Politecnico di Milano

School of Industrial and Information Engineering
Master of Science in Mathematical Engineering

MASTER THESIS

forse qui non starebbe bene un verbo? ad esempio
giving more flexibility...
adding more flexibility...
granting/providing/ecc more flexibility...

The DRPM Strikes Back: More Flexibility for a Bayesian Spatio-Temporal Clustering Model

Advisor
Prof. Alessandra Guglielmi

Coadvisor
Prof. Alessandro Carminati

Candidate
Federico Angelo Mor
Matr. 221429

Academic Year 2023–2024

Abstract

Clustering is a key technique for identifying patterns and structures in complex datasets, whose relevance is intensified in spatio-temporal contexts where observations are simultaneously influenced by multiple factors such as space, time, and covariates. This complexity can be effectively tamed by model-based clustering methods, which often provide more accurate and interpretable results with respect to traditional frequentist approaches thanks to the possibility of encoding data information directly inside the model. To this end, the Dependent Random Partition Model [PQD22] is one of the most relevant Bayesian models due to its explicit consideration of temporal dependence in the partitions. However, the current formulation of the model and the implementation of the associated MCMC algorithm lack of the inclusion of covariates, the handling of missing data, and the efficiency in execution times. Therefore, in this work we improve the original DRPM by addressing those issues trough updates on the model formulation and a brand new implementation in Julia. These advancements are then tested on synthetic and real-world datasets, including air quality data from the AgrImOnIA project [Fas+23] in Lombardy, Italy.

KEYWORDS: Bayesian modelling, clustering, spatio-temporal data, MCMC, Julia

Sommario

Il clustering è una tecnica fondamentale per identificare strutture e pattern in dataset complessi, la cui importanza è intensificata nei contesti spazio-temporali in cui le osservazioni sono influenzate simultaneamente da molteplici fattori come spazio, tempo e covariate. Questa complessità può essere efficacemente gestita da metodi di clustering basati su modelli, che spesso forniscono risultati più precisi e interpretabili rispetto agli approcci frequentisti tradizionali grazie alla possibilità di inserire informazioni riguardo ai dati direttamente all'interno del modello. In tal senso, il Dependent Random Partition Model [PQD22] (modello per partizioni aleatorie dipendenti) è uno dei modelli bayesiani più rilevanti in quanto tiene conto in modo esplicito della dipendenza temporale delle partizioni. Tuttavia, l'attuale formulazione del modello e la sua corrispondente implementazione dell'algoritmo di campionamento mancano dell'inclusione di covariate, della gestione dei dati mancanti, e di efficienza nei tempi di esecuzione. In questo lavoro abbiamo quindi migliorato l'originale DRPM affrontando tali problemi tramite aggiornamenti sulla formulazione del modello e una fiammante implementazione in Julia. Questi sviluppi sono stati poi testati su dataset sintetici e reali, compresi i dati sulla qualità dell'aria in Lombardia del progetto AgrImOnIA [Fas+23].

Parole chiave: modellistica bayesiana, clustering, dati spazio-temporali, MCMC, Julia

Introduction

Clustering is the canonical example of unsupervised learning, where a set of data points has to be divided into homogenous groups of units which exhibit a similar behaviour with respect to a target variable. It has always been a powerful tool to identify structures and patterns into data, especially in contexts where relationships between the observations are complex, e.g. when the target variable is affected by many factors simultaneously. For this reason, clustering techniques saw a continuos increase in popularity in a variety of scientific fields, including social sciences, climate and environmental analysis, economics, and healthcare.

In general, clustering approaches fall into two main categories: algorithmic and model-based methods. qui non ci va l' "a capo"

Algorithmic methods such as hierarchical, partition-based, or density-based methods treat the clustering problem as an optimization problem where a certain metric has to be minimised (or maximised). Partition-based methods as k-means generate the clusters around a set of centroids which are iteratively updated to minimize the within-cluster variance, i.e. the mean squared distance of the units from the cluster centroid. However, this method requires the desired number k of clusters to be set in advance and can only work on numerical data. Hierarchical clustering methods, on the other hand, build a tree of clustering solutions, represented as a dendrogram, which highlights the relationships among the possible clusters. This is done either trough an agglomerative (bottom-up) strategy, where each unit starts in her own cluster and gets iteratively combined with other units and clusters, or in a divisive (top-down) strategy, where units start together in a single cluster which is then iteratively subdivided into smaller ones. This method allows for the number of cluster to be unspecified a priori, but both strategies become highly sensitive to the choice of the distance metric, responsible for all the merge and split operations. Density based methods as DBSCAN define instead a density metric, which can provide less constraints in the structure of the generated clusters, frequently resulting in weird shapes and less interpretable configurations. Moreover, the density metric is again sensitive to the choice of the density parameters. In summary, all these algorithmic approaches are largely heuristic and are suited for specific cases such for well-separated groups and standard geometric forms, while less for more complex scenarios of overlapping clusters. Moreover, lacking a solid statistical basis, can often lead to unsatisfying results and do not provide assessments about the clustering uncertainties.

An alternative and more flexible approach is therefore proposed by model-based methods, which assume a probabilistic modelling of the data. This is typically done

trough a mixture model, where each mixture component corresponds to a cluster with its specific cluster parameters [Wad23] [GMR22]. In this way, each observation is assumed to have arisen from one of J possible groups which are mixed together in various proportions. More formally, for each unit $i = 1, \dots, n$ we have that

$$f(y_i|\boldsymbol{\pi}, \boldsymbol{\vartheta}_1^\star, \dots, \boldsymbol{\vartheta}_J^\star) = \sum_{j=1}^{J} \pi_j f_j(y_j|\boldsymbol{\vartheta}_j^\star) \tag{1}$$

where y_i is the data point of the i-th observation, $\boldsymbol{\pi}$ is the set of weights satisfying $\pi_j \in [0, 1]$ for $j = 1, \dots, J$ and $\sum_{j=1}^{J} \pi_j = 1$, $\boldsymbol{\vartheta}_j^\star$ are the cluster-specific parameters, and $f_j(\cdot, |\vartheta_j^\star)$ is the density of the j-th cluster. A common modelling choice is a gaussian mixture model (GMM), where each cluster follows a normal distribution and thus making $\boldsymbol{\vartheta_j^\star} = (\mu_j^\star, \sigma_j^{2\star})$, or $\boldsymbol{\vartheta_j^\star} = (\boldsymbol{\mu}_j^\star, \Sigma_j^\star)$ in the multivariate case. This choice is flexible and effective since, especially in the multivariate case, gaussian distribution are able to capture very different clustering structures. Anyway, in this model-based approach, the goal is to estimate the cluster-specific parameters $\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_J$ and the mixing proportions $\pi_1, \dots, \pi_J$. The estimation step is often performed trough the Expectation-Maximization (EM) algorithm, which iteratively refines the estimates of the parameters via maximum likelihood estimation (MLE). Once the cluster-specific parameters are estimated, each observation can be assigned to a component, i.e. to a cluster, based on the highest posterior probability of belonging to that component, which can be computed trough the Bayes rule.

This approach of mixture model can be naturally moved in a Bayesian framework where to each parameter is treated as a random variable with a corresponding prior distribution [Fra08]. This leads (1) to be reformulated into

$$\begin{aligned} y_i|c_i, \boldsymbol{\pi}, \boldsymbol{\vartheta}_1^\star, \dots, \boldsymbol{\vartheta}_J^\star &\overset{\text{iid}}{\sim} f_{c_i}(y_i|\boldsymbol{\vartheta}_{c_i}^\star) \\ c_1, \dots, c_n &\sim \text{Cat}(\pi_1, \dots, \pi_J) \\ \boldsymbol{\vartheta}_1^\star, \dots, \boldsymbol{\vartheta}_J^\star &\overset{\text{iid}}{\sim} F_0 \\ \pi_1, \dots, \pi_J &\sim \text{Dir}(\gamma, \dots, \gamma) \end{aligned} \tag{2}$$

where the cluster labels $c_1, \dots, c_n$ are assigned a multinomial distribution with probabilities given by the vector of weights $\boldsymbol{\pi}$, the cluster-specific parameters $\boldsymbol{\vartheta}_j^\star$ are assigned a prior distribution F_0, while the weights are assigned a Dirichlet distribution, where e.g. all parameters are the same $\gamma_j = \gamma$ if no prior information about cluster assignments is wished to be inserted into the model.

The Bayesian framework is however much powerful and allows for even more complex formulations. In fact, moving to a Bayesian non-parametric approach, an infinite mixture model, i.e. that does not require a predetermined number J of components, can be introduced. This extension often relies on the Dirichlet Process (DP) [Fer73], which leads to a model formulation in the form

$$\begin{aligned} y_i|\boldsymbol{\vartheta}_i &\overset{\text{ind}}{\sim} f(y_i|\boldsymbol{\vartheta}_i) \\ \boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_n|P &\overset{\text{iid}}{\sim} P \\ P &\sim \text{discrete RPM} \end{aligned}$$

where RPM denotes a random probability measure. The discreteness of P implies the presence of ties among the atoms of the process $\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_n$. These ties therefore induce a partition ρ_n which can be identified by the units manifesting the same values of the parameter $\boldsymbol{\vartheta}_i$. That is, denoting with $\boldsymbol{\vartheta}_1^\star, \ldots, \boldsymbol{\vartheta}_K^\star$ the unique values of $\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_n$, the generic h-cluster can be defined as $S_h = \{i \in \{1, \ldots, n\} : \boldsymbol{\vartheta}_i = \boldsymbol{\vartheta}_h^\star\}$. The Dirichlet Process plays a significant role in general in Bayesian nonparametrics [Gra23], but especially in clustering, thanks to its computationally-friendly implementation trough the stick-breaking representation [Teh10], the Pólya urn representation [BM73], and the Chinese restaurant process (CRP) [Ald85].

This increased flexibility in the clustering modelling becomes even more powerful when working on spatio-temporal datasets, in which observations are collected over time and across different spatial locations, possibly concealing trends behind both information levels. This type of data, in fact, is inherently complex due to this interaction between spatial and temporal dimensions; a complexity that is further increased if covariates are also available. As a result, model-based methods are fairly more suited to tackle this task, rather than traditional algorithmic methods, since they can combine all these different levels of information. A model-based analysis of such data should also be able to account for the temporal dependency among the partitions, in order to extract possible trends occurring in time, resulting in more gentle and interpretable temporal evolution of the clusters, while also granting efficient implementations to be possibly applied on large scale datasets which are commonly accessible in this context.

Recently, the use of Bayesian models to perform clustering has become more popular, particularly in this field of spatio-temporal datasets. Bayesian clustering, in fact, allows to incorporate prior information into the model enhancing the flexibility and interpretability of the results with respect, for example, to more traditional frequentist approaches. Throughout the years, several models have been proposed in the Bayesian literature, but one of the most relevant is the Dependent Random Partition Model [PQD22] which stands out for being able to handle explicitly the temporal dependence of partitions into the model formulation, while also possibly accounting for the spatial information. However, the current DRPM implementation of the MCMC algorithm, written in C and available trough an R interface, lacks some relevant utilities such as the inclusion of covariates, which could further improve the generation and informativity of the clusters, the handling of missing data, and an efficient implementation, which would speed up the model fitting to run multiple chains in parallel, or be more easily applied on large scale datasets.

In this work, we aim to address these three issues by making the original model more flexible. We will show how our updates can indeed be effective and also provide faster execution times. In fact, implementing the model using the Julia language, rather than C, we took advantage of its high-performance capabilities and well-equipped statistical ecosystem. Our comparison will focus on both synthetic and real-world datasets, with the latter case involving air quality measurements from the AgrImOnIA project [Fas+23], a comprehensive record of air pollutant levels and other environmental variables measured across the Lombardy region of Italy.

Chapter 1 briefly reviews the literature on Bayesian clustering models for spatio-temporal data, and then dives deeply into the analysis and description of the DRPM and of our generalization.

Chapter 2 provides insights about the computational aspects of the MCMC algorithm, motivating the choice of the Julia language and reporting some optimization possibilities emerged when developing the implementation.

Chapter 3 focuses on testing the original DRPM formulation and our generalized model. We firstly evaluate if they perform similarly at a common testing level, i.e. under the same data, hyperparameters values, and MCMC iterations setup. Then, we assess the performances of our updates by considering fits involving missing data and fits including covariates. A comparative analysis about execution times with respect to the size of the dataset and the type of the fit will also be provided.

Finally, in Chapter 4, we briefly review the strengths and drawbacks this work revealed and suggest possible further improvements.

changes and insertions that we made.

$$
\begin{aligned}
Y_{it}|Y_{it-1}, \boldsymbol{\mu}^\star_t, \boldsymbol{\sigma}^{2\star}_t, \boldsymbol{\eta}, \boldsymbol{c}_t &\overset{\text{ind}}{\sim} \mathcal{N}(\mu^\star_{c_{it}t} + \eta_{1i} Y_{it-1} + \boldsymbol{x}_{it}^T \boldsymbol{\beta}_t, \sigma^{2\star}_{c_{it}t}(1-\eta^2_{1i})) \\
Y_{i1} &\overset{\text{ind}}{\sim} \mathcal{N}(\mu^\star_{c_{i1}1} + \boldsymbol{x}_{i1}^T \boldsymbol{\beta}_1, \sigma^{2\star}_{c_{i1}1}) \\
\boldsymbol{\beta}_t &\overset{\text{ind}}{\sim} \mathcal{N}_p(\boldsymbol{b}, s^2 I) \\
\xi_i = \text{Logit}(\tfrac{1}{2}(\eta_{1i}+1)) &\overset{\text{ind}}{\sim} \text{Laplace}(a, b) \\
(\mu^\star_{jt}, \sigma^{2\star}_{jt}) &\overset{\text{ind}}{\sim} \mathcal{N}(\vartheta_t, \tau^2_t) \times \text{invGamma}(a_\sigma, b_\sigma) \\
\vartheta_t|\vartheta_{t-1} &\overset{\text{ind}}{\sim} \mathcal{N}((1-\varphi_1)\varphi_0 + \varphi_1 \vartheta_{t-1}, \lambda^2(1-\varphi_1^2)) \\
(\vartheta_1, \tau^2_t) &\overset{\text{iid}}{\sim} \mathcal{N}(\varphi_0, \lambda^2) \times \text{invGamma}(a_\tau, b_\tau) \\
(\varphi_0, \varphi_1, \lambda^2) &\sim \mathcal{N}(m_0, s_0^2) \times \mathcal{U}(-1, 1) \times \text{invGamma}(a_\lambda, b_\lambda) \\
\{\boldsymbol{c}_t, \dots, \boldsymbol{c}_T\} \sim \text{tRPM}(\boldsymbol{\alpha}, M) &\text{ with } \alpha_t \overset{\text{iid}}{\sim} \text{Beta}(a_\alpha, b_\alpha)
\end{aligned} \tag{1.5}
$$

A visual representation of this new version of the DRPM model is also present in Figure 1.1, to more clearly appreciate the hierarchical structure and the relations among the parameters.

qui devo ancora sistemare, spostando il modello completo prima e spiegando le sue parti dopo come ha indicato

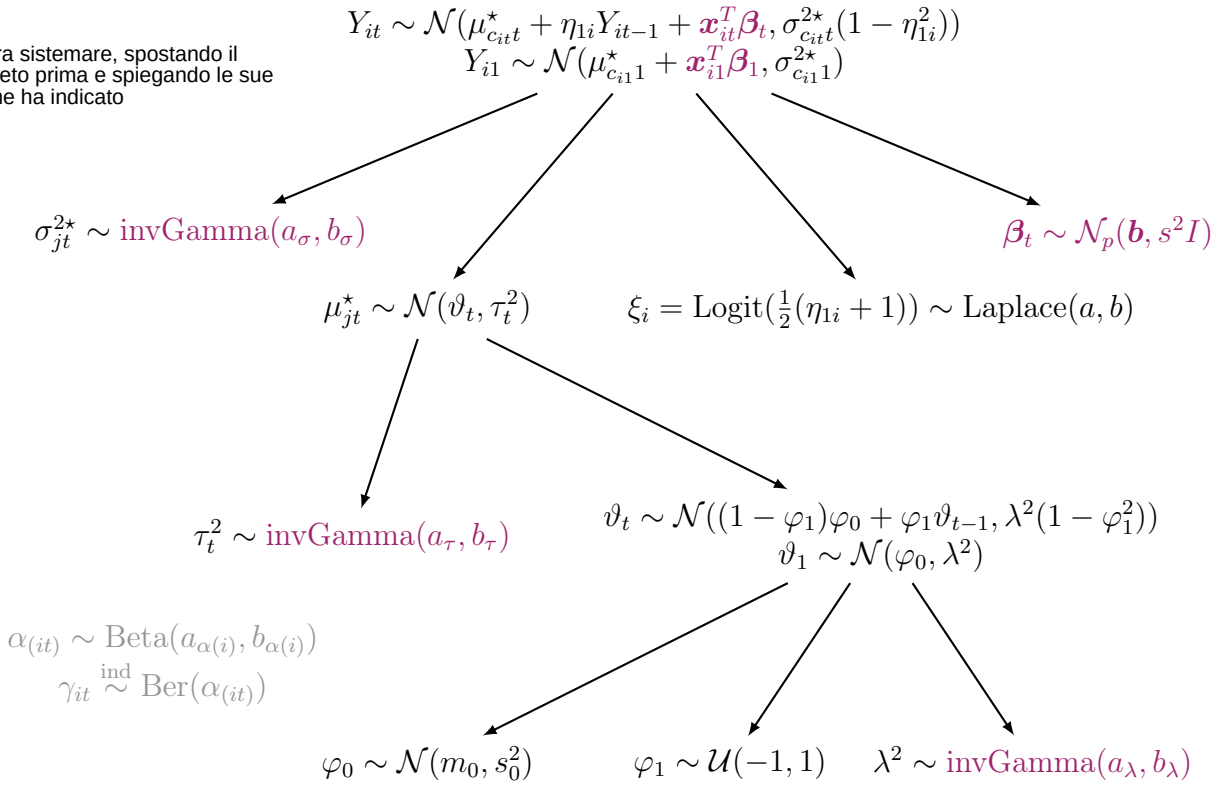

Figure 1.1: Graph visualization of the DRPM model, with highlighted in dark red the changes that we made to the original formulation and in gray the internal variables of the model.

In the course of this work, for the sake of clarity, we will refer to CDRPM for the original model formulation of [PQD22], and to JDRPM for our updated version. The starting letters C and J refer to the languages in which their corresponding MCMC algorithm has been implemented: C for the former, Julia for the latter.

We will now dive more deeply into the characteristics of our updated model by outlining the MCMC algorithm and, subsequently, inspecting the behaviours of the spatial cohesions and covariates similarities. This description refers to our generalization, the JDRPM model of (1.5), but we will remark analogies and differences with respect to the original CDRPM of (1.4).

1.1 MCMC sampling algorithm

We now report the full conditionals derivation for the parameters which had a conjugacy in the model (for the complete steps see Appendix A). Their computation is theoretically "simple", where we use the fact that *posterior* $\propto$ *likelihood* $\cdot$ *prior*, but followed useful suggestions and tricks from [Dun16]. The other variables not included here, namely η_{1i} and φ_1, involved instead the classical Metropolis update.

- update $\sigma^{2\star}_{jt}$. This full conditional derivation is characteristic of JDRPM only, since in CDRPM the variance had a uniform law and was therefore updated trough a Metropolis step.

  for $t = 1$: $f(\sigma^{2\star}_{jt}|-) \propto$ kernel of a invGamma$(a_{\sigma(\text{post})}, b_{\sigma(\text{post})})$ with

  $$a_{\tau(\text{post})} = a_\sigma + \frac{|S_{jt}|}{2} \quad b_{\tau(\text{post})} = b_\sigma + \frac{1}{2}\sum_{i\in S_{jt}} (Y_{it} - \mu^\star_{jt} - \boldsymbol{x}^T_{it}\boldsymbol{\beta}_t)^2$$

  for $t > 1$: $f(\sigma^{2\star}_{jt}|-) \propto$ kernel of a invGamma$(a_{\sigma(\text{post})}, b_{\sigma(\text{post})})$ with

  $$a_{\tau(\text{post})} = a_\sigma + \frac{|S_{jt}|}{2} \quad b_{\tau(\text{post})} = b_\sigma + \frac{1}{2}\sum_{i\in S_{jt}} (Y_{it} - \mu^\star_{jt} - \eta_{1i}Y_{it-1} - \boldsymbol{x}^T_{it}\boldsymbol{\beta}_t)^2$$

  (1.6)

- update $\mu^\star_{jt}$. This update rule is the same for both JDRPM and CDRPM.

  for $t = 1$: $f(\mu^\star_{jt}|-) \propto$ kernel of a $\mathcal{N}(\mu_{\mu^\star_{jt}(\text{post})}, \sigma^2_{\mu^\star_{jt}(\text{post})})$ with

  $$\sigma^2_{\mu^\star_{jt}(\text{post})} = \frac{1}{\frac{1}{\tau^2_t} + \frac{|S_{jt}|}{\sigma^{2\star}_{jt}}} \quad \mu_{\mu^\star_{jt}(\text{post})} = \sigma^2_{\mu^\star_{jt}(\text{post})}\left(\frac{\vartheta_t}{\tau^2_t} + \frac{\sum_{i\in S_{jt}}(Y_{i1} - \boldsymbol{x}^T_{it}\boldsymbol{\beta}_t)}{\sigma^{2\star}_{jt}}\right)$$

  for $t > 1$: $f(\mu^\star_{jt}|-) \propto$ kernel of a $\mathcal{N}(\mu_{\mu^\star_{jt}(\text{post})}, \sigma^2_{\mu^\star_{jt}(\text{post})})$ with

  $$\sigma^2_{\mu^\star_{jt}(\text{post})} = \frac{1}{\frac{1}{\tau^2_t} + \frac{\sum_{i\in S_{jt}}\frac{1}{1-\eta^2_{1i}}}{\sigma^{2\star}_{jt}}} \quad \mu_{\mu^\star_{jt}(\text{post})} = \sigma^2_{\mu^\star_{jt}(\text{post})}\left(\frac{\vartheta_t}{\tau^2_t} + \frac{\sum_{i\in S_{jt}}\frac{Y_{it} - \eta_{1i}Y_{i,t-1} - \boldsymbol{x}^T_{it}\boldsymbol{\beta}_t}{1-\eta^2_{1i}}}{\sigma^{2\star}_{jt}}\right)$$

  (1.7)

- update $\boldsymbol{\beta}_t$. This full conditional derivation is characteristic of JDRPM only, since the insertion of a regression term in the likelihood is a feature introduced by our generalized model.

  for $t = 1$: $f(\boldsymbol{\beta}_t|-) \propto$ kernel of a $\mathcal{N}(\boldsymbol{b}_{(\text{post})}, A_{(\text{post})})$ with

$$A_{(\text{post})} = \left(\frac{1}{s^2} I + \sum_{i=1}^{n} \frac{\boldsymbol{x}_{it} \boldsymbol{x}_{it}^T}{\sigma^{2\star}_{c_{it}t}} \right)^{-1} \quad \boldsymbol{b}_{(\text{post})} = A_{(\text{post})} \left(\frac{\boldsymbol{b}}{s^2} + \sum_{i=1}^{n} \frac{(Y_{it} - \mu^{\star}_{c_{it}t}) \boldsymbol{x}_{it}}{\sigma^{2\star}_{c_{it}t}} \right)$$

i.e. $f(\boldsymbol{\beta}_t|-) \propto$ kernel of a $\mathcal{N}\text{Canon}(\boldsymbol{h}_{(\text{post})}, J_{(\text{post})})$ with

$$\boldsymbol{h}_{(\text{post})} = \left(\frac{\boldsymbol{b}}{s^2} + \sum_{i=1}^{n} \frac{(Y_{it} - \mu^{\star}_{c_{it}t}) \boldsymbol{x}_{it}}{\sigma^{2\star}_{c_{it}t}} \right) \quad J_{(\text{post})} = \left(\frac{1}{s^2} I + \sum_{i=1}^{n} \frac{\boldsymbol{x}_{it} \boldsymbol{x}_{it}^T}{\sigma^{2\star}_{c_{it}t}} \right)$$

for $t > 1$: $f(\boldsymbol{\beta}_t|-) \propto$ kernel of a $\mathcal{N}(\boldsymbol{b}_{(\text{post})}, A_{(\text{post})})$ with

$$A_{(\text{post})} = \left(\frac{1}{s^2} I + \sum_{i=1}^{n} \frac{\boldsymbol{x}_{it} \boldsymbol{x}_{it}^T}{\sigma^{2\star}_{c_{it}t}} \right)^{-1} \quad \boldsymbol{b}_{(\text{post})} = A_{(\text{post})} \left(\frac{\boldsymbol{b}}{s^2} + \sum_{i=1}^{n} \frac{(Y_{it} - \mu^{\star}_{c_{it}t} - \eta_{1i} Y_{it-1}) \boldsymbol{x}_{it}}{\sigma^{2\star}_{c_{it}t}} \right)$$

i.e. $f(\boldsymbol{\beta}_t|-) \propto$ kernel of a $\mathcal{N}\text{Canon}(\boldsymbol{h}_{(\text{post})}, J_{(\text{post})})$ with

$$\boldsymbol{h}_{(\text{post})} = \left(\frac{\boldsymbol{b}}{s^2} + \sum_{i=1}^{n} \frac{(Y_{it} - \mu^{\star}_{c_{it}t} - \eta_{1i} Y_{it-1}) \boldsymbol{x}_{it}}{\sigma^{2\star}_{c_{it}t}} \right) \quad J_{(\text{post})} = \left(\frac{1}{s^2} I + \sum_{i=1}^{n} \frac{\boldsymbol{x}_{it} \boldsymbol{x}_{it}^T}{\sigma^{2\star}_{c_{it}t}} \right) \tag{1.8}$$

Here $\mathcal{N}\text{Canon}(\boldsymbol{h}, J)$ is the canonical formulation of the $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, with $\boldsymbol{h} = \Sigma^{-1}\boldsymbol{\mu}$ and $J = \Sigma^{-1}$. This other distribution facilitates the sampling, since these full conditional computations allow to derive directly the parameters of the canonical one, e.g. the inverse of the variance matrix rather than the variance matrix itself; and therefore sampling trough it does not require any inversion of matrices which would produce more computational load, numerical instabilities, and loss of accuracy. As a consequence, in Julia we can write `rand(MvNormalCanon(h_star, J_star))` rather than the riskier one `rand(MvNormal(inv(J_star)*h_star, inv(J_star)))`; which apart from the previously mentioned disadvantages would be a statistically equivalent form.

- update τ_t^2. This full conditional derivation is characteristic of JDRPM only, since in CDRPM the variance had a uniform law and was therefore updated trough a Metropolis step.

$$f(\tau_t^2|-) \propto \text{kernel of a invGamma}(a_{\tau(\text{post})}, b_{\tau(\text{post})}) \text{ with}$$
$$a_{\tau(\text{post})} = \frac{k_t}{2} + a_\tau \quad b_{\tau(\text{post})} = \frac{\sum_{j=1}^{k_t} (\mu^{\star}_{jt} - \vartheta_t)^2}{2} + b_\tau \tag{1.9}$$

- update ϑ_t. This update rule is the same for both JDRPM and CDRPM.

for $t = T$: $f(\vartheta_t|-) \propto$ kernel of a $\mathcal{N}(\mu_{\vartheta_t(\text{post})}, \sigma^2_{\vartheta_t(\text{post})})$ with

$$\sigma^2_{\vartheta_t(\text{post})} = \frac{1}{\frac{1}{\lambda^2(1-\varphi_1^2)} + \frac{k_t}{\tau_t^2}}$$
$$\mu_{\vartheta_t(\text{post})} = \sigma^2_{\vartheta_t(\text{post})} \left(\frac{\sum_{j=1}^{k_t} \mu^{\star}_{jt}}{\tau_t^2} + \frac{(1-\varphi_1)\varphi_0 + \varphi_1 \vartheta_{t-1}}{\lambda^2(1-\varphi_1^2)} \right)$$

for $1 < t < T$: $f(\vartheta_t|-) \propto$ kernel of a $\mathcal{N}(\mu_{\vartheta_t(\text{post})}, \sigma^2_{\vartheta_t(\text{post})})$ with

$$\sigma^2_{\vartheta_t(\text{post})} = \frac{1}{\frac{1+\varphi_1^2}{\lambda^2(1-\varphi_1^2)} + \frac{k_t}{\tau_t^2}}$$

$$\mu_{\vartheta_t(\text{post})} = \sigma^2_{\vartheta_t(\text{post})} \left(\frac{\sum_{j=1}^{k_t} \mu^\star_{jt}}{\tau_t^2} + \frac{\varphi_1(\vartheta_{t-1} + \vartheta_{t+1}) + \varphi_0(1-\varphi_1)^2}{\lambda^2(1-\varphi_1^2)} \right)$$

for $t = 1$: $f(\vartheta_t|-) \propto$ kernel of a $\mathcal{N}(\mu_{\vartheta_t(\text{post})}, \sigma^2_{\vartheta_t(\text{post})})$ with

$$\sigma^2_{\vartheta_t(\text{post})} = \frac{1}{\frac{1}{\lambda^2} + \frac{\varphi_1^2}{\lambda^2(1-\varphi_1^2)} + \frac{k_t}{\tau_t^2}}$$

$$\mu_{\vartheta_t(\text{post})} = \sigma^2_{\vartheta_t(\text{post})} \left(\frac{\varphi_0}{\lambda^2} + \frac{\varphi_1(\vartheta_{t+1} - (1-\varphi_1)\varphi_0)}{\lambda^2(1-\varphi_1^2)} + \frac{\sum_{j=1}^{k_t} \mu^\star_{jt}}{\tau_t^2} \right) \tag{1.10}$$

- update φ_0. This update rule is also the same for both JDRPM and CDRPM.

$$f(\varphi_0|-) \propto \text{kernel of a } \mathcal{N}(\mu_{\varphi_0(\text{post})}, \sigma^2_{\varphi_0(\text{post})}) \text{ with}$$

$$\sigma^2_{\varphi_0(\text{post})} = \frac{1}{\frac{1}{s_0^2} + \frac{1}{\lambda^2} + \frac{(T-1)(1-\varphi_1)^2}{\lambda^2(1-\varphi_1^2)}}$$

$$\mu_{\varphi_0(\text{post})} = \sigma^2_{\varphi_0(\text{post})} \left(\frac{m_0}{s_0^2} + \frac{\vartheta_1}{\lambda^2} + \frac{1-\varphi_1}{\lambda^2(1-\varphi_1^2)} \sum_{t=2}^{T} (\vartheta_t - \varphi_1 \vartheta_{t-1}) \right) \tag{1.11}$$

- update λ^2. This full conditional derivation is characteristic of JDRPM only, since in CDRPM the variance had a uniform law and was therefore updated trough a Metropolis step.

$$f(\lambda^2|-) \propto \text{kernel of a invGamma}(a_{\lambda(\text{post})}, b_{\lambda(\text{post})}) \text{ with}$$

$$a_{\lambda(\text{post})} = \frac{T}{2} + a_\lambda$$

$$b_{\lambda(\text{post})} = \frac{(\vartheta_1 - \varphi_0)^2}{2} + \sum_{t=2}^{T} \frac{(\vartheta_t - (1-\varphi_1)\varphi_0 - \varphi_1\vartheta_{t-1})^2}{2} + b_\lambda \tag{1.12}$$

- update α. This update rule is the same for both JDRPM and CDRPM.

if global α: $f(\alpha|-) \propto$ kernel of a $\text{Beta}(a_{\alpha(\text{post})}, b_{\alpha(\text{post})})$ with

$$a_{\alpha(\text{post})} = a_\alpha + \sum_{i=1}^{n}\sum_{t=1}^{T} \gamma_{it} \quad b_{\alpha(\text{post})} = b_\alpha + nT - \sum_{i=1}^{n}\sum_{t=1}^{T} \gamma_{it}$$

if time specific α: $f(\alpha_t|-) \propto$ kernel of a $\text{Beta}(a_{\alpha(\text{post})}, b_{\alpha(\text{post})})$ with

$$a_{\alpha(\text{post})} = a_\alpha + \sum_{i=1}^{n} \gamma_{it} \quad b_{\alpha(\text{post})} = b_\alpha + n - \sum_{i=1}^{n} \gamma_{it}$$

if unit specific α: $f(\alpha_i|-) \propto$ kernel of a $\text{Beta}(a_{\alpha(\text{post})}, b_{\alpha(\text{post})})$ with

$$a_{\alpha(\text{post})} = a_{\alpha i} + \sum_{t=1}^{T} \gamma_{it} \quad b_{\alpha(\text{post})} = b_{\alpha i} + T - \sum_{t=1}^{T} \gamma_{it}$$

if time and unit specific α: $f(\alpha_{it}|-) \propto$ kernel of a $\text{Beta}(a_{\alpha(\text{post})}, b_{\alpha(\text{post})})$ with

$$a_{\alpha(\text{post})} = a_{\alpha i} + \gamma_{it} \quad b_{\alpha(\text{post})} = b_{\alpha i} + 1 - \gamma_{it} \tag{1.13}$$

- update a missing observation Y_{it}. This full conditional derivation is characteristic of JDRPM only, since the handling of missing data feature introduced by our generalized model.

per la Y_{it} ci sono tutti i calcoli completi, che aveva richiesto nei suoi commenti, in appendice (come per il resto delle variabili)

$$\text{for } t = 1\text{: } f(Y_{it}|-) \propto \text{kernel of a } \mathcal{N}(\mu_{Y_{it}(\text{post})}, \sigma^2_{Y_{it}(\text{post})}) \text{ with}$$

$$\sigma^2_{Y_{it}(\text{post})} = \frac{1}{\frac{1}{\sigma^{2\star}_{c_{it}t}} + \frac{\eta^2_{1i}}{2\sigma^{2\star}_{c_{it+1}t+1}(1-\eta^2_{1i})}}$$

$$\mu_{Y_{it}(\text{post})} = \sigma^2_{Y_{it}(\text{post})} \left(\frac{\mu^\star_{c_{it}t} + \boldsymbol{x}^T_{it}\boldsymbol{\beta}_t}{\sigma^{2\star}_{c_{it}t}} + \frac{\eta_{1i}(Y_{it+1} - \mu^\star_{c_{it+1}t+1} - \boldsymbol{x}^T_{it+1}\boldsymbol{\beta}_{t+1})}{\sigma^{2\star}_{c_{it+1}t+1}(1-\eta^2_{1i})} \right)$$

$$\text{for } 1 < t < T\text{: } f(Y_{it}|-) \propto \text{kernel of a } \mathcal{N}(\mu_{Y_{it}(\text{post})}, \sigma^2_{Y_{it}(\text{post})}) \text{ with}$$

$$\sigma^2_{Y_{it}(\text{post})} = \frac{1-\eta^2_{1i}}{\frac{1}{\sigma^{2\star}_{c_{it}t}} + \frac{\eta^2_{1i}}{\sigma^{2\star}_{c_{it+1}t+1}}}$$

$$\mu_{Y_{it}(\text{post})} = \sigma^2_{Y_{it}(\text{post})} \left(\frac{\mu^\star_{c_{it}t} + \eta_{1i}Y_{it-1} + \boldsymbol{x}^T_{it}\boldsymbol{\beta}_t}{\sigma^{2\star}_{c_{it}t}(1-\eta^2_{1i})} + \frac{\eta_{1i}(Y_{it+1} - \mu^\star_{c_{it+1}t+1} - \boldsymbol{x}^T_{it+1}\boldsymbol{\beta}_{t+1})}{\sigma^{2\star}_{c_{it+1}t+1}(1-\eta^2_{1i})} \right)$$

$$\text{for } t = T\text{: } f(Y_{it}|-) \text{ is just the likelihood of } Y_{it} \tag{1.14}$$

Finally, we briefly highlight in Algorithm 1 the steps which compose the MCMC sampling algorithm. Regarding the computation of the LPML and WAIC metrics, they follow classic ideas from [Chr+10] and [GHV13].

The core of the clustering process happens in the updating steps of γ_{it} and ρ_t. Their update step is indeed quite complex, and as we said before involves the check of compatibility issues. In any case, the general idea is that, for each unit i currently belonging to cluster j, we simulate to assign her to one of the existing clusters, plus to a new singleton cluster, and compute for each case the likelihood of this to happen, deriving probability weights to finally sample the decision for the next iteration. The key elements participating into the definition of such weights are the spatial cohesions and, with the JDRPM update, also the covariate similarities, which we will now both investigate.

1.2 Spatial cohesions analysis

The clustering procedure revolves around the product partition model (PPM) [Har90] [BH93] [Cro97]. The simplest idea is to set $P(\rho_t) \propto \prod_{j=1}^{k_t} C(S_{jt})$, with the function $C(S_{jt})$ that measures how tightly grouped the elements in cluster S_{jt} are considered to be. Then, to include spatial information, the idea is to extend the PPM from being a function of just $C(S_{jt})$ to the more informed one $C(S_{jt}, \boldsymbol{s}^\star_{jt})$, where S_{jt} is the j-th cluster at time instant t and $\boldsymbol{s}^\star_{jt}$ is the subset of spatial coordinates of the units inside S_{jt}. For the sake of clarity, in this section where we are just interested in analysing the cohesions we employ the S_h notation to indicate a general h-th cluster, rather than the more pedantic S_{jt}.

Regarding the computation of spatial cohesion, several choices are available [PQ15]. The main common idea of the following formulas is to favour few spatially

Bibliography

[PQD22] Garrit L. Page, Fernando A. Quintana, and David B. Dahl. "Dependent Modeling of Temporal Sequences of Random Partitions". In: *Journal of Computational and Graphical Statistics* 31.2 (2022), pp. 614–627. DOI: 10.1080/10618600.2021.1987255 (cit. on pp. v, vii, 3, 5, 8, 31, 32).

[Fas+23] A. Fassò et al. *AgrImOnIA: Open Access dataset correlating livestock and air quality in the Lombardy region, Italy (3.0.0)*. 2023. DOI: https://doi.org/10.5281/zenodo.7956006 (cit. on pp. v, vii, 3, 33).

[Wad23] Sara Wade. "Bayesian cluster analysis". In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 381 (Mar. 2023), p. 20220149. DOI: 10.1098/rsta.2022.0149 (cit. on p. 2).

[GMR22] Isobel Gormley, Thomas Murphy, and Adrian Raftery. "Model-Based Clustering". In: *Annual Review of Statistics and Its Application* 10 (Oct. 2022). DOI: 10.1146/annurev-statistics-033121-115326 (cit. on p. 2).

[Fra08] Jessica Franzén. "Bayesian Cluster Analysis : Some Extensions to Non-standard Situations". In: 2008. URL: https://api.semanticscholar.org/CorpusID:12397258 (cit. on p. 2).

[Fer73] Thomas S. Ferguson. "A Bayesian Analysis of Some Nonparametric Problems". In: *The Annals of Statistics* 1.2 (1973), pp. 209–230. DOI: 10.1214/aos/1176342360. URL: https://doi.org/10.1214/aos/1176342360 (cit. on p. 2).

[Gra23] Clara Grazian. *A review on Bayesian model-based clustering*. Mar. 2023. DOI: 10.48550/arXiv.2303.17182 (cit. on p. 3).

[Teh10] Yee Whye Teh. "Dirichlet Process". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 280–287. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_219 (cit. on p. 3).

[BM73] David Blackwell and James B. MacQueen. "Ferguson Distributions Via Polya Urn Schemes". In: *The Annals of Statistics* 1.2 (1973), pp. 353–355. DOI: 10.1214/aos/1176342372 (cit. on p. 3).

[Ald85] David J. Aldous. "Exchangeability and related topics". In: *École d'Été de Probabilités de Saint-Flour XIII — 1983*. Ed. by P. L. Hennequin. Berlin, Heidelberg: Springer Berlin Heidelberg, 1985, pp. 1–198. ISBN: 978-3-540-39316-0 (cit. on p. 3).

[Bar+12] Luis E Nieto Barajas et al. "A Time-Series DDP for Functional Proteomics Profiles". In: *Biometrics* 68.3 (Jan. 2012), pp. 859–868. DOI: 10.1111/j.1541-0420.2011.01724.x (cit. on p. 5).

[AW15] Isadora Antoniano Villalobos and Stephen Walker. "A Nonparametric Model for Stationary Time Series". In: *Journal of Time Series Analysis* 63 (Aug. 2015). DOI: 10.1111/jtsa.12146 (cit. on p. 5).

[GMR16] Luis Gutiérrez, Ramsés H. Mena, and Matteo Ruggiero. "A time dependent Bayesian nonparametric model for air quality analysisa". In: *Computational Statistics & Data Analysis* 95.C (2016), pp. 161–175. DOI: 10.1016/j.csda.2015.10.00. URL: https://ideas.repec.org/a/eee/csdana/v95y2016icp161-175.html (cit. on p. 5).

[Jo+16] Seongil Jo et al. "Dependent Species Sampling Models for Spatial Density Estimation". In: *Bayesian Analysis* 12 (May 2016). DOI: 10.1214/16-BA1006 (cit. on p. 5).

[KG18] Maria Kalli and Jim Griffin. "Bayesian nonparametric vector autoregressive models". In: *Journal of Econometrics* 203 (Jan. 2018). DOI: 10.1016/j.jeconom.2017.11.009 (cit. on p. 5).

[DK18] Maria De Iorio and Athanasios Kottas. "Modeling for Dynamic Ordinal Regression Relationships: An Application to Estimating Maturity of Rockfish in California". In: *Journal of the American Statistical Association* 113.521 (2018), pp. 68–80. DOI: 10.1080/01621459.2017.1328357 (cit. on p. 5).

[De +19] Maria De Iorio et al. *Bayesian nonparametric temporal dynamic clustering via autoregressive Dirichlet priors*. Oct. 2019. DOI: 10.48550/arXiv.1910.10443 (cit. on p. 5).

[Car+17] François Caron et al. "Generalized Pólya urn for time-varying Pitman-Yor processes". In: *Journal of Machine Learning Research* 18.1 (Jan. 2017), pp. 836–867. ISSN: 1532-4435 (cit. on p. 5).

[Dun16] Earl Duncan. *Deriving the Full Conditionals*. BRAG: Bayesian Research & Application Group. 2016. URL: https://bragqut.wordpress.com/wp-content/uploads/2018/04/deriving-the-full-conditionals.pdf (cit. on p. 9).

[Chr+10] Ronald Christensen et al. *Bayesian ideas and data analysis. An introduction for scientists and statisticians*. CRC Press, 2010. DOI: 10.1201/9781439894798 (cit. on p. 12).

[GHV13] Andrew Gelman, Jessica Hwang, and Aki Vehtari. "Understanding predictive information criteria for Bayesian models". In: *Statistics and Computing* 24 (July 2013). DOI: 10.1007/s11222-013-9416-2 (cit. on p. 12).

[Har90] J.A. Hartigan. "Partition models". In: *Communications in Statistics - Theory and Methods* 19.8 (1990), pp. 2745–2756. DOI: 10.1080/03610929008830345 (cit. on p. 12).

[BH93] Daniel Barry and J. A. Hartigan. "A Bayesian Analysis for Change Point Problems". In: *Journal of the American Statistical Association* 88.421 (1993), pp. 309–319. ISSN: 01621459, 1537274X. URL: http://www.jstor.org/stable/2290726 (cit. on p. 12).

[Cro97] Evelyn M. Crowley. "Product Partition Models for Normal Means". In: *Journal of the American Statistical Association* 92.437 (1997), pp. 192–198. ISSN: 01621459. URL: http://www.jstor.org/stable/2291463 (cit. on p. 12).

[PQ15] Garritt Page and Fernando Quintana. "Spatial Product Partition Models". In: *Bayesian Analysis* 11 (Apr. 2015). DOI: 10.1214/15-BA971 (cit. on p. 12).

[DH01] D. Denison and C Holmes. "Bayesian Partitioning for Estimating Disease Risk". In: *Biometrics* 57 (Apr. 2001), pp. 143–9. DOI: 10.1111/j.0006-341X.2001.00143.x (cit. on p. 14).

[De +15] Pierpaolo De Blasi et al. "Are Gibbs-Type Priors the Most Natural Generalization of the Dirichlet Process?" In: *IEEE Trans. Pattern Anal. Mach. Intell.* 37.2 (Feb. 2015), pp. 212–229. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2013.217 (cit. on p. 15).

[MQR11] Peter Müller, Fernando Quintana, and Gary Rosner. "A Product Partition Model With Regression on Covariates". In: *Journal of computational and graphical statistics: a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America* 20 (Mar. 2011), pp. 260–278. DOI: 10.1198/jcgs.2011.09066 (cit. on p. 15).

[QMP15] Fernando Quintana, Peter Müller, and Ana Luisa Papoila. "Cluster-Specific Variable Selection for Product Partition Models". In: *Scandinavian Journal of Statistics* 42 (Mar. 2015). DOI: 10.1111/sjos.12151 (cit. on p. 15).

[PQ18] Garritt Page and Fernando Quintana. "Calibrating covariate informed product partition models". In: *Statistics and Computing* 28 (Sept. 2018), pp. 1–23. DOI: 10.1007/s11222-017-9777-z (cit. on p. 18).

[Gow71] J. C. Gower. "A General Coefficient of Similarity and Some of Its Properties". In: *Biometrics* 27.4 (1971), pp. 857–871. ISSN: 0006341X, 15410420. URL: http://www.jstor.org/stable/2528823 (cit. on p. 19).

[Bez+17] Jeff Bezanson et al. "Julia: A fresh approach to numerical computing". In: *SIAM Review* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671 (cit. on p. 23).

[Law+79] C. L. Lawson et al. "Basic Linear Algebra Subprograms for Fortran Usage". In: *ACM Trans. Math. Softw.* 5.3 (Sept. 1979), pp. 308–323. ISSN: 0098-3500. DOI: 10.1145/355841.355847 (cit. on p. 23).

[Bes+21] Mathieu Besançon et al. "Distributions.jl: Definition and Modeling of Probability Distributions in the JuliaStats Ecosystem". In: *Journal of Statistical Software* 98.16 (2021), pp. 1–30. ISSN: 1548-7660. DOI: 10.18637/jss.v098.i16 (cit. on p. 23).

[Lin+19] Dahua Lin et al. *JuliaStats/Distributions.jl: a Julia package for probability distributions and associated functions*. July 2019. DOI: 10.5281/zenodo.2647458 (cit. on p. 23).

[CR16] Jiahao Chen and Jarrett Revels. "Robust benchmarking in noisy environments". In: *arXiv e-prints*, arXiv:1608.04295 (Aug. 2016). arXiv: 1608.04295 [cs.PF] (cit. on p. 25).

[HA85] Lawrence J. Hubert and Phipps Arabie. "Comparing partitions". In: *Journal of Classification* 2 (1985), pp. 193–218. URL: https://api.semanticscholar.org/CorpusID:189915041 (cit. on p. 31).

[R C24] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2024. URL: https://www.R-project.org/ (cit. on p. 31).

[LHB22] Stefan Lenz, Maren Hackenberg, and Harald Binder. "The JuliaConnectoR: A Functionally-Oriented Interface for Integrating Julia in R". In: *Journal of Statistical Software* 101.6 (2022), pp. 1–24. DOI: 10.18637/jss.v101.i06 (cit. on p. 31).

[Gel04] Andrew Gelman. "Prior Distributions for Variance Parameters in Hierarchical Models". In: *Economics and Econometrics Research Institute (EERI), EERI Research Paper Series* 1 (Jan. 2004) (cit. on p. 62).

[GXG18] Hong Ge, Kai Xu, and Zoubin Ghahramani. "Turing: a language for flexible probabilistic inference". In: *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*. 2018, pp. 1682–1690. URL: http://proceedings.mlr.press/v84/ge18b.html (cit. on p. 90).

Acknowledgements

"This is to be my haven for many long years, my niche which I enter with such a mistrustful, such a painful sensation... And who knows? Maybe when I come to leave it many years hence I may regret it!"

— Fëdor Dostoevskij, *The House of the Dead*

I would like to thank my advisors Alessandra Guglielmi and Alessandro Carminati, primarily for having made me feel, during the entire course of the thesis, in a very calm, lovely atmosphere, in which I have been able to feel perfectly at ease (which is notoriously a NP-hard problem). The epigraphs of my beloved author Dostoevsky and the title inspired by the Star Wars films, as well as some poetic licenses and easter eggs scattered across the chapters (which they have kindly allowed me to keep) I think prove this feeling of sweet and respectful confidence.

I thank professor Alessandra Guglielmi for having guided me through the fog of uncertainty and puzzlement that often the thesis brings, or in general the final period of university. From the beginning she shared a genuine interest in the work that we would have done and, I believe, an overall confidence in my chances of really carrying it out.

I thank Alessandro Carminati for having assisted me in many of the most critical phases of the thesis. His action always precise and effective has been necessary to solve the various theoretical puddles in which I got stuck. Moreover, we share the same passion for Julia and, I believe, the delight for having overthrown the not-so-missed C of the old model implementation.

I thank my family for having helped me, in different ways, during these years of university, and my friends and fellows who have always and wisely brought to light my highest potential and resources. The passion for solving problems, inventing creative solutions, getting lost in calculations, wandering trough the magical world of the most advanced mathematics, sharing doubts and reasonings: all this has always found in you a wonderful and unforeseen companion.

I also thank all the boys and girls with whom I have had the honour to work and play in the various summer camps, for steadily reminding me of the cheerful and soft spirit with which all the various challenges can be faced.

Ringraziamenti

> *"Ecco il mio ponte d'approdo per molti lunghi anni, il mio angoletto, nel quale faccio il mio ingresso con una sensazione così diffidente, così morbosa... Ma chi lo sa? Forse, quando tra molti anni mi toccherà abbandonarlo, magari potrei anche rimpiangerlo!"*
>
> — Fëdor Dostoevskij, *Memorie da una casa di morti*

Vorrei ringraziare innanzitutto i miei relatori Alessandra Guglielmi e Alessandro Carminati, principalmente per avermi fatto sentire, durante l'intero svolgimento della tesi, in un clima molto tranquillo, sereno, in cui ho avuto modo di trovarmi pienamente a mio agio. Le epigrafi del mio caro autore Dostoevskij e il titolo ispirato alla saga di Star Wars, nonché alcune licenze poetiche sparse nei vari capitoli (che loro mi hanno gentilmente concesso di tenere) credo dimostrino questa atmosfera di piacevole ma rispettosa confidenza.

Ringrazio la professoressa Alessandra Guglielmi per avermi guidato attraverso la nebbia di incertezza e confusione che spesso accompagna la tesi, o in senso lato gli ultimi mesi di università. Fin dall'inizio ha infatti condiviso un sincero interesse per il lavoro che avremmo dovuto svolgere, e una credo generale fiducia nelle mie possibilità di condurlo a termine.

Ringrazio Alessandro Carminati per avermi aiutato in molte delle fasi più critiche della tesi. Il suo intervento sempre preciso e puntuale è stato necessario per risolvere le varie pozzanghere teoriche in cui mi ero impantanato. Inoltre, condividiamo la stessa passione per Julia e, credo, la soddisfazione per aver battuto il non-così-compianto C della vecchia implementazione del modello.

Ringrazio la mia famiglia per avermi aiutato, in vari modi, durante questi anni di università, e i miei amici e compagni di studio, i quali hanno sempre sapientemente fatto emergere le mie migliori potenzialità e risorse. La passione per risolvere problemi, ideare soluzioni creative, perdersi nei calcoli, addentrarsi nel magico mondo della matematica più avanzata, condividere dubbi e ragionamenti: tutto questo ha sempre trovato in voi un'ottima e inaspettata compagnia.

Ringrazio anche tutti i bambini e ragazzi con cui ho avuto modo di lavorare e giocare nei vari campi estivi, per farmi sempre ricordare dello spirito gioioso e leggero con cui si possono approcciare tutte le varie sfide.