\boldsymbol{j}}^{\boldsymbol{*}}) = \mathcal{N}(\cdot|m_j^*, v_j^*)$ and $q(\boldsymbol{\xi}_j^*) = \text{N-IG}(m_j^*, v_j^*|m_0, k_0, v_0, n_0)$. The user-supplied parameters are $\boldsymbol{\theta} = (m_0, k_0, v_0, n_0)$.
For categorical covariates a Multinomial-Dirichlet Model is applied, that is $q(\cdot|\boldsymbol{\xi}_j^*) = \text{Multinomial}(\cdot|\boldsymbol{\xi}_j^*)$ and $q(\boldsymbol{\xi}_j^*) = \text{Dirichlet}(\boldsymbol{\xi}_j^*|\boldsymbol{\alpha}_j \equiv a)$ where C is the number of cateogories. The user-supplied parameter is $\theta = a$.

4. *(Double Dipper)* With an auxiliary parameter $\boldsymbol{\xi}_j^*$, we have

$$g_5(\boldsymbol{x}_j^*|\boldsymbol{\theta}) := \int \prod_{i \in S_j} q(x_i|\boldsymbol{\xi}_j^*) q(\boldsymbol{\xi}_j^*|\boldsymbol{x}_j^*) \mathrm{d}\boldsymbol{\xi}_j^*$$

with the same options for the underlying models as for g_4. This option gives more weight on the local covariate structure compared to g_4.

3.3. DRPM Model: Dependent Modeling of Temporal Sequences of Random Partitions

In this model that is taken from [PQD21] and is implemented in the R-package `drpm` [Pag] finally a temporal evolvement of the partitions is considered. The overall model structure is the following:

$$\begin{aligned} Y_{it}|\boldsymbol{\mu}_t^*, \boldsymbol{\sigma}_t^{2*}, \mathbf{c}_t &\overset{\text{ind}}{\sim} \mathcal{N}(\mu_{c_{it}t}^*, \sigma_{c_{it}t}^{2*}) \quad \forall i = 1, \ldots, n; \, t = 1, \ldots, T \\ (\mu_{jt}^*, \sigma_{jt}^*)|\theta_t, \tau_t^2 &\overset{\text{ind}}{\sim} \mathcal{N}(\theta_t, \tau_t^2) \times \text{UN}(0, A_\sigma) \quad \forall j = 1, \ldots, k_t \\ (\theta_t, \tau_t) &\overset{\text{iid}}{\sim} \mathcal{N}(\phi_0, \lambda^2) \times \text{UN}\,(0, A_\tau) \quad \forall t = 1, \ldots, T \\ (\phi_0, \lambda) &\sim \mathcal{N}(m_0, s_0^2) \times \text{UN}(0, A_\lambda) \\ \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_T\} &\sim \text{tRPM}(\boldsymbol{\alpha}, M) \text{ with } \alpha_t \overset{\text{iid}}{\sim} \text{Beta}(a_\alpha, b_\alpha). \end{aligned}$$

The temporal random partition model models the temporal sequence of clusters as a first-order Markovian structure. We denote the clusters as $\rho_t = \{S_{1t}, \ldots, S_{k_t,t}\}$, $t = 1, \ldots, T$ or use the cluster-labeling notation.

The first ingredient for the model is an exchangeable probability function (EPPF) on the set of partitions of the measurement units. In our case

$$\mathrm{P}(\rho|M) = \frac{M^{k_\rho}}{\prod_{i=1}^n (M + i - 1)} \prod_{i=1}^{k_\rho} (|S_i| - 1)!$$

is applied. This is the marginal probability function for ρ derived from a Chinese Restaurant process with concentration parameter M. Smaller M favors less but larger clusters. This function is the prior for ρ_1.

Secondly, in order to define transition probabilities an auxiliary parameter γ_t is introduced. We define $\gamma_{it} \sim \text{Ber}(\alpha_t)$, i.e. $\gamma_t \in \{0,1\}^{\#\text{stations}}$ and give the following interpretation:

$$\gamma_{it} = \begin{cases} 1, & \text{station } i \text{ is } \textbf{not} \text{ relocated when moving from time } t-1 \text{ to } t \\ 0, & \text{else} \end{cases}.$$

The values of $\alpha_t \in [0,1]$ regulate the time-dependency, e.g. $\alpha_t = 1$ means $\rho_t = \rho_{t-1}$ with probability 1 and $\alpha_t = 0$ implies ρ_t is independent of ρ_{t-1}.

Given γ_t and ρ_{t-1} there is restriction to what partitions are compatible and can be considered for ρ_t. The transition probabilities are then

$$\mathbb{P}(\gamma_1, \rho_1, \ldots, \gamma_T, \rho_T) = \mathbb{P}(\rho_T|\gamma_T, \rho_{T-1}) \cdots \mathbb{P}(\rho_2|\gamma_2, \rho_1)\, \mathbb{P}(\rho_1)$$

and for $t \in \{1, \ldots, T\}$ the $\mathbb{P}(\rho_t|\gamma_t, \rho_{t-1})$ is given by the chosen EPPF from before truncated to the set of compatible partitions.

3.4. Extensions

For the DPRM Model we consider the extensions listed below. All of them were available in the `drpm` package [Pag]. The full extended model is taken from [PQD21, Section 4] which developed in the context of PM10 clustering.

Fixed α for each time step

Instead of drawing an α_t in each time step, we consider using the α_1 for the probability of non-relocation for each measurement unit in every time step. To be precise this is not a true extension, but is considered in order to elaborate, how the models' performance and computation time is effected.

Spatially informed DRPM

This extension combines the DRPM and the sPPM model. The EPPF in the temporal random partition model is changed to the function that was used as a prior in the sPPM-model, i.e. eq. (1). For similarity functions *Auxiliary Cohesion* and *Double Dipper* (g_3 and g_4 in section 3.1) are considered.

AR(1) structure in the Likelihood

As it is reasonable to assume that there is a temporal dependence for the time series at each measurement unit, a temporal structure in the likelihood is considered. For that purpose a measurement-unit specific time dependence parameter η_i with $|\eta_i| \leq 1$ is introduced. The resulting model is obtained when setting $\phi_1 = 0$ in the model presented in **??**.

AR(1) structure for θ_t

In addition a time-dependency in the time specific means θ_t can be added by bringing in another parameter ϕ_1. Combined with the AR(1) structure in the likelihood the fully extended model is the following:

$$\begin{aligned} Y_{it}|\boldsymbol{\mu}_t^*, \boldsymbol{\sigma}_t^{2*}, \mathbf{c}_t &\overset{\text{ind}}{\sim} \mathcal{N}(\mu_{c_{it}t}^* + \eta_i Y_{it-1}, \sigma_{c_{it}t}^{2*}(q - \eta_i^2)) \quad \forall i = 1, \ldots, n;\, t = 1, \ldots, T \\ Y_{i1} &\overset{\text{ind}}{\sim} \mathcal{N}(\mu_{c_{i1}1}^*, \sigma_{c_{i1}1}^{2*}) \quad \forall i = 1, \ldots, n \\ \xi_i = \text{Logit}(0.5(\eta_i + 1)) &\overset{\text{iid}}{\sim} \text{Laplace}(a, b) \quad \forall i = 1, \ldots, n \\ (\mu_{jt}^*, \sigma_{jt}^*)|\theta_t, \tau_t^2 &\overset{\text{ind}}{\sim} \mathcal{N}(\theta_t, \tau_t^2) \times \text{UN}(0, A_\sigma) \quad \forall j = 1, \ldots, k_t; t = 1, \ldots, T \\ \theta_t|\theta_{t-1} &\overset{\text{ind}}{\sim} \mathcal{N}((1-\phi_1)\phi_0 + \phi_1\theta_{t-1}, \lambda^2(1-\phi_1^2)) \\ (\theta_1, \tau_t) &\sim \mathcal{N}(\phi_0, \lambda^2) \times \text{UN}\,(0, A_\tau) \quad \forall t = 1, \ldots, T \\ (\phi_0, \phi_1, \lambda) &\sim \mathcal{N}(m_0, s_0^2) \times \text{UN}(-1, 1) \times \text{UN}(0, A_\lambda) \\ \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_T\} &\sim \text{tRPM}(\boldsymbol{\alpha}, M) \text{ with } \alpha_t \overset{\text{iid}}{\sim} \text{Beta}(a_\alpha, b_\alpha). \end{aligned} \tag{2}$$

Note that setting $\eta_i = 0$ and $\phi_1 = 0$ yields exactly the DRPM model presented in the preceding section. In our analysis we consider the full model and all possible combinations of the listed extensions to comprehensively assess the model's performance and computational time with respect to the additional introduced complexity.

4. Data Preparation and Evaluation

4.1. Data Aggregation

siamo passati dai dati giornalieri che avevamo a quelli settimanali, per ogni covariata abbiamo deciso se fare la media o prendere altri valori come massimo, minimo, moda ecc.

4.2. Data Exploration

The initial phase of data exploration necessitates a correlation analysis (**??**). The analysis reveals a strong negative correlation between one covariate and the PM2.5 levels (fig. 2). This covariate denotes the maximum Planetary Boundary Layer Height (BLH), which represents the lowermost part of the troposphere. The observed outcome aligns with expectations as the BLH encapsulates the entirety of our planet's pollution. Its negative correlation with PM2.5 levels signifies a significant relationship warranting further investigation.

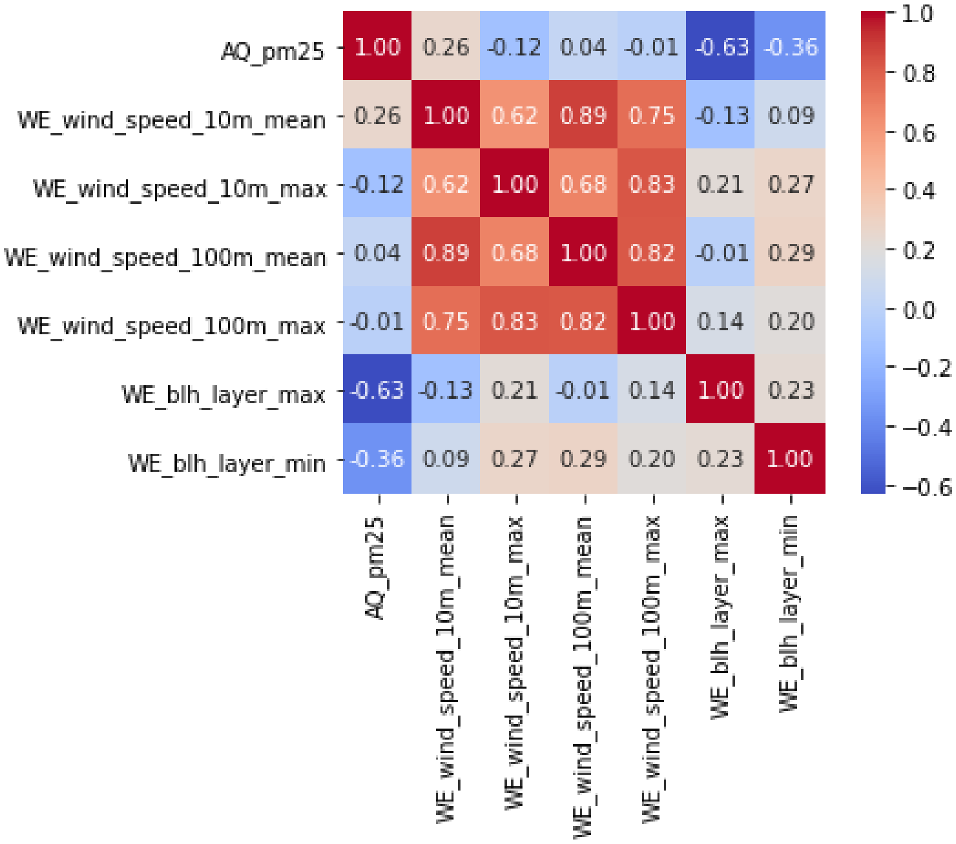


Figure 2: Correlation matrix between the different covariates of the dataset.

The subsequent step involves analyzing the variation in PM2.5 levels throughout the year across different stations and years. Stations in Bolzano, Milano, and Cremona (**??**) were chosen due to the distinct morphological characteristics of their territories and varying levels of industrialization. Additionally, these stations provide complete PM2.5 data spanning four different

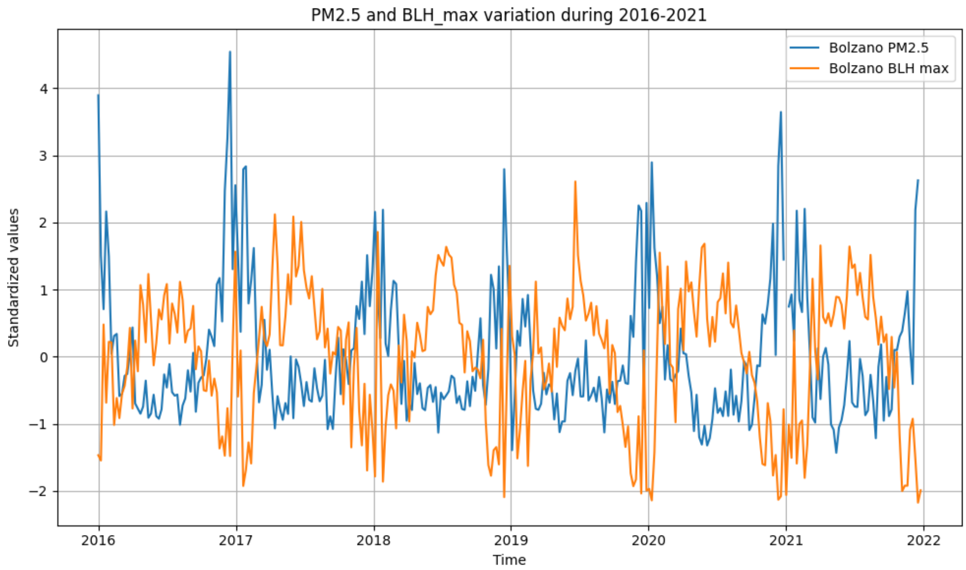


Figure 3: Correlation between PM2.5 Concentrations and Boundary Layer Height (BLH) max which represents the maximum depth of air next to the Earth's surface that is most affected by the resistance to the transfer of momentum, heat, or moisture across the surface.

years. The findings indicate a discernible U-shaped trend, suggesting pronounced seasonality; PM2.5 levels tend to be lower during the summer months. This pattern could potentially be attributed to reduced household heating impact and differing weather conditions compared to winter. Furthermore, notable differences between the year 2020 (amidst the COVID-19 pandemic) and other years were not evident, indicating that the primary factor might not be mobility-related. The plotted data distinctly illustrates that the Bolzano station, situated in a sparsely populated mountainous region, consistently exhibits lower PM2.5 levels compared to the other two stations. However, this station records PM2.5 levels below the recommended threshold set by the World Health Organization only during brief summer periods.

4.3. Imputation of Missing Data

per prima cosa abbiamo deciso quale anno analizzare in base alla completezza dei nostri valori di pm2.5 –> anno 2019. scrivo come abbiamo considerato i missing data –> prima eliminando le stazioni senza pm2.5, poi eliminando le covariate rimaste con meno valori e successivamente fissando i valori rimasti nulli interpolando con i valori delle settimane vicine

4.4. Evaluation

4.4.1 Goodness-of-fit

To indicate goodness-of-fit of the used models the following criteria are considered.

LPML = Log Pseudo Marginal Likelihood[1] (higher is better) The LPML is a predictive information criterion based on the idea of leave-on-out cross validation, the following definition is taken from [Gug23]. Let $\mathbf{y} = (y_1, \ldots, y_n)$ denote our data, $\mathbf{y}_{(-i)} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ and $m(y_i|\mathbf{y}_{(-i)})$ the marginal likelihood of y_i given $\mathbf{y}_{(-i)}$ in the considered model. Then the LPML is defined as

$$\mathrm{LPML} = \sum_{i=1}^{n} \log m(y_i|\mathbf{y}_{(-i)}).$$

[1]We refer to [PQD21, Section 4.1] for details on numerical problems when computing the LPML.

./imgs/maps/bolzano_milano_cremona_2017_2020.pdf

Figure 4: Time series of PM2.5 in Bolzano, Milano and Cremona (2017-2020).

WAIC = Widely Applicable Information Criterion (lower is better) The WAIC is another predictive information criteria that accounts for the over-estimation of log-pointwise predictive density $\sum_{i=1}^{n} \log m(y_i|\mathbf{y})$ by subtracting a penalization term p_{WAIC}, see [Gug23]. In our definition the WAIC is obtained by multiplying this difference with -2. In the following θ denotes the parameters and f the likelihood of the given model. The WAIC is defined as

$$\mathrm{WAIC} = -2\left(\sum_{i=1}^{n} \log m(y_i|\mathbf{y}) - p_{\mathrm{WAIC}}\right),$$

$$p_{\mathrm{WAIC}} = \sum_{i=1}^{n} \mathrm{Var}_{\theta|\mathbf{y}} \log f(y_i|\theta).$$

MSE = Mean Squared Error (lower is better) Denote by $\hat{y}_i = \mathbb{E}(Y_i|\mathbf{y})$ the posterior mean of the i-th observation, then

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

MaxDev = Maximal Deviation (lower is better) Given a partition $\rho = \{S_1, \dots, S_k\}$ we compute the maximum of the cluster-internal deviation of PM2.5 values over all cluster, namely

$$\mathrm{MaxDev} = \max_{i=1,\dots,k}\left(\max_{i\in S_k} y_i - \min_{i\in S_k} y_i\right)$$

as an indication on how closely related the target variable's values are given the partition ρ.

4.4.2 Cluster Estimation

Once a model is fit to the data the question remains how to summarize the obtained posterior for the clusters in a meaningful way. One number that we report is the posterior mean of the number of clusters. Furthermore, a general idea is to find an estimate $\hat{\rho}^*$ that minimizes a certain partition loss function L. Assuming that there is a "true" ρ (we take the posterior distribution), this becomes

$$\hat{\rho}^* = \operatorname{argmin}_{\hat{\rho}} \mathbb{E}(L(\rho, \hat{\rho})|\mathbf{y}) \approx \frac{1}{M} \sum_{m=1}^{M} L(\rho^{(m)}, \hat{\rho}),$$

where $\left(\rho^{(1)}, \ldots, \rho^{(M)}\right)$ are MCMC samples from the posterior. For that approach the R-package `salso` [DJa23] provides a number of possibilities. In [DJM22] different loss functions are explained as well as the `SALSO` algorithm that is implemented in the package.

Binder Loss One of the most widely used loss functions is the Binder loss function that considers pairwise misclassifications. Switching to the equivalent cluster notation for partitions the definition is

$$\mathrm{L}_{\text{Binder}}(\mathbf{c}, \hat{\mathbf{c}}) = \sum_{i<j} a \cdot I(\{c_i = c_j\}) I(\{\hat{c}_i \neq \hat{c}_j\}) + b \cdot I(\{c_i \neq c_j\}) I(\{\hat{c}_i = \hat{c}_j\}).$$

For $a = b = 1$ a measure of similarity between partitions, the Rand Index, is obtained by $\mathrm{RI}(\rho, \hat{\rho}) = 1 - \mathrm{L}_{\text{Binder}}(\rho, \hat{\rho}) / \binom{n}{2}$. Maximizing the the posterior expectation of the Rand Index is equivalent to minimizing the expected loss for a Binder loss function with $a = b = 1$. As this index fails to account for chance agreements, there exists a generalization that we will consider.

Adjusted Rand Index The Adjusted Rand Index (ARI) is defined as

$$\mathrm{ARI}(\rho, \hat{\rho}) = \frac{\sum_{s \in \rho} \sum_{E \in \hat{\rho}} \binom{|S \cap E|}{2} - \left(\sum_{S \in \rho} \binom{|S|}{2} \sum_{E \in \hat{\rho}} \binom{|E|}{2}\right) \frac{1}{\binom{n}{2}}}{\frac{1}{2}\left(\sum_{S \in \rho} \binom{|S|}{2} + \sum_{E \in \hat{\rho}} \binom{|E|}{2}\right) - \left(\sum_{S \in \rho} \binom{|S|}{2} \sum_{E \in \hat{\rho}} \binom{|E|}{2}\right) \frac{1}{\binom{n}{2}}}.$$

Large values mean a larger similarity between the partitions. We obtain another point estimate of the posterior expectation of the partition by maximizing the posterior expectation of the ARI.

5. Results

We used the same seed for every experiments to make the results reproducible and comparable. To obtain a single partition from all of our MCMC-samples, we use the `SALSO` function provided by the `salso` package with the Binder loss function the parameters $a = 1$. The upper bound for the number of clusters is set to `None` and thus capped by the maximum number of clusters provided in the MCMC iterates itself.

5.1. sPPM

M	Cohesion	WAIC	LPML	MSE	Clusters
10^{-3}	$C_{1_{\alpha=1}}$	24.72	14.94	0.12	1
	$C_{1_{\alpha=2}}$	24.08	13.01	0.12	2
	$C_{2_{a=0.5}}$	90.05	-25.80	0.17	15
	$C_{2_{a=1.5}}$	24.92	13.42	0.13	1
10^{-2}	$C_{1_{\alpha=1}}$	26.39	14.20	0.12	1
	$C_{1_{\alpha=2}}$	25.09	11.99	0.16	5
	$C_{2_{a=0.5}}$	∞	-23.35	0.15	15
	$C_{2_{a=1.5}}$	24.63	14.20	0.12	1
10^{-1}	$C_{1_{\alpha=1}}$	27.72	12.35	0.41	6
	$C_{1_{\alpha=2}}$	31.82	10.35	0.19	7
	$C_{2_{a=0.5}}$	85.11	-28.14	0.15	16
	$C_{2_{a=1.5}}$	25.27	13.66	0.13	1
1	$C_{1_{\alpha=1}}$	34.40	6.94	0.11	9
	$C_{1_{\alpha=2}}$	48.05	-1.26	0.12	13
	$C_{2_{a=0.5}}$	93.75	-28.31	0.19	18
	$C_{2_{a=1.5}}$	29.24	9.85	0.11	3
10	$C_{1_{\alpha=1}}$	70.37	-16.25	0.12	15
	$C_{1_{\alpha=2}}$	97.07	-30.71	0.16	20
	$C_{2_{a=0.5}}$	152.56	-57.61	0.31	25
	$C_{2_{a=1.5}}$	57.98	-9.95	0.11	12

Table 1: sPPM performances for different cohesion and M parameter values.

5.2. PPMx

5.3. DRPM

5.3.1 Non-spatially informed: Hyperparameter Gridsearch

In order to test the model's sensitivity and response with respect to different values of the hyperparameters M and the starting value α_0 we provide a large grid-search-like experiment. To investigate the model's dependency on different values of the prior parameters, we conduct each of the grid-search-like experiment for three different prior believes presented in table 1. The first prior values, namely DRPM-Paper, are directly taken from [PQD21, Section 4.1] in the context of monthly PM10 data and we consider this model as a baseline. Since our early explorations showed that these prior parameters lead to quite large clusters (most of the times the model only returned one or two clusters), we modified the prior values to incorporate a lower standard deviation for the likelihood, nameley Lower Std, and we provide a third set of prior values which additionally integrate the mean PM2.5 value of the year 2018 as a prior value for the predictive mean. To make the experiments comparable and reproducable, we use the same random seed for each of the experiments.

The results are comprehensively available in the folder `/report/tables/results` of the Github repository.[2] For the sake of simplicity and limited space available, we only present the best hyperparameters for each of the model in the summary table 2. Interestingly, as already mentioned, the baseline model using the DRPM-Paper prior values performs poorly and, as shown in fig. 3, all stations are in the same cluster for each time step. In contrast, the models using our two tuned prior values perform reasonably well, despite requiring longer computational times of factor 5 and 7 respectively, most probably due to a less informative prior on the α_t values. Additionally, we were surprised that the WAIC and MSE performance metrics do not correlate. Notwithstanding looking promising on paper with the lowest MSE of 1.271, the DRPM-Paper informed model performs poorly in practice, as it is obvious that a clustering of all stations in one cluster is absolutely not desirable. Consequently, despite having an higher MSE, our prior values are favorable. An exemplary clustering of the three models is visualized in fig. 5. In fig. 3 we analyzed the three different models with respect to their number and sizes of clusters. Although the time-evolvement of α_t is somehow similar for all three models, the number of clusters significantly differ and the mean-informed Mean 2018 version favors slightly more clusters than the zero-centered mean version Lower Std. The MSE of the model using our two priors is nearly equal.

In order to analyze the convergence behaviour of the MCMC, we exemplarily visualize trace plots for the parameters of our Lower Std prior model, as it was the best performing one in our initial test. The plots in fig. 4 clearly exhibit the desirable "fat caterpillar" structure for the parameters $\mu^*_{c_1t}$, τ^2_t and ϕ_0, nonetheless for the rest of the shown parameters the convergence behaviour could be improved. Given the results presented in [PQD21, Section 4.1] with 50.000 MCMC samples, we expect this behaviour to vanish when increasing the number of samples as well as the burn-in and the thinning. Owing to constraints in computational resources, the exploration of this aspect is deferred to a future investigation.

Name \ Prior Parameters	m_0	s_0^2	A_σ	A_r	A_λ	b	a_α	b_α
DRPM-Paper [PQD21]	0.0	100^2	10.0	5.0	5.0	1.0	2.0	2.0
Lower Std (ours)	0.0	200	0.1	1.0	1.0	1.0	1.0	1.0
Mean 2018 (ours)	2.91	200	0.1	1.0	1.0	1.0	1.0	1.0

Table 2: Different Prior Parameters for the three models we used for our initial large test.

[2] See `https://github.com/Flo-Wo/PM25-Clustering/tree/main/report/tables/results`

Prior	M	α_0	LPML	WAIC	Time [s]	MSE	MaxDev
DRPM-Paper	100.0	0.25	$-1.234 \cdot 10^{+03}$	$2.445 \cdot 10^{+03}$	**12.70**	**1.271**	1.753
Lower Std	0.1	0.25	–	$\mathbf{-1.285 \cdot 10^{+03}}$	68.71	1.696	**1.495**
Mean 2018	0.1	0.25	–	$-9.548 \cdot 10^{+02}$	90.26	1.699	1.679

Table 3: Non-spatially informed DRPM model performances for different prior values. A dash value indicates that the `drpm_fit` function was not able to compute the corresponding values due to unknown reasons. Bold values indicate the column-wise best result.

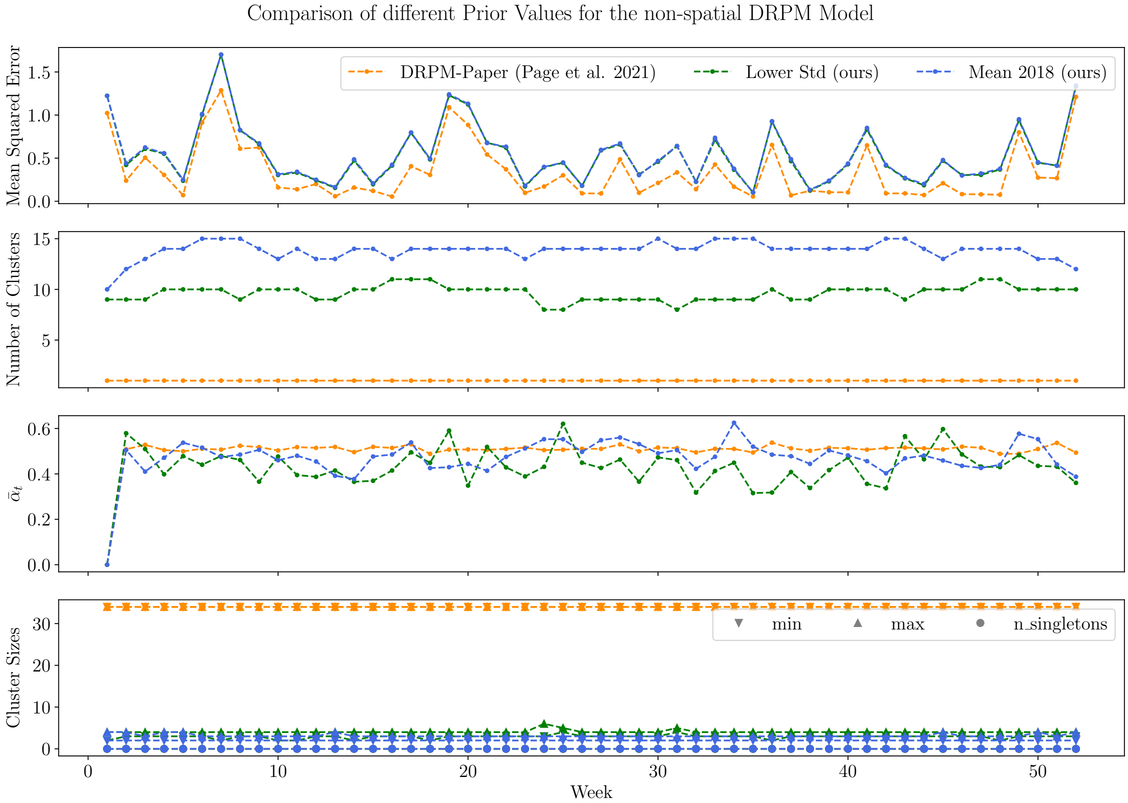


Figure 5: Comparison of the best model for each prior for DRPM without spatial cohesion. The used prior values are listed in table 1. For the DRPM model itself we use $M = 0.1$ as the concentration parameter. The MCMC uses 10000 draws with a burn-in of 1000 and a thinning of 10, resulting in 900 total MCMC samples. The DRPM-Paper model reached a WAIC (lower is better) of $3.103 \cdot 10^{+03}$ while our models achieved a WAIC score of $-1.285 \cdot 10^{+03}$ and $-9.548 \cdot 10^{+02}$ for the Lower Std and Mean 2018 models respectively. For the cluster sizes, the minimum (min) and maximum (max) number of stations for each timestep is shown, as well as the number of singletons, i.e. clusters consisting of only one station. $\bar{\alpha}_t$ indicates the mean of the MCMC samples at each time step $t = 1, \ldots, T$.

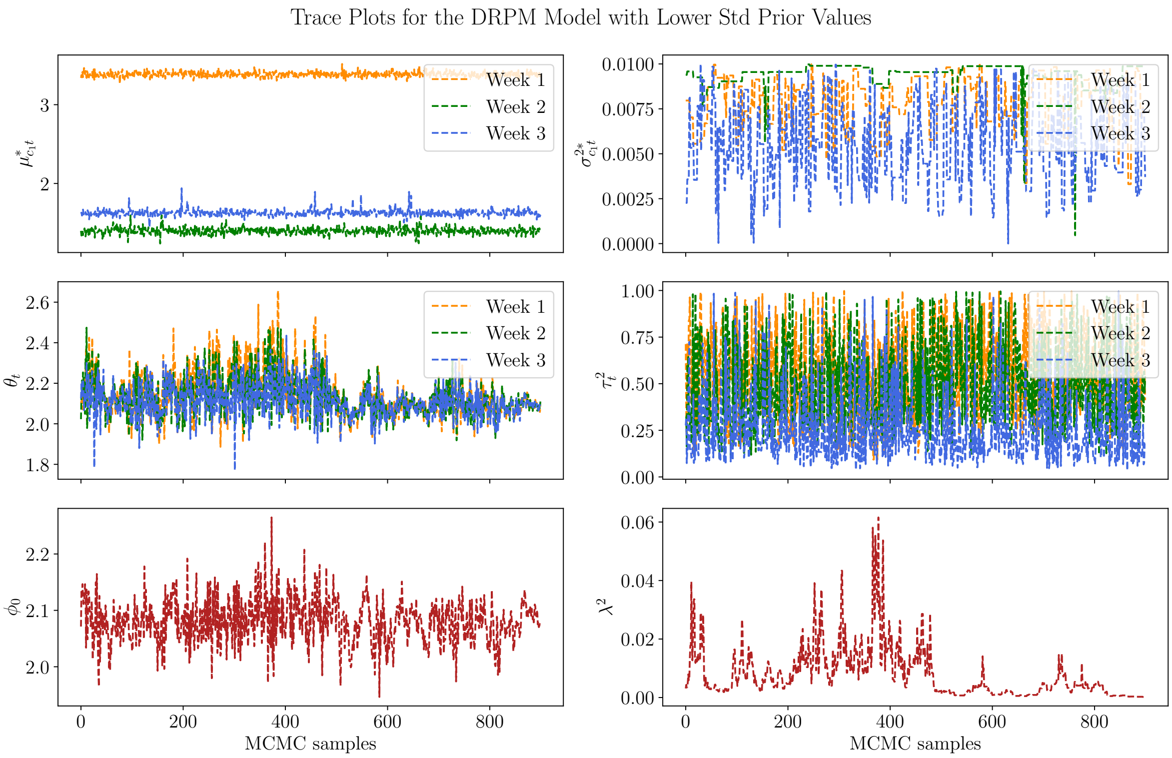


Figure 6: Trace Plots for the DRPM Model without spatial information parameters with the Lower Std Prior values. Week-specific parameters are shown for the first three weeks of the year and cluster specific parameters are presented for the first cluster.

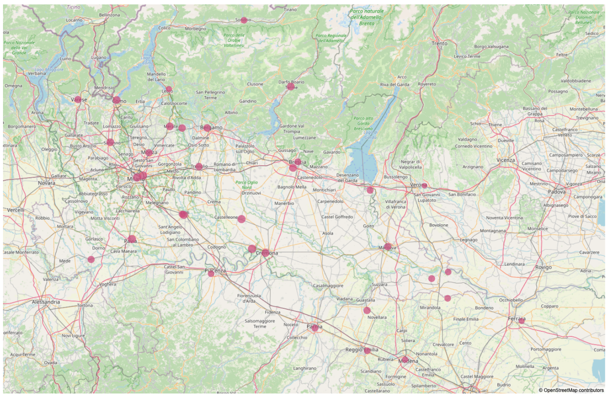
5.3.2 Spatially informed: Hyperparameter Gridsearch

We performed the same hyperparameter grid search for M and α_0 as in section 5.3.1 but now the DRPM model is spatially informed, i.e. has access to the stations' longitude and latitude, allowing us to simultaneously test the two cohesion functions 3 and 4 (indicated by g_i) of section 3.2 in our large experiment.[3] As before, the results are comprehensively available in the Github repository and table 3 provides a summarizing overview of the best performing models for each of the three prior combinations. We picked the best performing case for each model and for the reason of comparability, we display the results using the secondary cohesion function as well. Interestingly, the spatial-informed models prefer a slightly higher value of α_0 but the same value for the concentration parameter M. Overall, but especially for the Lower Std Prior values, the performance is slightly decreased and simultaneously the higher model complexity caused a significant increase in the computational time. As can be seen from **??** the number of clusters in the models with the Mean 2018-prior is clearly smaller for the best spatially-informed model compared to the best model without spatial information.

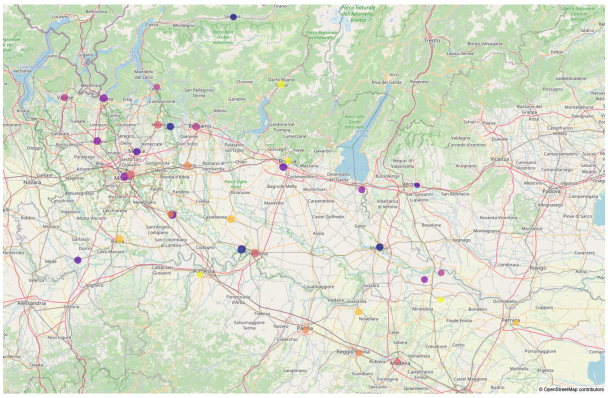
Similarly to [PQD22] we provide a lagged ARI value comparison for spatially and non-spatially informed models using the three different priors. Each matrix in **??** contains at the entry (i, j) the lagged ARI value between the partition at week i and week j, for $i, j = 1, \ldots, T$. Here, it is evident that the DRPM-Paper prior is not informative, since all the stations are in the same cluster for all of the weeks.

Regarding the overall structure for the Lower Std Prior, both, the non-spatial and spatial informed ARI matrices, exhibit a similar structure. Nevertheless, the spatial informed version show a generally higher time-connectivity.

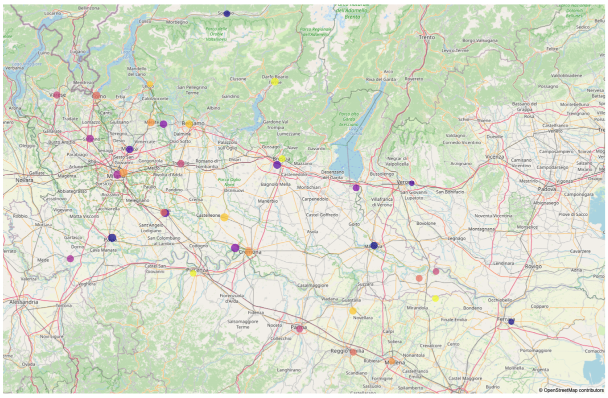
[3]Cf. [PQD21, Section 4.2] for more details.



(a) DRPM-Paper [PQD21]: 1 cluster in total.



(b) Lower Std (ours): 9 clusters in total.



(c) Mean 2018 (ours): 13 clusters in total.

Figure 7: Exemplary clustering for week three of the year 2019 using all three of our base models without spatial information. All use the concentration parameter $M = 0.1$ and the prior parameters listed in table 1. The color of the bubble indicates the cluster and the size of the bubble the weekly-average PM2.5 value.

Prior	M	α_0	g_i	LPML	WAIC	Time [s]	MSE	MaxDev
DRPM-Paper	0.1	0.5	3	$\mathbf{-1.217 \cdot 10^{+03}}$	$2.422 \cdot 10^{+03}$	$\mathbf{2.672 \cdot 10^{+01}}$	**1.257**	1.753
DRPM-Paper	0.1	0.5	4	$-1.509 \cdot 10^{+03}$	$2.988 \cdot 10^{+03}$	$3.235 \cdot 10^{+01}$	1.348	1.753
Lower Std	0.1	0.5	3	–	$1.705 \cdot 10^{+01}$	$1.200 \cdot 10^{+02}$	1.688	1.621
Lower Std	0.1	0.5	4	–	$-4.022 \cdot 10^{+02}$	$2.208 \cdot 10^{+02}$	1.700	**1.495**
Mean 2018	0.1	0.0	3	–	$2.042 \cdot 10^{+02}$	$1.200 \cdot 10^{+02}$	1.697	1.679
Mean 2018	0.1	0.0	4	–	$\mathbf{-5.403 \cdot 10^{+02}}$	$2.536 \cdot 10^{+02}$	1.708	1.541

Table 4: Spatially informed DRPM model performances for different prior values. A dash value indicates that the `drpm_fit` function was not able to compute the corresponding values due to unknown reasons. Bold values indicate the column-wise best result.

Interestingly, for the non-spatially informed Mean 2018 prior, we observe an outlier-like behavior for week 23, which seems to be unrelated with the previous and the upcoming weeks. This behavior vanishes for the spatially informed version, for which we observe a clean and nearly time-uniform relationship for all weeks, which is similar to the corresponding Lower Std Prior version.

5.3.3 Spatially informed: Time-dependency Extensions

We fix the values of $M = 0.1$ and $\alpha_0 = 0.5$ for all of the experiments in this section, since our previous results emphasized these two values as a good trade-off for all methods. Furthermore, the MSE is computed for all values in the log-space and is the average MSE computed over all weekly MSEs. Given our extension described in section 3.4, we compare our three base models by using a spatial-informed version of DRPM and contrasting all possible combinations of temporal-dependency extensions.

In our notation η_{10} = True indicates an AR(1)-type temporal-dependency in the likelihood, ϕ_1 = True a temporal-dependency within the stations and α_t = True a temporal-dependency within the partitions, i.e. α_t = False implies a constant value of $\alpha_t = \alpha_1$ over the entire time horizon $t = 1, \ldots, T$. Finally, g_i is the type of cohesion function used by the algorithm.

The results for our baseline version using the DRPM-Paper prior values is shown in table 4 and for the Lower Std and Mean 2018 version in table 5 and table 6 respectively.

Upon analyzing the results from table 4 to table 6, it becomes evident that more complex models come with increased computational complexity. Unfortunately, similar to the previous chapter, more complex temporal-dependencies cannot significantly enhance the models' performance. The intra-group comparison reveals the same pattern as in the non-spatially informed case: the DRPM-Paper version performs poorly compared to the Lower Std and Mean 2018 versions, with the latter performing slightly better. Nevertheless, in comparison to our base models in table 2, none of the three models surpass the performance of the non-spatially informed Lower Std version. Interestingly, we observe that for some hyperparameter combinations, the choice of cohesion functions strongly influences the WAIC score (e.g., in table 5 with all values being False). In contrast, the different temporal-dependencies only negligibly affect the MSE.

5.3.4 DRPM Conclusion

Based on the preceding sections, our exploration of various temporal-dependency extensions within the DRPM framework reveals that more complex models entail higher computational complexity without a significant improvement in performance. Despite variations in hyperparameter combinations and cohesion functions, the non-spatially informed Lower Std version consistently outperforms the DRPM-Paper and Mean 2018 versions. It is noteworthy that in the case of clustering,

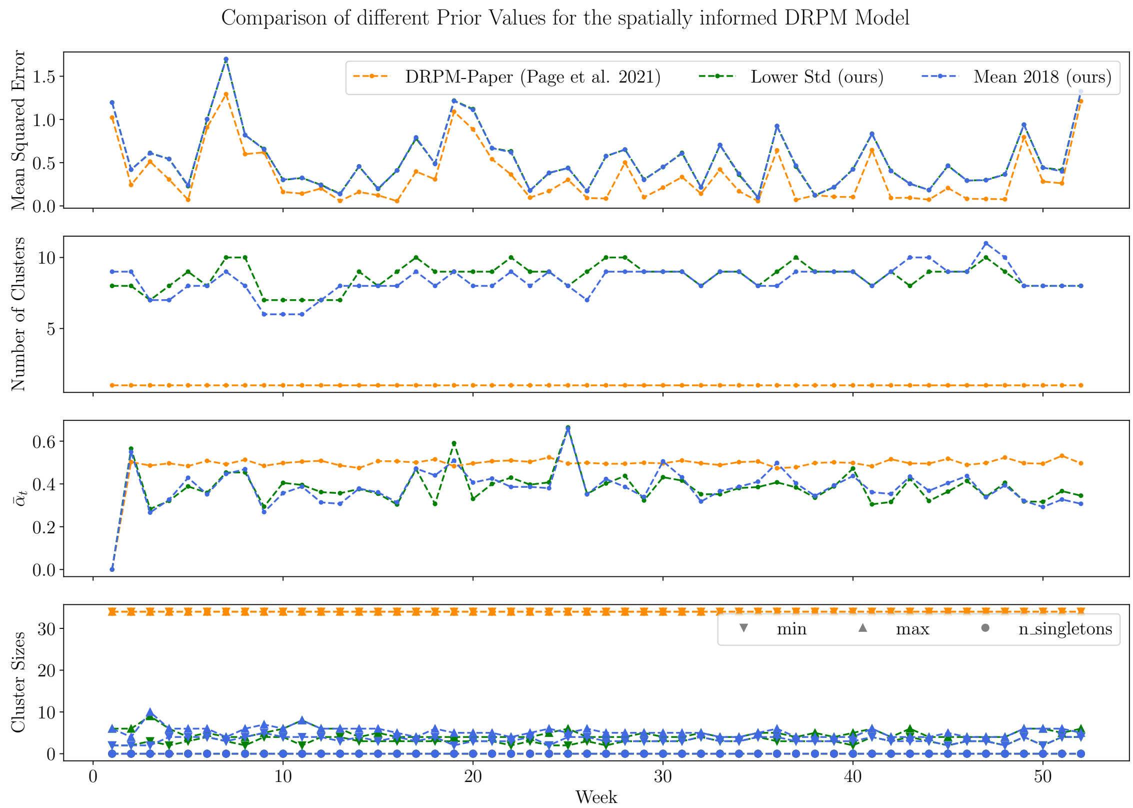


Figure 8: Comparison of the best model based on WAIC value for each prior for DRPM with spatial cohesion. The used prior values are listed in table 1. For the DRPM model itself we use $M = 0.1$ as the concentration parameter. The chosen similarity function was g_3 for the DRPM-Paper prior and g_4 for the other two priors. The MCMC uses 10000 draws with a burn-in of 1000 and a thinning of 10, resulting in 900 total MCMC samples. The DRPM-Paper model reached a WAIC (lower is better) of $2.422 \cdot 10^{+03}$ while our models achieved a WAIC score of $-4.022 \cdot 10^{+03}$ and $-5.403 \cdot 10^{+02}$ for the Lower Std and Mean 2018 models respectively. For the cluster sizes, the minimum (min) and maximum (max) number of stations for each timestep is shown, as well as the number of singletons, i.e. clusters consisting of only one station. $\bar{\alpha}_t$ indicates the mean of the MCMC samples at each time step $t = 1, \ldots, T$.

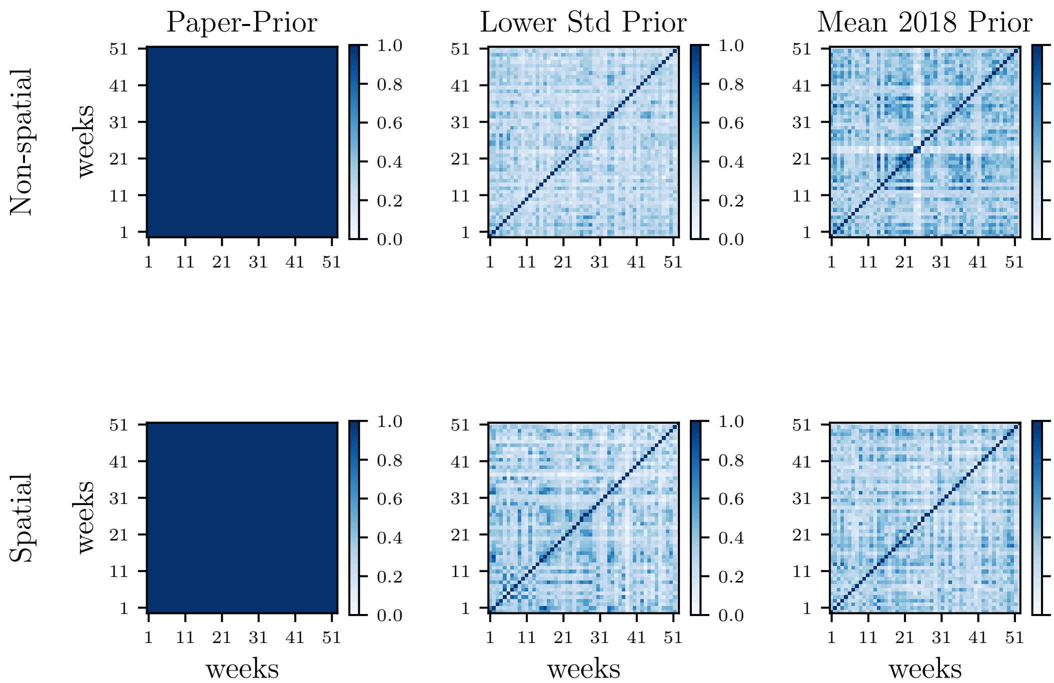


Figure 9: Comparison of the lagged Adjusted Rand Index (ARI). For better comparability all models use parameters $M = 0.1$ and $\alpha_0 = 0.5$. For the spatially-informed models the optimal cohesion functions for each prior are used, that is g_3 for the Paper-Prior and g_4 for the Lower Std-Prior and the Mean 2018-Prior. The ARI is computed for the partition estimates that we obtain using the Binder Loss Function with parameters $a = b = 1$ in the `salso`package. The ARI is computed using the python-package `scikit-learn`.

η_{10}	ϕ_1	α_t	g_i	LMPL	WAIC	Time [s]	MSE
False	False	True	3	$-1.202 \cdot 10^{+03}$	$2.382 \cdot 10^{+03}$	$2.719 \cdot 10^{+01}$	$1.248 \cdot 10^{+00}$
False	False	True	4	$-1.217 \cdot 10^{+03}$	$2.407 \cdot 10^{+03}$	$2.854 \cdot 10^{+01}$	$1.238 \cdot 10^{+00}$
False	True	True	3	$-1.199 \cdot 10^{+03}$	$2.378 \cdot 10^{+03}$	$2.698 \cdot 10^{+01}$	$1.256 \cdot 10^{+00}$
False	True	True	4	$-1.226 \cdot 10^{+03}$	$2.421 \cdot 10^{+03}$	$3.068 \cdot 10^{+01}$	$\mathbf{1.197 \cdot 10^{+00}}$
True	False	True	3	$-1.329 \cdot 10^{+03}$	$2.644 \cdot 10^{+03}$	$2.726 \cdot 10^{+01}$	$1.279 \cdot 10^{+00}$
True	False	True	4	$-1.256 \cdot 10^{+03}$	$2.486 \cdot 10^{+03}$	$3.311 \cdot 10^{+01}$	$1.235 \cdot 10^{+00}$
True	True	True	3	$-1.710 \cdot 10^{+03}$	$3.407 \cdot 10^{+03}$	$2.721 \cdot 10^{+01}$	$1.376 \cdot 10^{+00}$
True	True	True	4	$-1.766 \cdot 10^{+03}$	$3.500 \cdot 10^{+03}$	$3.397 \cdot 10^{+01}$	$1.391 \cdot 10^{+00}$
False	False	False	3	$\mathbf{-1.198 \cdot 10^{+03}}$	$\mathbf{2.375 \cdot 10^{+03}}$	$\mathbf{2.647 \cdot 10^{+01}}$	$1.221 \cdot 10^{+00}$
False	False	False	4	$-1.275 \cdot 10^{+03}$	$2.491 \cdot 10^{+03}$	$4.507 \cdot 10^{+01}$	$1.200 \cdot 10^{+00}$
False	True	False	3	$-1.199 \cdot 10^{+03}$	$2.376 \cdot 10^{+03}$	$2.757 \cdot 10^{+01}$	$1.237 \cdot 10^{+00}$
False	True	False	4	$-1.264 \cdot 10^{+03}$	$2.473 \cdot 10^{+03}$	$4.155 \cdot 10^{+01}$	$1.215 \cdot 10^{+00}$
True	False	False	3	$-1.279 \cdot 10^{+03}$	$2.545 \cdot 10^{+03}$	$2.700 \cdot 10^{+01}$	$1.262 \cdot 10^{+00}$
True	False	False	4	$-1.958 \cdot 10^{+03}$	$3.878 \cdot 10^{+03}$	$4.503 \cdot 10^{+01}$	$1.488 \cdot 10^{+00}$
True	True	False	3	$-1.601 \cdot 10^{+03}$	$3.189 \cdot 10^{+03}$	$2.686 \cdot 10^{+01}$	$1.343 \cdot 10^{+00}$
True	True	False	4	$-1.937 \cdot 10^{+03}$	$3.837 \cdot 10^{+03}$	$4.398 \cdot 10^{+01}$	$1.430 \cdot 10^{+00}$

Table 5: **Spatially informed** DRPM Model for different hyperparameter configurations with the following prior values: $m_0 = 0.0$, $s_0^2 = 10000.0$, $A_\sigma = 10.0$, $A_\tau = 5.0$, $A_\lambda = 5.0$, $b = 1.0$, $a_\alpha = 2.0$, $b_\alpha = 2.0$ (**DRPM-Paper**).

η_{10}	ϕ_1	α_t	g_i	LMPL	WAIC	Time	MSE
False	False	True	3	$-\infty$	$-3.154 \cdot 10^{+02}$	$1.289 \cdot 10^{+02}$	$1.690 \cdot 10^{+00}$
False	False	True	4	–	$-1.931 \cdot 10^{+02}$	$2.315 \cdot 10^{+02}$	$1.698 \cdot 10^{+00}$
False	True	True	3	–	$-3.572 \cdot 10^{+02}$	$1.455 \cdot 10^{+02}$	$1.695 \cdot 10^{+00}$
False	True	True	4	–	$3.747 \cdot 10^{+03}$	$2.449 \cdot 10^{+02}$	$1.705 \cdot 10^{+00}$
True	False	True	3	–	$4.923 \cdot 10^{+02}$	$1.789 \cdot 10^{+02}$	$1.708 \cdot 10^{+00}$
True	False	True	4	–	$-4.251 \cdot 10^{+02}$	$2.020 \cdot 10^{+02}$	$1.697 \cdot 10^{+00}$
True	True	True	3	$\mathbf{-1.627 \cdot 10^{+03}}$	$-3.593 \cdot 10^{+02}$	$\mathbf{1.086 \cdot 10^{+02}}$	$\mathbf{1.682 \cdot 10^{+00}}$
True	True	True	4	–	$3.429 \cdot 10^{+03}$	$3.682 \cdot 10^{+02}$	$1.702 \cdot 10^{+00}$
False	False	False	3	–	$-3.490 \cdot 10^{+02}$	$1.345 \cdot 10^{+02}$	$1.682 \cdot 10^{+00}$
False	False	False	4	–	$9.056 \cdot 10^{+00}$	$2.713 \cdot 10^{+02}$	$1.710 \cdot 10^{+00}$
False	True	False	3	–	$-4.750 \cdot 10^{+02}$	$2.016 \cdot 10^{+02}$	$1.695 \cdot 10^{+00}$
False	True	False	4	–	$\mathbf{-6.328 \cdot 10^{+02}}$	$2.183 \cdot 10^{+02}$	$1.697 \cdot 10^{+00}$
True	False	False	3	–	$2.319 \cdot 10^{+02}$	$1.721 \cdot 10^{+02}$	$1.697 \cdot 10^{+00}$
True	False	False	4	–	$-3.644 \cdot 10^{+02}$	$2.026 \cdot 10^{+02}$	$1.695 \cdot 10^{+00}$
True	True	False	3	–	$1.314 \cdot 10^{+02}$	$1.452 \cdot 10^{+02}$	$1.693 \cdot 10^{+00}$
True	True	False	4	–	$-1.356 \cdot 10^{+02}$	$2.083 \cdot 10^{+02}$	$1.695 \cdot 10^{+00}$

Table 6: **Spatially informed** DRPM Model for different hyperparameter configurations with the following prior values: $m_0 = 0.0$, $s_0^2 = 200.0$, $A_\sigma = 0.1$, $A_\tau = 1.0$, $A_\lambda = 1.0$, $b = 1.0$, $a_\alpha = 1.0$, $b_\alpha = 1.0$ (**Lower Std**). A dash indicates that the package was not able to calculate the corresponding value.

η_{10}	ϕ_1	α_t	g_i	LMPL	WAIC	Time	MSE
False	False	True	3	–	$-6.109 \cdot 10^{+02}$	$1.245 \cdot 10^{+02}$	$1.690 \cdot 10^{+00}$
False	False	True	4	–	$-3.894 \cdot 10^{+02}$	$2.151 \cdot 10^{+02}$	$1.700 \cdot 10^{+00}$
False	True	True	3	$-5.115 \cdot 10^{+03}$	$-4.875 \cdot 10^{+02}$	$1.412 \cdot 10^{+02}$	$1.685 \cdot 10^{+00}$
False	True	True	4	–	$-2.315 \cdot 10^{+02}$	$2.505 \cdot 10^{+02}$	$1.702 \cdot 10^{+00}$
True	False	True	3	–	$4.319 \cdot 10^{+02}$	$\mathbf{1.073 \cdot 10^{+02}}$	$\mathbf{1.674 \cdot 10^{+00}}$
True	False	True	4	–	$-1.461 \cdot 10^{+02}$	$2.288 \cdot 10^{+02}$	$1.702 \cdot 10^{+00}$
True	True	True	3	–	$5.938 \cdot 10^{+01}$	$1.098 \cdot 10^{+02}$	$1.684 \cdot 10^{+00}$
True	True	True	4	–	$-3.021 \cdot 10^{+02}$	$2.043 \cdot 10^{+02}$	$1.699 \cdot 10^{+00}$
False	False	False	3	–	$\mathbf{-6.850 \cdot 10^{+02}}$	$1.817 \cdot 10^{+02}$	$1.690 \cdot 10^{+00}$
False	False	False	4	–	$-4.635 \cdot 10^{+02}$	$2.898 \cdot 10^{+02}$	$1.701 \cdot 10^{+00}$
False	True	False	3	–	$-5.577 \cdot 10^{+02}$	$1.536 \cdot 10^{+02}$	$1.688 \cdot 10^{+00}$
False	True	False	4	–	$-2.545 \cdot 10^{+02}$	$2.438 \cdot 10^{+02}$	$1.708 \cdot 10^{+00}$
True	False	False	3	$\mathbf{-2.353 \cdot 10^{+03}}$	$-2.660 \cdot 10^{+02}$	$1.189 \cdot 10^{+02}$	
True	False	False	4	–	$-5.078 \cdot 10^{+02}$	$2.501 \cdot 10^{+02}$	$1.701 \cdot 10^{+00}$
True	True	False	3	–	$1.608 \cdot 10^{+04}$	$4.449 \cdot 10^{+02}$	$1.715 \cdot 10^{+00}$
True	True	False	4	–	$3.923 \cdot 10^{+03}$	$2.871 \cdot 10^{+02}$	$1.716 \cdot 10^{+00}$

Table 7: **Spatially informed** DRPM Model for different hyperparameter configurations with the following prior values: $m_0 = 2.91$, $s_0^2 = 200.0$, $A_\sigma = 0.1$, $A_\tau = 1.0$, $A_\lambda = 1.0$, $b = 1.0$, $a_\alpha = 1.0$, $b_\alpha = 1.0$ (**Mean 2018**). A dash indicates that the package was not able to calculate the corresponding value.

one should not solely rely on computational metrics such as WAIC or MSE. As evidenced in the non-spatially informed DRPM-Paper case, where some metric-based results appear promising, the actual proposed clustering is rather poor and uninformative. In conclusion, we want to highlight the improved model performance induced by the temporal-dependency of the partitions while emphasizing the limited enhancements of more sophisticated models.

6. Conclusions

References

[DJa23] David B. Dahl, Devin J Johnson, and et al. *Package 'salso': Search Algorithms and Loss Functions for Bayesian Clustering*. version 0.3.35. 2023. URL: `https://github.com/dbdahl/salso`.

[DJM22] David B. Dahl, Devin J. Johnson, and Peter Müller. "Search Algorithms and Loss Functions for Bayesian Clustering". In: *Journal of Computational and Graphical Statistics* 31.4 (2022), pp. 1189–1201. DOI: `10.1080/10618600.2022.2069779`. eprint: `https://doi.org/10.1080/10618600.2022.2069779`. URL: `https://doi.org/10.1080/10618600.2022.2069779`.

[Gug23] Alessandra Guglielmi. *Bayesian Statistics: Lecture Notes*. 2023.

[MQR11] Peter Müller, Fernando A Quintana, and Gary L. Rosner. "A Product Partition Model With Regression on Covariates". In: *Journal of Computational and Graphical Statistics* 20.1 (2011). PMID: 21566678, pp. 260–278. DOI: `10.1198/jcgs.2011.09066`. eprint: `https://doi.org/10.1198/jcgs.2011.09066`. URL: `https://doi.org/10.1198/jcgs.2011.09066`.

[Pag] Garritt L. Page. *Package 'drpm': Dependent Random Partition Model*. version 0.1.2. URL: `https://github.com/gpage2990/drpm`.

[Pag+23] Garritt L. Page et al. *Package 'ppmSuite': A Collection of Models that Employ Product Partition Distributions as a Prior on Partitions*. version 0.3.4. 2023. URL: `https://cran.r-project.org/web/packages/ppmSuite/index.html`.

[PQ16] Garritt L. Page and Fernando A. Quintana. "Spatial Product Partition Models". In: *Bayesian Analysis* 11.1 (2016), pp. 265–298. DOI: `10.1214/15-BA971`. URL: `https://doi.org/10.1214/15-BA971`.

[PQ17] Garritt L. Page and Fernando A. Quintana. "Calibrating covariate informed product partition models". In: *Statistics and Computing* 28 (2017), pp. 1009–1031. DOI: `10.1007/s11222-017-9777-z`. URL: `https://doi.org/10.1007/s11222-017-9777-z`.

[PQD21] Garritt L. Page, Fernando A. Quintana, and David B. Dahl. *Dependent Modeling of Temporal Sequences of Random Partitions*. 2021. arXiv: `1912.11542 [stat.ME]`.

[PQD22] Garritt L. Page, Fernando A. Quintana, and David B. Dahl. "Spatio-Temporal Random Partition Models". In: (2022). URL: `https://arxiv.org/pdf/1912.11542v1.pdf`.

[R C23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2023. URL: `https://www.R-project.org/`.

[RLJ21] G.L. Rosner, P.W. Laud, and W.O. Johnson. *Bayesian Thinking in Biostatistics*. 1st edition. Chapman and Hall/CRC, 2021. URL: `https://doi.org/10.1201/9781439800102`.

[Rod+23] Alessandro Fassò Jacopo Rodeschini et al. "Agrimonia: a dataset on livestock, meteorology and air quality in the Lombardy region, Italy". In: *Scientific Data* 10.1 (Mar. 2023). ISSN: 2052-4463. DOI: `10.1038/s41597-023-02034-0`. URL: `http://dx.doi.org/10.1038/s41597-023-02034-0`.