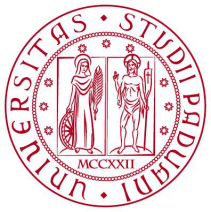


UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXVIII

NONPARAMETRIC MODELS FOR DEPENDENT FUNCTIONAL DATA

Direttore della Scuola: Ch.ma Prof.ssa. MONICA CHIOGNA

Supervisore: Prof. BRUNO SCARPA

Co-supervisore: Prof. MICHELE GUINDANI

Dottorando: RONALDO ROUVHER GUEDES SILVA

31 Gennaio 2016

Acknowledgements

This research was supported by the scholarship number 245612/2012-2 from CNPq/CAPES Brazilian Science Without Borders Program. Thank you Brazil.

In real life, unlike MCMC algorithms, different choices of starting points may lead you to have different results. As I am happy with my results after the end of this Ph.D. I can tell that choose to work with Bruno Scarpa, Antonio Canale, Michele Guindani, Marina Vannucci and Francesco Versace were good choices and definitely contributed to my good starting point. Special thanks to Monica Chiogna that was always there to give me support since the very first e-mail sent to the Ph.D. school. Beyond the scientific community, my friend Tahir played a great role to make possible the Ph.D. dream.

The real life beyond this Ph.D. existed and was enjoyable thanks to Giulio, Roberta, Giovanni, Daniele, Arianna, Giulia, Elena and people I got know through them.

My family, yes, they forgave me for spending such long time far from home.

Sommario

Nel contesto dell'analisi di dati funzionali, in questa tesi, vengono presentati due modelli Bayesiani non parametrici. Il primo modello è motivato da un problema di analisi di neuroimmagini, e considera funzioni correlate spazialmente e raggruppate seguendo un processo di Dirichlet dipendente funzionale (Functional Dependent Dirichlet process) che include una struttura di dipendenza autoregressiva condizionata per modellare la selezione spaziale. Un tale modello permette di considerare in maniera appropriata simmetrie spaziali delle risposte funzionali nel cervello. Il secondo modello è invece motivato dalla piattaforma italiana di bilanciamento del mercato del gas naturale e include una dipendenza temporale tra funzioni attraverso i pesi di un processo di Dirichlet dipendente funzionale basato su un modello lineare dinamico definito su una partizione dello spazio funzionale. Forme caratteristiche tipiche delle funzioni vengono modellate da curve flessibili basate su splines che formano gli atomi del processo di Dirichlet. In entrambe le applicazioni vengono usate tecniche bayesiane di selezione di variabili per scegliere le funzioni di base per le splines in ciascun cluster. Algoritmi di tipo Gibbs sampling sono sviluppati per il calcolo delle distribuzioni a posteriori. Vengono proposti studi di simulazione e applicazioni a dati reali per verificare l'appropriatezza degli approcci proposti.

Abstract

In the framework of functional data analysis we propose two Bayesian Nonparametric models. In the first model, motivated by an application in neuroimaging, functions are assumed to be spatially correlated and clustered together by an underlying Functional Dependent Dirichlet process which encodes a conditional autoregressive dependence structure to guide the spatial selection. Spatial symmetries of the functional responses in the brain can be appropriately accounted for in our framework. Motivated by the Italian natural gas balancing platform, in the second model time dependence are induced in the weights of the underlying Functional Dependent Dirichlet process through a dynamic linear model defined over a partitioned function space. Typical shape characteristics of the functions are modeled by flexible spline-based curve estimates as atoms of the process. In both applications Bayesian variable selection techniques are used to select significant sets of bases coefficients in each cluster. Gibbs sampling algorithms are developed for posterior computation, simulation studies and application to real data assess the performance of our approaches.

Table of contents

List of Algorithms

List of figures

Chapter 1

Introduction

1.1 Overview

In many applications the sample space Ω where the random variables take its values is a set of functions. Often the mathematical properties of the functions (curves) represent information of the features of the statistical units and prior knowledge informs of their possible temporal or spatial dependence. As examples, in the Event-Related Potentials (ERP) problem (Section 2.1) statistical units are waveforms presenting symmetries and spatial dependence; in the Italian Natural Gas Balancing Platform problem (Section 2.2) statistical units are time dependent monotonic bounded curves. In both motivating applications as the sample elements are curves our framework is within Functional Data Analysis (FDA). In this branch of statistics the dataset is called functional dataset, the set where its elements are functions, which is a suitable way to collect information of phenomena that actually are described by curves. For a introduction to FDA refer to Section 2.3.

ERPs represent a safe and non-invasive approach to directly measure brain activity, and constitute an ideal tool to investigate the neural correlates of cognitive and emotional processes in humans. ERPs are voltage fluctuations in the electroencephalogram (EEG) that are time-locked to the presentation of external events (e.g., pictures). Thanks to their high temporal resolution (in the order of the milliseconds), ERPs allow researchers to investigate fast-changing brain processes and constitute a sensitive and reliable biomarker of motivational and cognitive processing (Kenemans and Kähkönen, 2001; Lang and Bradley, 2010; Luck et al., 2011).

Versace et al. (2015) in an experiment with emotional and cigarette-related pictures identified that smokers with blunted brain responses to pleasant stimuli had lower rates of long-term smoking abstinence constituting a new biomarker to identify smokers at higher risk of relapse; used statistical methods include cluster analysis using k-means methods based on

time-windows placed on the amplitude of the late positive potential and two-way ANOVA to analyze between cluster difference. Versace et al. (2011) obtained results, in experiment with smokers before the beginning of a smoking cessation program, to support the hypothesis that cigarette-related cues are motivationally relevant stimuli that active brain circuits as those of that are normally used in the processing of intrinsically emotional stimuli. Randomized tests on time regions of interest, identified by temporal principal component analysis, was used to compare the differences ERP waveforms recorded during the presentation of pictures with pleasant, unpleasant, neutral and cigarette-related content.

Besides the frequent use of ANOVA on mean activity in specific time-windows others statistical tools in the analysis of ERP data include, for example, filtering methods to remove noise in the data from eye blink, analysis of covariance and correlation analysis. Moreover, Maris (2012) in his paper review evaluated the use of Neyman-Pearson, permutation-based, bootstrap-based and Bayesian approach to the statistical testing of electrophysiological data. Groppe et al. (2011) instead review approach on mass univariate analyses consisting of thousands of statistical tests and powerful corrections for multiple comparisons such as strong control of the family-wise error rate via permutation tests (FWER), weak control of FWER via cluster-based permutation tests, false discovery rate control, and control of the generalized FWER.

Modern EEG amplifiers allow for the simultaneous recording of ERPs from hundreds of electrodes. Although these high-density electrode arrays can improve ERPs' spatial resolution (Junghöfer et al., 2006), researchers not interested in uncovering the ERPs' brain sources often average signals (Jackson and Bolger, 2014) from neighboring electrodes. Indeed, the spatial smearing of the EEG signals justifies the computation of average voltages across electrodes, e.g., Bloom et al. (2013) compared within the dopaminergic system the P2a and P3 components during an attention selection cue exposure task using cigarette-related and neutral images. Statistical tools to their comparison constitute visual inspection of plots, ANOVA and reduction of the ERP data using a Region of Interest (ROI) method by averaging voltage across the electrodes in a ROI.

The previous issue is also important when conducting multiple hypotheses testing of brain activity based on the ERP signals, since different topographies and groupings of neighboring electrodes typically can lead to the computation of significantly different cluster-based statistics and hence possibly to conflicting inferences on the channels' activations. For a review on randomization tests for comparison of topographies in ERP studies refer to Maris (2004).

Within FDA framework Meyer et al. (2015) proposed a function-on-function regression which focus on characterizing the time-varying relationship between ERP outputs from specific electrode pairs.

However, there are no established methods to guide the decision of which electrodes should be clustered together. Here we propose a Bayesian nonparametric approach that allows to group electrodes together based on the observed multichannel ERP waveforms. Our approach allows to flexibly model the ERP temporal dynamics, takes into account spatial dependencies in the brain and locate symmetries in the brain response to stimuli.

In the second motivating application, the Italian natural gas balancing platform (PB-GAS), a system where gas traders virtually sell and buy natural gas to balance the common pipeline give rise to daily demand and supply curves, i.e., a time series in which a curve is observed at each time, namely a time dependent functional time series. Managed by the energy regulatory *Gestore Mercati Energetici* (GME) with SNAM acting as central counterpart for all daily offers. Every day SNAM submits a demand bid or supply offer for a volume of gas corresponding to the overall imbalance of the system while the operators submit demand bids and supply offer for the storage resource they have available.

The daily exchanged price and quantity is in the intersection of supply and demand curves. Additionally, the curves are bounded, monotonic and clearly time dependent due to the nature of the market; one would expect the exchanged price and quantity of today affect those of tomorrow.

The statistical problem of predicting the price and quantity exchanging in this functional dataset or more general in any two-curves system can be solved in many ways, e.g., by a) separately forecast the two curves and then via their intersection pick up the predicted price and quantity or b) by joint estimation of demand and supply curves. Another possibility is instead c) use classical time series approach considering as dataset the intersection of the curves. Note that, in this approach, the resulting time series is not functional.

Canale and Vantini (2014) using the approach a) bypassed the monotonicity and boundedness constraints on the supply and demand curves by mapping them into an unconstrained space. The new unconstrained curves are them modeled by a functional-to-functional autoregressive model and mapped back to the original space. In statistical problems involving unconstrained curves the functional autoregressive (FAR) model (Bosq, 1991) plays a crucial role. In the nonparametric approach to unconstrained curves to model functional time series refer to Ferraty and Vieu (2006) and references there in.

Regards to the approach b) to our knowledge there is no proposed method within FDA framework. Readers interested in this statistical problem may refer to the unpublished work

by Brumm and Mccallum (2008) where a method is proposed to solve the "simultaneous equation problem" with application to supply and demand system.

Though the approach c) may sounds appealing, for the considered application, the whole information present in the curves is lost and, consequently, that about the behavior of the marketing for the next day impoverishing the analysis by the gas trader. The previous also apply for any application where information about the curves is intended to be preserved.

Canale and Vantini (2014) cleverly bypassed curves constraints; nevertheless to our knowledge there is no proposed method, in the context of dependent functional data, where constraints are actually taken into account in the model. Here we pursue the approach a) and propose a nonpararametric Bayesian model that actually can be used to model dependent functional dataset arising from a number of different phenomena. Our model allows to group curves within classes based on the features of the functions while taking into account time dependence and curves constraints.

1.2 Main Contributions of the Thesis

In the framework of functional data analysis we propose two Bayesian Nonparametric models briefly described below. Besides the models, in Section 2.8 we present a method, based on definitions and properties of splines in De Boor (2001), to be used as a guide to those who want do input constraints in B-spline basis functions. In Section 4.1.1 we extend the probit model applied to heterogeneous systems in Heckman (1981) and applied to stick-breaking prior in Dunson and Rodríguez (2011) to a probit dynamic model to accommodate time dependence over a partitioned sample space Ω by conditioning the process to proportion of samples arising from each partition of the function space.

1.2.1 A Bayesian Nonparametric approach for the analysis of functional ERP data in neuroimaging

Let $y(s,t)$ denote the functional ERP response observed at time $t = 1, \dots, T$ in each electrode s, $s = 1, \dots, S$. We assume that

$$y(s,t) \;=\; f(s,t) + \epsilon(s,t)$$

where the collection $\{f(s,t)\}_{t=1}^{T}$ is considered to be the realization of a random function $f(s,\cdot) : \mathscr{R}^{+} \to \mathscr{R}$ and $\epsilon(s,t)$ is a spatio-temporal error process. The functions f's are assumed to be spatially correlated and clustered together by an underlying Functional Dependent Dirichlet process which encodes a conditional autoregressive dependence structure

to guide the spatial selection, and uses flexible spline-based curve estimates as atoms of the process to describe the temporal dependence. In Section 2.4 and 2.5, we review the essential about Functional Dependent Dirichlet process and conditionally autoregressive models needed for the development of our proposed model. Bayesian variable selection techniques are used to select significant sets of bases coefficients in each cluster. Typical shape characteristics of the ERP functionals and spatial symmetries of the functional responses in the brain can be appropriately accounted for in our framework. We illustrate the clustering behavior of the process and the related improvements in the detection of brain activity for the multi-comparison problem on a simulated dataset which mimic the typical ERP waveforms observed in real EEG experiments. We then apply our method to the analysis of a multiple pictures-viewing task in a group of smokers interested in quitting, but prior to the onset of treatment, and show differential activations in response to cigarette-related cues.

1.2.2 A Bayesian Nonparametric approach for the analysis of the balancing marketing

Let $y(t,q)$ denote the supply or demand curves collected at days t, $t = 1,\dots,T$ for each quantity $q = 1,\dots,Q$. We assume that

$$y(t,q) \;=\; f(t,q) + \epsilon(t,q)$$

where the collection $\{f(t,q)\}_{q=1}^{Q}$ is considered to be the realization of a random function $f(t,\cdot) : \mathscr{R}^+ \to \mathscr{R}$ and $\epsilon(t,q)$ is a specific residual to time t. The functions f's time dependence are induced in the weights of an underlying Functional Dependent Dirichlet process through a dynamic linear model. In Section 2.4 and 2.6, we review the essential about Functional Dependent Dirichlet process and dynamic linear model needed for the development of our proposed model.

The problem of forecasting $f(T+1,\cdot)$ (both for supply and demand curves) is solved by predicting the latent process involved in the specification of the weights followed by the calculation of the weights itself. Once the functions are forecasted, the predict price and quantity are in their intersection. We illustrate the clustering behavior of the process on a simulated dataset which mimic the typical supply curves and then apply our method to the Italian natural gas balancing platform functional dataset.

In Chapters 3 and 4 the proposed models are presented along with their Gibbs sampling schemes.

Chapter 2

A brief theoretical background

In addition to the presentation of the motivational applications this section is thought to be an introduction to the main statistical theories needed in our proposed Bayesian Nonparametric models. Concepts such as Functional Data Analysis, Dependent Dirichlet Process, Conditionally Autoregressive, Dynamic Linear models, B-splines and Control Polygon, Feature Classes and Binder's loss are presented.

2.1 Application in neuroimaging

Event-Related Potentials (ERP) represent a safe and non-invasive approach to directly measure brain activity, and constitute an ideal tool to investigate the neural correlates of cognitive and emotional processes in humans. ERPs are voltage fluctuations in the electroencephalogram (EEG) that are time-locked to the presentation of external events (e.g., pictures). Thanks to their high temporal resolution (in the order of the milliseconds), ERPs allow researchers to investigate fast-changing brain processes and constitute a sensitive and reliable biomarker of motivational and cognitive processing (Kenemans and Kähkönen, 2001; Lang and Bradley, 2010; Luck et al., 2011).

The application here considered is a multiple pictures-viewing task experiment in a group of smokers interested in quitting, but prior to the onset of treatment, to study the differential activations in response to cigarette-related cues. The experiment was performed in the Department of Behavioral Sciences at University of Texas M. D. Anderson Cancer Center by Dr. Francesco Versace and his researching group as follows: the hydrocel geodesic sensor net, see scheme in Figure 2.1, is placed on the scalp of the subject; a picture with cigarette-related, emotional (pleasant and unpleasant) or neutral content is showed on the monitor and the brain activity is recorded by electrodes placed in each of the 129 spots in the sensor net.

Figure 2.2 illustrates examples of brain activities recorded by four of the electrodes to highlight the different behavior of the curves between the time interval 50-100 ms; the dashed line located at 26 ms marks the stimulus starting point, i.e., when the picture is showed on the screen. Observe that the electrodes 1 and 15 presents maximum values within this interval while electrodes 65 and 81 their minimum values classifying the curves in two symmetric subclasses. This well known symmetry (see Dickter and Kieffaber, 2013; Nunez and Srinivasan, 2006) is an important information to be accounted for and is included in our model by informative prior with the proportion of curves falling into each subclass. In Figure 2.3 is presented the spatial configuration of some of the electrodes together with the recorded brain activities.

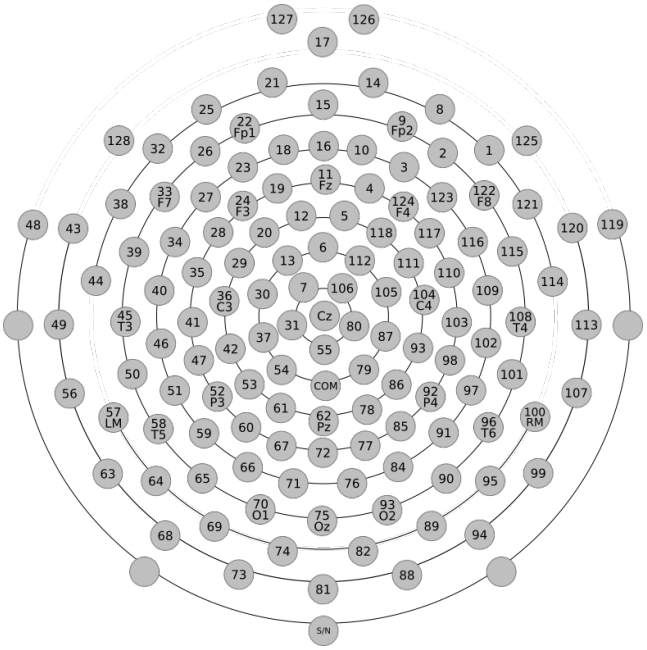
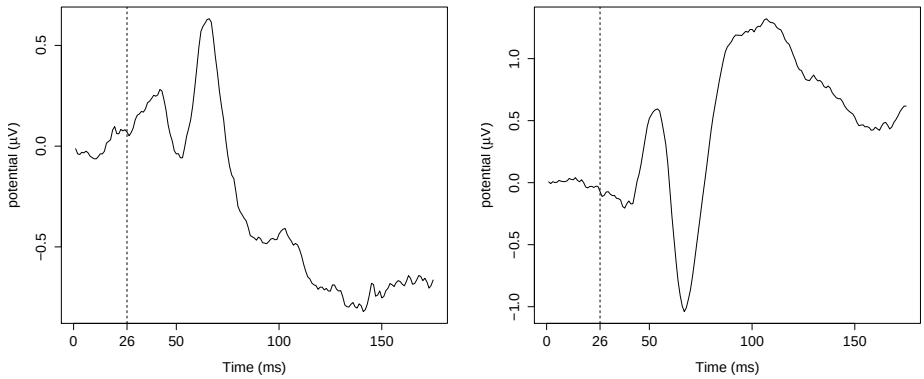


Fig. 2.1 Geodesic sensor net, 128 channels

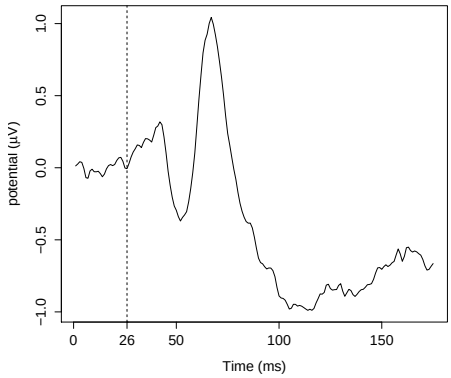
The spatial configuration confer to the data the label areal. In areal data nomenclature the electrodes are areal units and group of electrodes regions. In this type of experiment the response has its origin in the occipital lobe located in the back of the brain and spread out suggesting that the waveforms recorded by each electrode are spatial correlated.

The ERP dataset are the outcomes resulting after the experiment described above. Here we consider a single subject (grand average across subjects (Dickter and Kieffaber, 2013) could also be used) and a single stimulus, organized in a matrix of order number of electrodes $\times$ number of time points, a 129×175 matrix. In this setting a statistical unit is a curve (outcome of one of the channels) stored in one of the rows $i, i = 1, \ldots, 129$, of the dataset matrix, called functional dataset.

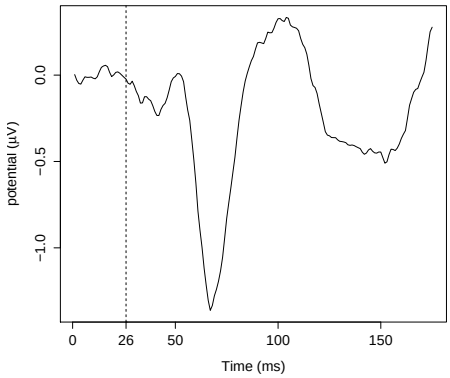


(a) Channel 1

(b) Channel 65



(c) Channel 15



(d) Channel 81

Fig. 2.2 Real data: ERP waveforms recorded by electrodes 1, 65, 15 and 81.

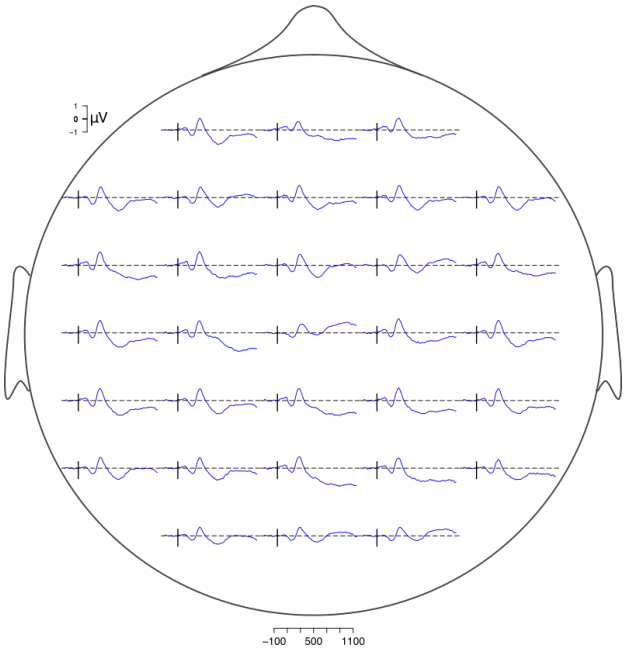
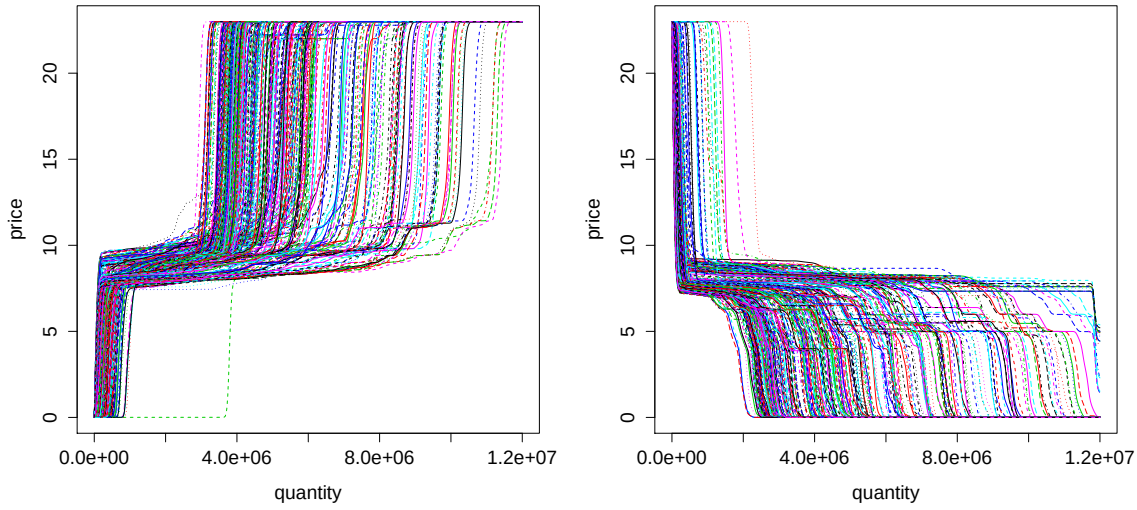


Fig. 2.3 Real data: spatial configuration of some selected electrodes.

2.2 Application in balancing marketing

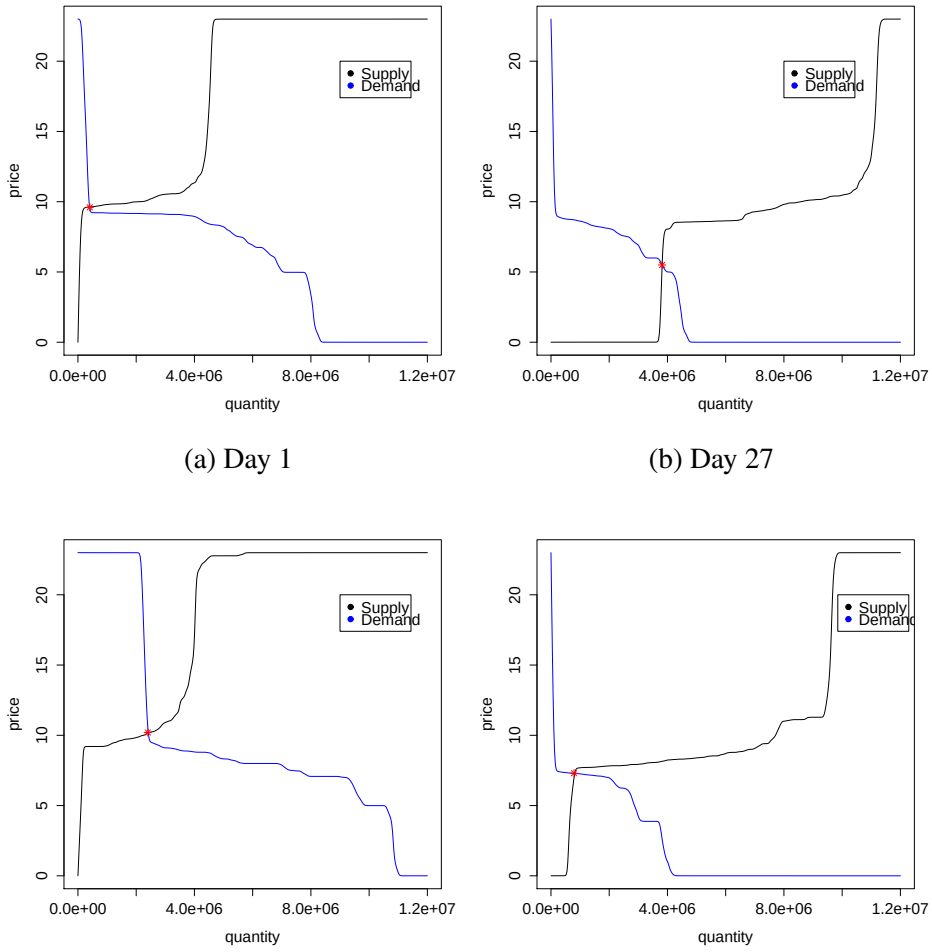
The Italian natural gas balancing platform (PB-GAS) is a system in which gas traders virtually sell and buy natural gas to balance the common pipeline. The *Gestore Mercati Energetici* (GME) organizes and manages the PB-GAS while SNAM act to keep the gas system balance. Every day SNAM submits a demand bid or supply offer for a volume of gas corresponding to the overall imbalance of the system while the operators submit demand bids and supply offer for the storage resource they have available.

Each dataset, stored in a 397×501 matrix, are the observed demand and supply curves for 397 days in the period of December 1st, 2011 to December 31rd, 2012. Figure 2.4 illustrates the supply and demand curves observed in 397 days. Note that the curves are bounded, monotonic and clearly time dependent due to the nature of the market; one would expect the exchanged price and quantity of today affect those of tomorrow. In this dataset a single curve contains 501 observations with quantity ranging from 0 to 1.2×10^7 gigajoules (GJ) and price from 0 to 23 Euro/GJ. In this setting a statistical unit is a curve stored in one of the rows $i, i = 1, \ldots, 397$, of the dataset matrix, called functional dataset (see Section 2.3).



(a) Supply (b) Demand

Fig. 2.4 Real data: Italian natural gas balancing platform dataset



(a) Day 1 (b) Day 27

(c) Day 68 (d) Day 394

Fig. 2.5 Real data: two-function system, supply and demand curves. The red star represents the exchanged quantity and price.

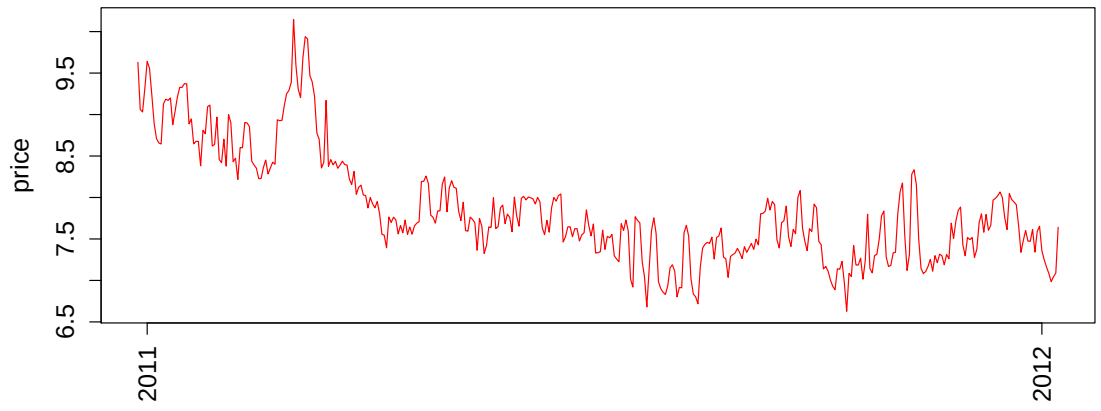


Fig. 2.6 Real data: time series of exchanged quantity and price

The daily exchanged quantity and price is in the intersection of the supply and demand curves as illustrated in Figure 2.5 for days 1, 27, 68 and 394. The intersections of the all 397 days, the red stars in the plot, give rise to the time series illustrated in Figure 2.6.

2.3 Functional Data Analysis

In the field of Functional Data Analysis (FDA) one works with the triplet $(\Omega, \mathscr{B}, P)$ where Ω is a function space, B and P are the usual σ-algebra of subsets of Ω and a probability measure, respectively. An observation f of a random variable F, also named functional variable, in Ω is called functional data. Unlike other authors, Ferraty and Vieu (2006) present formal definition of key concepts in functional data analysis and is recommended for whom that would like to study it with mathematical rigor. For more intuitive definitions and motivational examples refer to Ramsay and Silverman (2006). In the following we alternate between the mathematical rigor and intuition depending on what is the idea to be presented.

In FDA each sample element is a curve, surface or any mathematical object varying over a continuum and provide information of an underlying process generating the data, e.g, a single curve recorded by an electrode in the ERP dataset or a supply curve observed in one of the 397 days in the Italian natural gas dataset.

At this point one, to make descriptive analyses on the datasets, could have the intuition what is the functional mean and others statistical moments to analyze functional dataset. These descriptive statistics and basic statistical tools to analyze and model functional data including smoothing, time-warping, functional linear models, functional principal components and basis function may be studied in Ramsay and Silverman (2002). Of these statistical methods basis function is an outstanding tool for this thesis as it allows to write the functional observation by a linear combination of L known basis functions b_l,

$$f(\cdot) = \sum_{l=1}^{L} \beta_l b_l(\cdot), \tag{2.1}$$

where the coefficients $\beta = (\beta_1, \ldots, \beta_L)$ specifies the function $f(\cdot)$. Usually the number of basis functions is lower than the number of observations of f allowing a efficient storage of functional data. Examples of basis function include Fourier series, B-spline, I-spline and wavelet. One choose the basis accordingly to the features of the data; the Fourier basis allows to write functions presenting periodicity, I-spline is suitable for functions presenting monotonicity, wavelets are suitable for bounded functions and B-spline are flexible enough to allow constraint in the coefficients such that features as monotonicity and boundedness could be accommodated. A dense material about B-splines is found in de Boor (1978).

2.4 Functional Dependent Dirichlet Process

In the Bayesian Nonparametric approach to statistical problems the probability measure P on (Ω, B, P), defined in section (2.3), is allowed to be uncertain. To set a prior to the random measure P define it on probability space $(\Upsilon, \mathscr{B}, \mathscr{P})$, where Υ is the space of probability measures over (Ω, B), $\mathscr{B}$ the usual σ-algebra of subsets of Υ and $\mathscr{P}$ a probability measure, i.e., $P \sim \mathscr{P}$.

In this thesis we focus on the stick-breaking prior and write, following Ishwaran and James (2001),

$$P = \sum_{k=1}^{K} p_k \delta_{\theta_k}, \tag{2.2}$$

where δ_θ is the dirac probability measure with density only in θ_k, the so called atom, $0 \leq p_k \leq 1$ are random weights with $\sum_{k=1}^{K} p_k = 1$ almost surely. Define,

$$\begin{aligned} p_1 &= V_1; \\ p_k &= (1 - V_1)(1 - V_2) \cdots (1 - V_{k-1}) V_k, \ k \geq 2, \end{aligned} \tag{2.3}$$

where $V_k \overset{iid}{\sim} \text{Beta}(a_k, b_k)$, $\boldsymbol{a} = (a_1, a_2, \cdots, a_K)$ and $\boldsymbol{b} = (b_1, b_2, \cdots, b_K)$.

Potentially $1 \leq K \leq \infty$; computationally a finite truncation is required. Though some experienced statistician have intuition of the suitable truncation for the general P_∞ theoretical approach may be used to this task as presented in Theorem 1 and 2 in Ishwaran and James (2001). Once a truncation level is chosen it is necessary to guarantee that $\sum_{k=1}^{K} p_k = 1$ with probability 1, in the case of stick-breaking prior it is sufficient to setting $V_K = 1$ because $1 - \sum_{k=1}^{K-1} p_k = (1 - V_1) \cdots (1 - V_{K-1})$ with $\boldsymbol{a} = (a_1, a_2, \cdots, a_{K-1})$, $\boldsymbol{b} = (b_1, b_2, \cdots, b_{K-1})$, $p_k = V_k \prod_{l<k}(1 - V_l)$. From now on P is used to indicate a truncated stick-breaking prior as opposite to P_∞ that indicates the infinite one.

Using the general definition of P a number of process may be characterized. Examples include the Dirichlet multinomial processes (Muliere and Secchi, 2001), m-spike models (Liu, 1996) two-parameter Poisson-Dirichlet process (Pitman and Yor, 1997), finite dimensional Dirichlet priors (Ishwaran and Zarepour, 2000a,b) and beta two-parameter processes (Ishwaran and Zarepour, 2000c). Our foucs is on the Ferguson Dirichlet process (Ferguson, 1973a,b) in which the parameters in the beta distribution are set to $\boldsymbol{a} = \mathbf{1}$ and $\boldsymbol{b} = \boldsymbol{\alpha}$.

The process just described produce independent and identically distributed draws of the $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ from a base measure P_0 and $\boldsymbol{V} = (V_1, \ldots, V_{K-1})$ from Beta$(1, \alpha)$. In Dirichlet Process (DP) P_0 is the expected value of the process and α the concentration parameter

expressing how close P is to P_0. Potentially a prior can be set to α and let the data inform about it (Escobar and West, 1995).

MacEachern (2000) proposed the Dependent Dirichlet Process (DDP) by replacing the single random probability measure P for a dependent collection of random probability measures $P_\chi = \{P_x : x \in \chi\}$ with mass V_x assigned to the location θ_x for each value of the covariate $x \in \chi$ where x refer to time, space, or predictors.

A simplification (computationally speaking) of the full DDP (where dependence is induced in the atoms and weights of the process) is the single-p DDP model (MacEachern, 2000), i.e., $p_{k(x)}$ does not vary over χ as do the atoms $\theta_{k(x)}$ of the process, viz

$$P_x(1,\alpha) = \sum_{k=1}^{K} p_k \delta_{\theta_{k(x)}}(\cdot), \quad \theta_{k(x)} \sim P_{0(x)} \tag{2.4}$$

where $p_k = V_k \prod_{l<k}(1 - V_l)$ and $V_k \overset{iid}{\sim} \text{Beta}(1,\alpha)$.

As illustration of this type of model, consider the ANOVA DDP model proposed in De Iorio et al. (2009), where the dependence across distributions is introduced by modeling the atoms $\theta_{k(x)} = m_k + A_{\nu k} + B_{wk}$ with $x = (\nu, w)$ a bivariate covariate, $\nu \in \{1,\dots,V\}$ and $w \in \{1,\dots,W\}$. $x = (\nu, w)$ is seen a as level of two treatments in a clinical trial and P_x the random distributions of outcomes for each patient, where $m_h \overset{iid}{\sim} p^0_{m_k}$, $A_{k(\nu)} \overset{iid}{\sim} p^0_{A_{k(\nu)}}$ and $B_{k(w)} \overset{iid}{\sim} p^0_{B_{k(w)}}$ may be interpreted as an overall mean, main effects for covariate level ν and w, respectively. Jointly, the probability model on $\wp = \{P_x, x \in \chi\}$ is an ANOVA DDP(α, p^0) where α is the DP concentration parameter and p^0 the base measure. Moreover, marginally P_x is a DP process with base mesure P_{0x} given by the convolution of $p^0_{m_k}, p^0_{A_{k(\nu)}}$ and $p^0_{B_{k(w)}}$ with dependence among $\theta_{k(x)}$ defined by its covariance structure. Note that a characteristic of the single-p model is that it does not produce independent distributions.

Others examples of the single-p model include the spatial Dirichlet process mixture by Gelfand et al. (2005) and the extension of the DDP model incorporating dependence in the random effects of distribution across groups by De la Cruz-Mesia et al. (2007).

Another option to the full DDP model is instead leave just the weights $p_{k(x)}$ varying with x, that is,

$$P_x(1,\alpha) = \sum_{k=1}^{K} p_{k(x)} \delta_{\theta_k}(\cdot), \quad \theta_k \sim P_0 \tag{2.5}$$

where $p_{k(x)} = V_{k(x)} \prod_{l<k}(1 - V_{l(x)})$ and $V_{k(x)} \overset{iid}{\sim} \text{Beta}(1,\alpha)$. As example, Reich and Fuentes (2007) replaced V_k in the stick-breaking prior by $w_{k(x)}V_k$, where $w_{k(x)}$ is a kernel function

restricted to the interval $[0, 1]$. In this model the random distributions P_x are dependent through the weights and θ_k independent across k allowing borrowing of information at different sites. They applied their model in a multivariate spatial setting to analyze wind field data.

Other examples of varying-p models include the Generalized spatial Dirichlet process models by Duan et al. (2007), the order based dependent DP by Griffin and Steel (2006) and the Kernel stick-breaking processes by Dunson and Park (2008) and the Hybrid Dirichlet mixture to model functional dependence in the label process through Gaussian Copula by Petrone et al. (2009).

Due to the functional nature of our datasets we go further on these definitions towards the Functional Dependent Dirichlet Process (FDP). In this process Ω is a function space as in Section 2.3. To be more precise, the atoms in the DDP are suitably replaced by functional atoms according to the application. To our purpose the base measure P_0 is chosen to be B-Splines (de Boor, 1978) due to its flexibility and easy of implementation of curve constraints. See Section 2.8 for details.

2.5 Conditionally Autoregressive Model

The conditionally autoregressive (CAR) model, in particular the auto-normal case, is commonly used for incorporating the dependence present in spatially referenced data into the covariance structure of a gaussian distribution. The CAR auto-normal model is presented as a case on the broad class of conditional probability models to describe spatial process developed by Besag (1974). Key aspects on its development are the assumption of a system with finite set of sites associated with random variables $S_1, \ldots, S_n$ with conditional probability $P(s_i|s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ and the positivity condition by Hammersley and Clifford (1971) stating that if $P(s_i) > 0$ for each i, then $P(s_1, \ldots, s_n) > 0$. To obtain the joint probability distribution consider for any two given realizations $\boldsymbol{s} = (s_1, \ldots, s_n)$ and $\boldsymbol{t} = (t_1, \ldots, t_n)$ with $\boldsymbol{s}$ and $\boldsymbol{t} \in \Omega = \{\boldsymbol{s} : P(\boldsymbol{s}) > 0\}$, then the required joint probability is written as

$$\frac{P(\boldsymbol{s})}{P(\boldsymbol{t})} = \prod_{i=1}^{n} \frac{P(s_i|s_1, \ldots, s_{i-1}, t_{i+1}, \ldots, t_n)}{P(t_i|s_1, \ldots, s_{i-1}, t_{i+1}, \ldots, t_n)}, \tag{2.6}$$

where the non-zero denominator is guaranteed by the positivity condition. It is know that the joint probability distribution determines the conditional probabilities, to guarantee that the joint probability is uniquely determined restrictions on the conditional probabilities must be imposed to avoid many possible factorizations of $P(\boldsymbol{s})/P(\boldsymbol{t})$.

The just mentioned factorization problem due to an arbitrary labelling of the sites in the considered system is addressed by Hammersley and Clifford (1971) theorem in which the set of neighbors for each site i is determined. The neighboring scheme is valid for a broad class of lattice (known as Bravais lattice due to Auguste Bravais).

In the most general case to define that the site j is a neighbor of site i it is necessary and sufficient that the conditional probability $P(s_i|\text{all other variables})$ depends only upon to x_j. At this point one may intuit we are heading to the definition of a Markov random field and therefore obtain valid conditional probabilities to the spatial scheme. To do so a crucial concept is the so called clique: the neighborhood configuration in which, given a set of sites every site is a neighbor of all the other sites in the set. To understand the importance of a clique consider $Q(\boldsymbol{x}) \equiv \ln\{P(\boldsymbol{x})/P(\mathbf{0})\}$, assuming $P(\mathbf{0}) > 0$ to allow realization of values equal zero in any site, is unique expanded as

$$\begin{aligned} Q(\boldsymbol{x}) &= \sum_{1\leqslant i\leqslant n} x_i G_i(x_i) + \sum_{1\leqslant i<j\leqslant n}\sum x_i x_j G_{i,j}(x_i, x_j) + \sum_{1\leqslant i<j<k\leqslant n}\sum\sum x_i x_j x_k G_{i,j,k}(x_i, x_j, x_k) \\ &\quad + \ldots + x_1 x_2 \cdots x_n G_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \end{aligned} \tag{2.7}$$

where the function $G_{i,j,\ldots,k}(x_i, x_j, \ldots, x_k)$ may be non-null if and only if the sites $i, j, \ldots, k$ form a clique.

Before define the auto-normal case it is useful pass by the general class of auto-models, i.e., the conditional probability $p_i(\cdot)$ associated with site i belongs to the exponential family of distributions for all i.

Defined our scope, consider cliques are of size 2, meaning that if i is neighbor of site j, j is neighbor of site i, denoted by $i \sim j$[1] in which the sites form a finite rectangular lattice. Then the expansion (2.7) is written as

$$Q(\boldsymbol{x}) = \sum_{1\leqslant i\leqslant n} x_i G_i(x_i) + \sum_{1\leqslant i<j\leqslant n}\sum \beta_{i,j} x_i x_j \tag{2.8}$$

where $G_{i,j}(x_i, x_j) \equiv \beta_{i,j} H_i(x_i) H_j(x_j)$ may be non-null if $i \sim j$, $x_i H_i(x_i) = B_i(x_i) - B_i(0)$ and B_i is linear in x_i. In this scenarios the conditional probability is

$$\frac{p_i(x_i;\ldots)}{p_i(0;\ldots)} = \exp\left[x_i\left\{G_i(x_i) + \sum_{j=1}^{n} b_{i,j} x_j\right\}\right], \tag{2.9}$$

[1] in this case $\sim$ stand for equivalence relation in mathematics

where $\beta_{i,j} \equiv \beta_{j,i}$ and $\beta_{i,j} = 0$ unless $i \sim j$. In the auto-normal case we assume the conditional probability

$$p_i(\cdot) = (2\pi\tau^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\tau^2}\left\{x_i - \mu_i - \sum b_{i,j}(x_j - \mu_j)\right\}^2\right] \tag{2.10}$$

with mean equal zero and multivariate normal joint probability, using the factorization 2.6,

$$P(\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}^T D^{-1}(I - B)\boldsymbol{x}\right\}, \tag{2.11}$$

$B = \{b_{ij}\}$ and D is diagonal with $D_{ii} = \tau_i^2$. To ensure $D^{-1}(I - B)$ is symmetric set $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, where $w_{ij} = 1$ if $i \sim j$ and 0 otherwise are entries in the adjancency matrix $W_{n\times n}$ and write

$$P(\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}^T(D_w - \rho W)\boldsymbol{x}\right\}, \tag{2.12}$$

where D_w is diagonal with elements w_{i+} and $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$, where $\lambda_{(1)} < \cdots < \lambda_{(n)}$ are the eigenvalues of $D_w^{-1/2} W D_2^{-1/2}$ guarantee $(D_w - \rho W)^{-1}$ is not singular. Since $\rho = 0$ leads to independence of the sites, say i, and j, it is often called spatial correlation or spatial dependence. With this restrictions the full conditional probability is therefore $p(x_i|x_j, i \neq j) \sim N(\rho \sum_j w_{ij} y_j/w_{i+}, \tau^2/w_{i+})$. From the next Sections we adopt the notation $CAR(\rho, \tau^2)$ to refer to this model.

For a general definition of conditionally autoregressive models and examples of neighboring schemes refer to Besag (1974) from Section 3. For a simpler exposition of conditionally autoregressive models see Chapter 3 in Banerjee et al. (2004).

2.6 Dynamic Linear Model

To motivate this section consider a dynamic system, anything that evolves over time, governed by a process $z_t \sim N(\zeta_t, v_t)$, $t = 1, \ldots, t$. Together, the equations

$$z_t = \zeta_t + v_t, \quad v_t \overset{iid}{\sim} N(0, V_t) \tag{2.13}$$

$$\zeta_t = \zeta_{t-1} + \omega_t, \quad \omega_t \overset{iid}{\sim} N(0, W_t) \tag{2.14}$$

provide understanding about the system, where Equation (2.14) inform its state, and Equation (2.13) inform what the next state of the system will be in the next instant of time given the current state. If the Equations 2.13 and 2.14 satisfy the following assumptions,

1. ζ_t is a Markov chain

2. Conditionally on ζ_t, the z_t's are independent and z_t depends on ζ_t only,

and in addition if at time $t = 0$ the system is started with a gaussian prior, $\zeta_0 \sim N(m_0, C_0)$, 2.13 and 2.14 are said to be a dynamic linear model (Durbin and Koopman, 2001; Petris et al., 2009).

For our purposes we conveniently set $V_t = 1$ and to the unknown variance W_t of the state of the system, focusing in a Bayesian approach, set a prior for its precision of the form

$$W_t^{-1} = \varrho \sim \text{gamma}(a_w, b_w), \tag{2.15}$$

where gamma(a, b) stands for the gamma distribution with shape $a > 0$ and rate $\beta > 0$.

Considering the joint probability distribution $p(\zeta_{0:T}, \varrho | z_{1:T})$, where $\zeta_{0:T}$ and $z_{1:T}$ are finite sequence of consecutive values of ζ and z, the Bayesian statistical problem of drawing the states ζ_t and state precision ϱ given the observations z_t is solved within Gibbs sampler steps by drawing them from their full conditional distributions $p(\zeta_{0:T}|\varrho, z_{1:T})$ and $p(\varrho|\zeta_{0:T}, z_{1:T})$, respectively.

While the full conditional of ϱ_t has the closed form

$$\begin{aligned}
p(\varrho|\zeta_{0:T}, z_{1:T}) &= p(\varrho, \zeta_{0:T}, z_{1:T}) \\
&= p(z_{1:T} \mid \zeta_{0:T}, \varrho)\pi(\zeta_{0:T}|\varrho)p(\varrho) \\
&= \prod_{t=1}^{T} \pi(\zeta_t|\zeta_{t-1}, \varrho)p(\varrho) \\
&\propto \varrho^{a-1} \exp(-\varrho b) \times \varrho^{T/2} \exp\left(-\frac{\varrho}{2}(\theta_t - \theta_{t-1})^2\right) \\
&= \varrho^{a+\frac{T}{2}-1} \exp\left(-\varrho\left[b + \frac{1}{2}\sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2\right]\right),
\end{aligned}$$

therefore

$$(\varrho|\zeta_{0:T}, z_{1:T}) \sim \text{gamma}\left(a + \frac{T}{2}, b + \frac{1}{2}\sum_{t=1}^{T}(\zeta_t - \zeta_{t-1})^2\right), \tag{2.16}$$

sampling from $p(\zeta_{0:T}, |\varrho, z_{1:T})$ is not straightforward. To do so, write

$$\begin{aligned}
p(\zeta_{0:T} \mid \varrho, z_{1:T}) &= p(\zeta_0, \ldots, \zeta_T \mid \varrho, z_{1:T}) \\
&= p(\zeta_T \mid \varrho, z_{1:T})p(\zeta_{T-1} \mid \theta_T, \varrho, z_{1:T}) \ldots \pi(\theta_0 \mid \theta_1, \ldots, \theta_T, \varrho, z_{1:T}) \\
&= \prod_{t=0}^{T} p(\theta_t \mid \theta_{t+1:T}, \varrho, z_{1:T}),
\end{aligned} \tag{2.17}$$

to call attention to the factor $p(\zeta_T \mid \varrho, z_{1:T})$, the filtering distribution of $\zeta_T \sim N(m_T, C_T)$, where m_T and C_T, the mean and variance of a gaussian distribution, is updated by the Kalman filter (Kalman, 1960). To obtain a draw from $p(\zeta_{0:T} \mid \varrho, z_{1:T})$ start by drawing ζ_T and then, for $t = T-1, \ldots, 0$, recursively draw ζ_t from $p(\theta_t \mid \theta_{t+1:T}, \varrho, z_{1:T})$. Next, note that ζ_t and $z_{t+1:T}$ are conditionally indepdent given ζ_{t+1}, therefore $p(\theta_t \mid \theta_{t+1:T}, \varrho, z_{1:T}) = p(\theta_t \mid \theta_{t+1}, \varrho, z_{1:t})$ has distribution $N(h_t, H_t)$, with

$$\begin{aligned} h_t &= m_t + C_t G_{t+1}^{'} R_{t+1}^{-1}(\zeta_{t+1} - a_{t+1}), \\ H_t &= C_t - C_t G_{t+1}^{'} R_{t+1}^{-1} G_{t+1} C_t, \end{aligned}$$

obtained via Kalman smoother, where $a_t = E(\zeta_t | z_{1:t-1}) = G_t m_{t-1}$, $R_t = Var(\zeta_t | z_{1:t-1}) = G_t C_{t-1} G_t^{'} + W_t$, $m_t = E(\zeta_t | z_{1:t}) = a_t + R_t F_t^{'} Q_t^{-1} e_t$, $C_t = Var(\zeta_t | z_{1:t}) = R_t - R_t F_t^{'} Q_t^{-1} F_t R_t$, $f_t = E(z_t | z_{1:t-1}) = F_t a_t$, $Q_t = Var(z_t | z_{1:t-1}) = F_t R_t F_t^{'} + V_t$, and $e_t = z_t - f_t$ is the forecast error. The process just described is the so called forward filtering backward sampling (FFBS) Algorithm, see Algorithm 1.

Algorithm 1 Forward Filtering Backward Sampling

1. Run Kalman filter.
2. Draw $\zeta_T \sim N(m_T, C_T)$.
3. For $t = T-1, \ldots, 0$, draw $\zeta_t \sim N(h_t, H_t)$.

Within a Gibbs sampler, draws from $p(\zeta_{0:T}, |\varrho, z_{1:T})$ and $p(\varrho | \zeta_{0:T}, z_{1:T})$ are summarized in Algorithm 2.

Algorithm 2 Forward Filtering Backward Sampling within a Gibbs Sampler

0. Initialize: set $\varrho = \varrho^{(0)}$.
1. For $i = 1, \ldots, N$:
   a) Draw $\zeta_{0:T}^{(i)}$ from $p(\zeta_{0:T} | z_{1:T}, \varrho = \varrho^{(i-1)})$ using FFBS.
   b) Draw $\varrho^{(i)}$ from $p(\varrho \mid z_{1:T}, \zeta_{0:T} = \zeta_{0:T}^{(i)})$.

2.7 Feature Classes and the Enriched Stick-Breaking Prior

To introduce feature classes as proposed by Scarpa and Dunson (2013), let Ω be the function space where a generic curve f takes its value be expressed as $\Omega = \cup_{a=1}^{\sharp\mathscr{A}} \Omega_a$, where each $\Omega_a = \{f : \mathscr{T}(f) = \mathscr{A}_a, f \in \Omega\}$ congregates functions within feature class a, $\mathscr{T} : \Omega \mapsto$

$\{0,1\}^M$ denote a feature class indicator function and $\sharp\mathscr{A} = 2^M$ are the number of partitions of Ω. Though the number of partitions grows exponentially with the number of features M, with few classes it is possible to accommodate a number of features since for $m = 1, \ldots, M$, $\mathscr{I}_m(f) = 1$ denotes that function f possesses feature m and $\mathscr{I}_m(f) = 0$ otherwise, i.e., each class is the disjoint union of two characteristic-specific subclasses. Information about the features are given a priori and the proportions of functions falling into different feature classes a is given by

$$\mathbf{\Pi} = (\Pi_a, a = 1, \ldots, \sharp\mathscr{A})^T \sim Diri(a_{\Pi,1}, \ldots, a_{\Pi,\sharp\mathscr{A}}). \tag{2.18}$$

Set up the feature classes, Equation 2.2 can be written as

$$P = \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{k=1}^{N_a} p_{a,k}\delta_{\theta_{a,k}} \tag{2.19}$$

where Π_a is the proportion of curves falling into feature class a, $p_{a,1} = \Pi_a V_{a,1}$, $p_j = \Pi_a V_{a,j}$ $\prod_{k<j}(1 - V_{a,k})$, $V_{a,k} \overset{i.i.d.}{\sim} Beta(1, \alpha)$, for $k = 1 \ldots, N_a - 1$ with $\boldsymbol{N} = \{N_a\}_{a=1}^{\sharp\mathscr{A}}$ the truncation level of each class. To truncate the stick-breaking prior to K terms set $V_{a,N_a} = 1$, for each a, to guaranty, by Lemma 1 in Scarpa and Dunson (2013), $\sum_{k=1}^{K} p_k = 1$ almost surely, where $K = (1, \ldots, N_1, N_1 + 1, \ldots, N_1 + N_2, \ldots, \sum N_{a\,a=1}^{\sharp\mathscr{A}-1}, \ldots, N_{\sharp\mathscr{A}})$. Note that, now the base measure is class specific and allow borrow of information through the clusters of the process.

2.8 A note on B-spline and base measure specification

To generate $\theta_{a,k}$ of Equation 2.19, for each k, from the base measure $P_{0,a}$, for each a, write

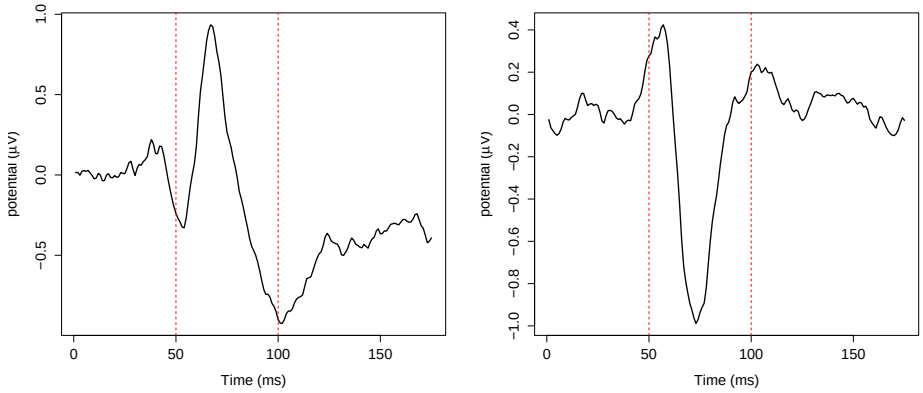
$$\theta_{a,k}(\cdot) = \sum_{l=1}^{L} \beta_{a,kl} b_l(\cdot), \tag{2.20}$$

where $\beta_{a,kl} = (\beta_{a,k1}, \ldots, \beta_{a,kl})^T$ are basis coefficients specific to the kth function atom within class a and $\{b_l, l = 1, \ldots, L\}$ B-spline basis functions having L equally-spaced knots in [0,1], $b_1(0) = 1$ and $b_L(1) = 1$ with the basis coefficients drew from

$$\beta_{a,kl} \sim \nu_0\delta_0(\cdot) + (1 - \nu_0)\phi_{kl}(\cdot; 0, c),\ l = 1, \ldots, L, \tag{2.21}$$

where $\phi_{kl}(\cdot, \mu, \sigma^2)$ denotes a gaussian density and $\nu_0 \sim \text{beta}(a_{\nu_0}, b_{\nu_0})$.

A key aspect is obtain draws of $\theta_{a,k}$ from $P_{0,a}$, i.e, draw the functional atoms such that $f \in \Omega_a$ and therefore f falls within feature class a. As an illustration, consider our application in neuroimaging, see Section 2.1, where the interest is to take into account prior information of symmetries in the brain response to stimuli. To see this, in the Figure 2.7 two ERP responses recorded by electrodes 3 and 57 located in the frontal and occipital brain lobes, respectively. The red dashed lines mark the interval between 50 to 100 ms where the curves $f(3,\cdot)$ and $f(57,\cdot)$ change their behavior presenting its maximum and minimum value within this interval, respectively. Therefore all curves belongs to the same feature class, $M = 1$, the class in which a maximum or a minimum values of the f's are observed within the considered interval. In this feature class the curves are expected to fall within one of two subclasses, $\sharp\mathscr{A}$=2, as follows: using the same notation for feature classes as above, $\mathscr{I}_1(f) = 1$ if f, within the interval 50 to 100 ms, has a maximum value and $\mathscr{I}_1(f) = 0$ otherwise. The definition of this feature class split the function space Ω into $2^M = 2$, where Ω_1 contains functions with maximum values within the considered interval and Ω_2 functions with minimum values within the same interval.



(a) Electrode 3 (b) Electrode 57

Fig. 2.7 Real data: ERP waveforms recorded by electrodes.

.

To obtain functions belonging to each subclass we restrict draws of B-splines coefficients, specifically we restrict $\phi_{kl}(\cdot)$ in Equation 2.21 to a region A_a, for selected a's in $a = 1, \ldots, \sharp\mathscr{A}$, resulting in $\theta_{a,k}$ belonging to the feature class a. Specifically, since the curve's feature we are interested in is within the time interval 50–100 ms we want to control the shape of the curve in this time interval by localizing the knots controlling this shape. Below we introduce some B-splines definitions and properties, see Chapter IX and XI in

De Boor (2001), leading to the definition of control point sequence and control polygon used to tackle this task.

Let $\boldsymbol{r} := \{r_j\}_1^{L+\kappa}$ be real values, called knots, of a κ-order B-spline $B_{\kappa,t}$, $B_{\kappa,t} := \{\sum_j \beta_j b_j : \beta_j \in \mathbb{R}, \text{ for all } j\}$, with $L \geq \kappa > 1$ such that $r_j < r_{j+\kappa}$ for all j and $r_1 = \ldots = r_\kappa = a_B$, $r_{L+1} = \ldots = r_{L+\kappa} = b_B$; call $[a_B, b_B] = [r_\kappa, r_{L+1}]$ the basic interval. The convex hull property states that for $r_j < x < r_{j+1}$, $x \in [a_B, b_B]$, the spline $B^*_{\kappa,r} = \sum_j \beta_j b_j$ is a strictly convex combination of the $\beta_{j+1-\kappa}, \ldots, \beta_j$, i.e., spline values are bounded by the κ nearby numbers and $B_j(x) = 0$ for $x \notin [r_j, r_{j+\kappa}]$ leading to the definition of control polygon; indeed, $x = \sum_j r^*_{j\kappa} b_j(x)$, $x \in [a_B, b_B]$, where

$$r^*_{j\kappa} = \frac{r_{j+1} + \cdots + r_{j+\kappa-1}}{\kappa - 1}, \text{ for all } j. \tag{2.22}$$

Essentially, the Equation 2.22 is a method to localize knots controlling the shape of the curve in the interval $[j + 1 - \kappa, j]$, $P_j := (r^*_{j\kappa}, \beta_j) \in \mathbb{R} : j = 1, \ldots L$ is the control point sequence of the spline $\sum_j \beta_j b_j \in B_{\kappa,t}$ and $C_{\kappa,r}$ is the control polygon of a spline $B^*_{\kappa,r} \in B_{\kappa,r}$ with P_j as vertex sequence.

Back to the illustration in Figure 2.7, given a sequence of knots $\boldsymbol{r}$ one localize the knots r^* controlling the shape of the curve between the time interval 50–100 ms and draw the coefficients with 0 lower bound values for curves falling into subclass $a = 1$ and $\mathscr{I}_1(f) = 1$, e.g., the case of Figure 2.7a, and draw coefficients with opposite constraints for curves falling into sublcass $a = 2$ and $\mathscr{I}_1(f) = 0$, e.g., the case of Figure Figure 2.7b. For a combination of curve's shape constraints, e.g., whether the curves in subclass 1 were also monotonic within the time interval 50-75 ms Abraham and Khadraoui (2015) is a good reference.

2.9 Binder's Loss

The Bayesian clustering problem in this thesis, for each of the two proposed models, is classify each curve in classes, subclasses and groups, in this order. To give an example, considering the neuroimaging application, the ERP responses, see Figure 2.7 above, are allocated in one of two subclasses and groups within each subclass. To tackle this problem consider, as before, $\Omega = \cup_{a=1}^{\sharp\mathscr{A}} \Omega_a$ the partitioned function space, $K = \sum_{a=1}^{\sharp\mathscr{A}} N_a$ the upper bound of number of cluster of the Dirichlet process, with N_a the truncation level of each class for all a, and for this Section $\sharp n$ the total number of subjects. Introduce the allocation variables $\boldsymbol{\mathscr{K}} = (\mathscr{K}_1, \ldots, \mathscr{K}_{\sharp n})^T$ and $\boldsymbol{\mathscr{C}} = (\mathscr{C}_1, \ldots, \mathscr{C}_K)^T$ such that $\mathscr{K}_i = k$, $k \in (1, \ldots, N_1, N_1 + 1, \ldots, N_1 + N_2, \ldots, \sum N_{a\,a=1}^{N_{\sharp\mathscr{A}-1}}, \ldots, N_{\sharp\mathscr{A}})$, and $\mathscr{C}_k = a$ indicates that $f(i, \cdot)$ is allocated to cluster k within feature class a, implying $f(i, \cdot) = \theta_{a,k}$. In this setting

cluster estimates are based on the posterior distributions (see Ishwaran and James, 2001) $(\mathscr{K}|\text{data})$ and $(\mathscr{C}|\text{data})$. The last is induced from the former by relabelling the indexes accordingly.

The loss of estimating the true clustering $\mathscr{K}$ and $\mathscr{C}$ by the MCMC samples $\widehat{\mathscr{K}}$ and $\widehat{\mathscr{C}}$ is measured by the loss functions $L_k(\mathscr{K}, \widehat{\mathscr{K}})$ and $L_c(\mathscr{C}, \widehat{\mathscr{C}})$; here the $L(\cdot)$'s are the Binder's Loss (see Binder, 1978; Fritsch and Ickstadt, 2009; Lau and Green, 2007), a quadratic function of the counts, without loss of generality, given by

$$L_k(\mathscr{K}, \widehat{\mathscr{K}}) = \sum_{i<j} l_1 I\{\mathscr{K}_i = \mathscr{K}_j\} I\{\widehat{\mathscr{K}}_i = \widehat{\mathscr{K}}_j\} + l_2 I\{\mathscr{K}_i = \mathscr{K}_j\} I\{\widehat{\mathscr{K}}_i = \widehat{\mathscr{K}}_j\}, \tag{2.23}$$

with $l_1 = l_2 = 1$ to penalize the two allocation errors equally, i.e., the error of allocating two observations to different clusters when they should be in the same cluster or allocating them to the same cluster when they should be in different clusters, for all possible pair of observations $i, j \in \{1, \ldots, \sharp n\}$. The optimal clustering configuration $\mathscr{K}^*$ is the partition which minimizes

$$\sum_{i<j} \left(I\{\widehat{\mathscr{K}}_i = \widehat{\mathscr{K}}_j\} - \pi_{i,j} \right)^2, \tag{2.24}$$

where $\pi_{i,j} = \Pr\left(\widehat{\mathscr{K}}_i = \widehat{\mathscr{K}}_j\}|\text{data}\right)$ is the posterior probability that i and j belongs to the same cluster.

Chapter 3

A Bayesian Nonparametric approach for the analysis of functional ERP data in neuroimaging

In this Chapter we propose a Bayesian nonparametric approach that allows to group electrodes together based on the observed multichannel ERP waveforms. Our approach allows to flexibly model the ERP temporal dynamics, takes into account spatial dependencies in the brain and feature classes allows us to locate symmetries in the brain response to stimuli.

3.1 Proposed Model

Let $y(s,t)$ denote the functional ERP response observed at time $t = 1, \ldots, T$ in each electrode s, $s = 1, \ldots, S$. We assume that

$$y(s,t) \;=\; f(s,t) + \epsilon(s,t)$$

where the collection $\{f(s,t)\}_{t=1}^{T}$ is considered to be the realization of a random function $f(s,\cdot) : \mathscr{R}^{+} \to \mathscr{R}$ and $\epsilon(s,t) \sim t_{\nu}(\sigma^2)$ is a spatio-temporal error process with ν the degrees of freedom and σ the scalar parameter of a Student-t distribution. To take into account prior information of symmetries in the brain response to stimuli the function space where the functions f's take its values is expressed as $\Omega = \cup_{a=1}^{\sharp\mathscr{A}} \Omega_a$, where each $\Omega_a = \{f : \mathscr{J}(f) = \mathscr{A}_a, f \in \Omega\}$ congregates functions within feature class a, $\mathscr{J} : \Omega \mapsto \mathscr{A} = \{0,1\}^M$ denote a feature class indicator function and $\sharp\mathscr{A} = 2^M$ are the number of partitions of Ω (Scarpa and Dunson, 2013). Moreover, the functions are assumed to be spatially correlated and clustered

together by an underlying Functional Dependent Dirichlet Process (FDDP), viz

$$\boldsymbol{f} \;=\; \{f_s\}_{s=1}^{S} | P_S \sim P_S \tag{3.1}$$

with

$$P_S = \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{u_1,\ldots,u_S} \pi_{u_1,\ldots,u_S} \delta_{\theta_{a,u_1}} \cdots \delta_{\theta_{a,u_S}} \tag{3.2}$$

a FDDP defined over a lattice space. The stochastic weights

$$\pi_{u_1,\ldots,u_S} = Pr\left(\sum_{k=1}^{u_1-1} p_{a,k} < F^{(1)}(\eta_1) < \sum_{k=1}^{u_1} p_{a,k}, \ldots, \sum_{k=1}^{u_S-1} p_{a,k} < F^{(S)}(\eta_S) < \sum_{k=1}^{u_S} p_{a,k} \right), \tag{3.3}$$

where $F^{(1)}(\eta_1), \ldots, F^{(S)}(\eta_S)$ are the cumulative distributions of the marginal distribution of $\boldsymbol{\eta}$, Π_a the proportion of curves falling into feature class a, $p_{a,1} = \Pi_a V_{a,1}$, $p_j = \Pi_a V_{a,j} \prod_{k<j}(1 - V_{a,k})$, $V_{a,k} \overset{i.i.d.}{\sim} Beta(1, \alpha)$, for $k = 1 \ldots, N_a - 1$ with N_a the truncation level within each class. To truncate the stick-breaking prior to K terms set $V_{a,N_a} = 1$, for each a to guaranty, by Lemma 1 in Scarpa and Dunson (2013), $\sum_{k=1}^{K} p_k = 1$ almost surely, where $K = (1, \ldots, N_1, N_1 + 1, \ldots, N_1 + N_2, \ldots, \sum N_{a\,a=1}^{\sharp\mathscr{A}-1}, \ldots, N_{\sharp\mathscr{A}})$.

The Equation 3.3 is a copula type model with CAR prior placed on $\boldsymbol{\eta}$ to accommodate the spatial dependence,

$$\boldsymbol{\eta} = \{\eta_s\}_{s=1}^{S} \sim N_s(\mathbf{0}, \boldsymbol{\Sigma}), \tag{3.4}$$

where $\boldsymbol{\Sigma} = \sigma_\eta^2 (D - \rho W)^{-1}$ with $W_{S\times S}$ an adjacency matrix with entries $w_{ij} = 1$ if $i \sim j$ and 0 otherwise and D is diagonal with elements w_{s+}. Note that, the dependence structure assigned to $\boldsymbol{\eta}$ are induced in the marginals $F^{(1)}(\eta_1), \ldots, F^{(S)}(\eta_S)$.

Use a copula type model to define the weights of a DDP is not new; Petrone et al. (2009) proposed a Hybrid Dirichlet mixture model where the aim was to propose a Bayesian mixture model to reduce dimensionality of curves by representing the sample of curves through a recombination of smaller set of canonical curves. Dependencies on the weights are accommodated through a continuous spatial process; Li et al. (2014) proposed Bayesian models for detecting difference boundaries in areal data by adapting the Hybrid Dirichlet mixture model by modeling the weights using Markov random fields. They apply their model to counting data using Poisson likelihood that could be replaced by other discrete distribution in the ex-

ponential family. Both approaches ensure that the marginal distribution of their model is a Dirichlet process.

The differential of our approach is to model curves over an areal space and include prior information of curve's features. Moreover our model accommodate curves constraints and spatial correlation through a conditional autoregressive model. In particular, it is a novel method to guide the decision of which electrodes should be clustered together with bayesian variable selection techniques used to select significant set of bases coefficients in each cluster.

The marginal distribution of our model (Equation 3.2) follows the enriched functional dirichlet process proposed by Scarpa and Dunson (2013); to see that, for each s,

$$\begin{aligned} P^{(s)} &= \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{k=1}^{N_a} \sum_{u_1,\ldots,u_s=k,\ldots,u_S} \pi_{u_1,\ldots,u_s=k,\ldots,u_S} \delta_{\theta_{u_1}} \ldots \delta_{\theta_{u_s=k}} \ldots \delta_{\theta_{u_S}} \\ &= \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{k=1}^{N_a} p_{a,k} \delta_{\theta_{a,k}}, \end{aligned} \tag{3.5}$$

where $p_{a,k} = \sum_{k=1}^{N_a} Pr\left(\sum_{l=1}^{k-1} p_l < F^{(s)}(\eta_s) < \sum_{l=1}^{k} p_l\right)$ and $\theta_{a,k}$ is a functional atom draw from $P_{0,a}$, a base measure for our functional dirichlet process, for all a specified as in Section 2.8. In 3.5 is used that $\pi_{u_1,\ldots,u_{s-1},\ldots,u_S} = \pi_{u_1,\ldots,u_{s-1},\cdot,u_{s+1},\ldots,u_S} = \sum_{u_1,\ldots,u_s=k,\ldots,u_S} \pi_{u_1,\ldots,u_s=k,\ldots,u_S}$.

Moreover, if $\Pi_a = 1$ and $M = 1$ Equation 3.2 is a functional dirichlet process over a lattice and it is an extension of the discrete process by Li et al. (2014).

The first and second moments of the process is given by

$$\mathrm{E}(\theta_s|P_S) = \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{k=1}^{N_a} p_{a,k}(s) \delta_{\theta_{a,k}}(s) \tag{3.6}$$

$$\mathrm{var}(\theta_s|P_S) = \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{k=1}^{N_a} p_{a,k}(s) \delta^2_{\theta_{a,k}}(s) - \left\{ \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{k=1}^{N_a} p_{a,k}(s) \delta_{\theta_{a,k}}(s) \right\}^2, \tag{3.7}$$

with covariance

$$\begin{aligned}
&\mathrm{Cov}(\theta_r, \theta_s | P_S) = \sigma_r^2 \sigma_s^2 \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{l=1}^{N_a} Pr(u_r = u_s = l) \\
= \;& \sigma_r^2 \sigma_s^2 \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{l=1}^{N_a} Pr\left(\sum_{k=1}^{l-1} p_k < F^{(r)}(\eta_r) < \sum_{k=1}^{l} p_k, \sum_{k=1}^{l-1} p_k < F^{(s)}(\eta_s) < \sum_{k=1}^{l} p_k \right) \\
= \;& \sigma_r^2 \sigma_s^2 \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{l=1}^{N_a} p_l Pr\Bigg(F^{(r)^{-1}} \sum_{k=1}^{l-1} p_k < \eta_r < F^{(r)^{-1}} \sum_{k=1}^{l} p_k \mid F^{(s)^{-1}} \sum_{k=1}^{l-1} p_k < \eta_s < \\
& F^{(s)^{-1}} \sum_{k=1}^{l} p_k \Bigg),
\end{aligned} \tag{3.8}$$

where (η_r, η_s) is distributed as a bivariate normal distribution with covariance matrix given by the CAR model.

3.2 Posterior Computation

Given the cluster allocation variables $\mathscr{K}$ and $\mathscr{C}$ introduced in Section 2.9, the MCMC algorithm, after initial values are set, iterate with the following steps (see Appendix A for some calculations):

1. *update* $\mathbf{\Pi}$ by sampling from the full conditional posterior

$$(\mathbf{\Pi} | \mathscr{K}, \mathscr{C}, \boldsymbol{y}) \sim \mathrm{Diri}\left(a_{\Pi,1} + \sum_{s=1}^{S} 1(\mathscr{C}_{\mathscr{K}_s} = 1), \;\ldots, a_{\Pi,\sharp\mathscr{A}} + \sum_{s=1}^{S} 1(\mathscr{C}_{\mathscr{K}_s} = \sharp\mathscr{A}) \right), \tag{3.9}$$

2. *update parameters for* $\epsilon(s,t)$. For computational reasons, write the Student-t distribution as $\epsilon(s,t) \sim N(0, \psi_{st}^{-1}\sigma_\epsilon^2)$, see, e.g., Congdon (2001) for a proof of this result. Set $\psi_{st} \sim \mathrm{gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, assume $\sigma_\epsilon^{-2} = \tau_\epsilon \sim \mathrm{gamma}(a_\tau, b_\tau)$ and $\nu \sim \mathrm{gamma}(a_\nu, b_\nu)$. (a) Update ψ_{st}

$$(\psi_{st} | \boldsymbol{V}, \tau_\eta, \boldsymbol{\eta}, \tau_\epsilon, \Theta, \boldsymbol{y}) \sim \mathrm{gamma}\left(\frac{\nu}{2} + \frac{1}{2}, \frac{\nu}{2} + \frac{1}{2}\sigma_\epsilon^{-2}\{y(s,t) - f(s,t)\}^2 \right) \tag{3.10}$$

(b) Update ν using Metropolis random walk, by generating a candidate

$$\nu^* \sim N(\nu^{t-1}, 0.25), \tag{3.11}$$

accept with probability

$$\min\left\{1, \frac{\text{gamma}(\nu^*; a_\nu, b_\nu)\prod_{s=1}^{S}\prod_{t=1}^{T}\text{gamma}\left(\psi_{st}; \frac{\nu^*}{2}, \frac{\nu^*}{2}\right)}{\text{gamma}(\nu^{(t-1)}; a_\nu, b_\nu)\prod_{s=1}^{S}\prod_{t=1}^{T}\text{gamma}\left(\psi_{st}; \frac{\nu^{(t-1)}}{2}, \frac{\nu^{(t-1)}}{2}\right)}\right\}, \tag{3.12}$$

(c) Update $\tau_\epsilon = \sigma_\epsilon^{-2}$ by sampling from the full conditional posterior

$$(\tau_\epsilon | \boldsymbol{V}, \tau_\eta, \Theta, \boldsymbol{S}, \boldsymbol{y}, \boldsymbol{\eta}) \sim \text{gamma}\left(a_\tau + \frac{1}{2}\sum_{s=1}^{S} T, b_\tau + \frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\{y(s,t) - f(s,t)\}^2\right). \tag{3.13}$$

3. *update* $(\boldsymbol{\eta}|\Theta, \boldsymbol{V}, \tau_\eta, \tau_\epsilon, \boldsymbol{y})$ by sampling candidate η_s^* from $N(\eta_s, K_\eta)$, compute the corresponding $\boldsymbol{u}^*$ through

$$u_s^* = \sum_{j=1}^{K} jI\left(\sum_{k=1}^{j-1} p_k < F^{(s)}(\eta_s^*) < \sum_{k=1}^{j} p_k\right), \tag{3.14}$$

where $I(\cdot) = 1$ if η_s^* belongs to cluster k, $p_k = \Pi_a V_{a,k}\prod_{k<h}(1 - V_{a,h})$, for a, k : $I_k(a, k) = 1$. Accept the candidate η^* with probability

$$\min\left\{1, \frac{\exp(-\frac{1}{2}\boldsymbol{\eta}^{*\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}^*)\prod_{t=1}^{T}\phi\left(y(s,t); \Theta_{u_s^*}, \psi_{st}^{-1}\tau_\epsilon^{-1}\right)}{\exp(-\frac{1}{2}\boldsymbol{\eta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\eta})\prod_{t=1}^{T}\phi\left(y(s,t); \Theta_{u_s}, \psi_{st}^{-1}\tau_\epsilon^{-1}\right)}\right\}, \tag{3.15}$$

with $\boldsymbol{\eta}^* = \{\eta_1, \ldots, \eta_s^*, \ldots, \eta_S\}$.

4. *update the* $(\boldsymbol{V}|\tau_\eta, \Theta, \boldsymbol{\eta}, \tau_\epsilon, \boldsymbol{y})$ *the stick-breaking weights*. Sample candiate $\boldsymbol{V}^*$ from beta(a_V, b_V), where $a_V = max(\delta + \varepsilon V, 1)$ and $b_V = max(\delta + \varepsilon(1 - V), 1)$. This is the random beta walk. Compute the corresponding $\boldsymbol{p}^*$ and $\boldsymbol{u}^*$ where

$$u_s^* = \sum_{j=1}^{K} jI\left(\sum_{k=1}^{j-1} p_{a,k} < F^{(s)}(\eta_s^*) < \sum_{k=1}^{j} p_{a,k}\right), \tag{3.16}$$

and $p_{a,1} = \Pi_a V_{a,1}$; $p_j = \Pi_a V_{a,k} \prod_{k<j}(1 - V_{a,j})$. Then accept the candidate $\boldsymbol{V}^*$ with probability

$$\min \left\{ 1, \frac{\prod_{k=1}^{K}(1 - V_k^*)^{\alpha-1} \prod_{s=1}^{S} \prod_{t=1}^{T} \phi\left(y(s,t); \Theta_{u_s^*}, \psi_{st}^{-1}\tau_\epsilon^{-1}\right)}{\prod_{k=1}^{K}(1 - V_k)^{\alpha-1} \prod_{s=1}^{S} \prod_{t=1}^{T} \phi\left(y(s,t); \Theta_{u_s}, \psi_{st}^{-1}\tau_\epsilon^{-1}\right)} \right\}. \tag{3.17}$$

5. *update* $(\tau_\eta|\Theta, \boldsymbol{V}, \boldsymbol{\eta}, \tau_\epsilon, \boldsymbol{y})$ the spatial dispersion parameter. Sample from

$$\text{gamma}\left(\frac{n}{2} + c_\tau, \frac{1}{2}\boldsymbol{\eta}'(D - \rho W)\boldsymbol{\eta} + d_\tau\right) \tag{3.18}$$

which is the conjugate gamma full conditional distribution for τ_η.

6. *update* $\theta_{a,k}(\cdot)$, for $k = 1, \ldots, N_a$, by updating the basis coefficients, $\boldsymbol{\beta}_{a,k}$ by sampling from the full conditional posterior distribution

$$(\beta_{kl}|\boldsymbol{V}, \boldsymbol{\beta}_{a,k(-l)}, \tau_\eta, \boldsymbol{\eta}, \tau_\epsilon, \boldsymbol{y}) \sim \hat{\nu}_{kl}\delta_0(\cdot) + (1 - \hat{\nu}_{kl})\phi_{A_{kl}}(\cdot; \hat{\beta}_{kl}, \widehat{\sigma}^2_{\beta_{kl}}) \tag{3.19}$$

where $\phi_{Akl}(\cdot, \mu, \sigma^2)$ denotes the truncated gaussian density to the region A resulting in $\theta_{a,k}$ belonging to the feature class a and

$$\widehat{\sigma}^2_{\beta_{kl}} = \left(c_\nu^{-1} + \sum_{s:S_s=k} \sum_{t=1}^{T} \psi_{st}\tau_\epsilon b_l(t_s)^2\right)^{-1},$$

$$\hat{\beta}_{kl} = \widehat{\sigma}^2_{\beta_{kl}} \left(\sum_{s:S_s=k} \sum_{t=1}^{T} \psi_{st}\tau_\epsilon y(s,t)^{(k,-l)} b_l(t_s)\right),$$

$$\text{where } y(s,t)^{(k,-l)} = y(s,t) - \sum_{m \neq l} \beta_{km} b_m(t_s),$$

$$\hat{\nu}_{kl} = \left(1 + \frac{\nu_0}{1 - \nu_0} \frac{\phi_{A_{kl}}(\cdot; 0, c_\nu)}{\phi_{A_{kl}}(\cdot; \hat{\beta}_{kl}, \widehat{\sigma}^2_{\beta_{kl}})}\right)^{-1}.$$

7. *update the probability* of including a basis function by sampling as follows:

$$(\nu_0|-) \sim \text{beta}\left(a_\nu + \sum_k \sum_l 1(\beta_{kl} = 0), b_\nu + \sum_k \sum_l 1(\beta_{kl} \neq 0)\right). \tag{3.20}$$

8. *update* α as in Escobar and West (1995) - Assuming gamma(a_α, b_α) hyperprior for α, with the gamma paremeterized to have mean $a_\alpha b_\alpha$ and variance $a_\alpha b_\alpha^2$ the conditional

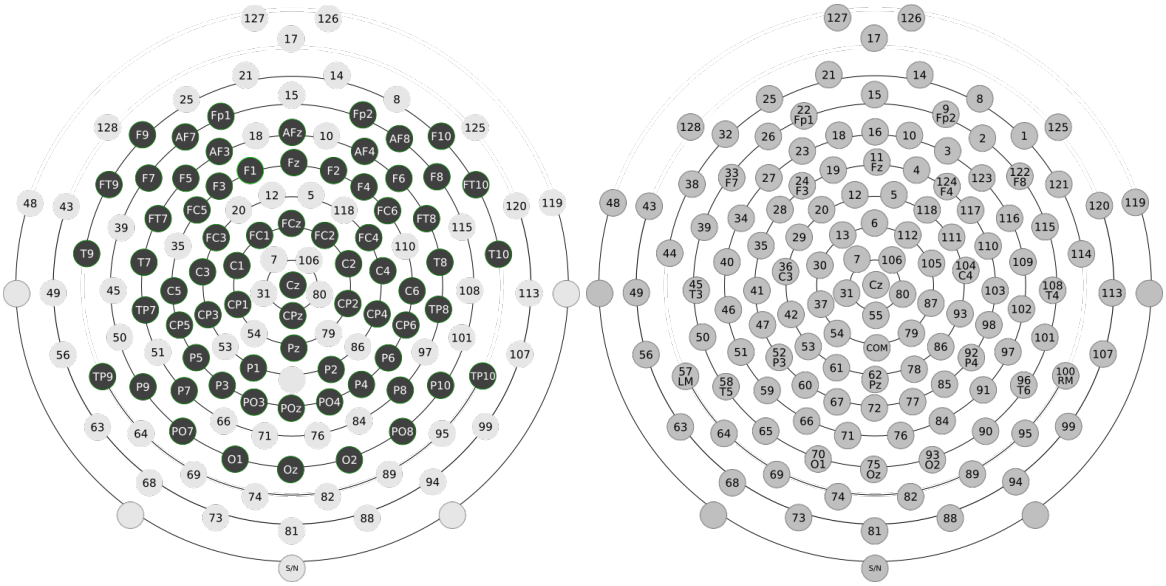
posterior is

$$\begin{aligned}(\alpha|\eta,k) \sim \pi_\eta \text{gamma}\left(a_\alpha + k, b_\alpha - \log(\eta)\right) \\ + (1-\pi_\eta)\text{gamma}\left(a_\alpha + k - 1, b_\alpha - \log(\eta)\right),\end{aligned} \tag{3.21}$$

with $\pi_\eta/(1-\pi_\eta) = (a_\alpha + k - 1)/(K\{b_\alpha - \log(\eta)\})$ and $(\eta|\alpha,k) \sim \text{beta}(\alpha+1, K)$.

3.3 Simulation study

We simulated data for 71 electrodes, $s = 1,\ldots,71$, using the 10-10 international system. In Figure 3.1a is illustrated the position and approximate correspondence between the international 10-10 system, dark gray color, with those of the 128-channel, illustrated in Figure 3.1b, as in Luu and Ferree (2005). We use a smaller set of data, 71 electrodes, in the simulation study instead 128 electrodes, as in the real data, to assess the performance of the model in detect activated regions when a smaller set of data is available. The 71 electrodes were di-



(a) Layout 10-10 system, 71 electrodes, equivalent on the 128-channel HydrcoCel

(b) Layout 128 channels

Fig. 3.1 Geodesic sensor nets.

vided in 3 regions arbitrary chosen to be the frontal ($r_b = 1$, 32 electrodes), parietal ($r_b = 2$, 18 electrodes) and occipital ($r_b = 3$, 21 electrodes) brain lobes though in the real data the activated regions might not be coincident with this configuration. Specifically, following

Jansen and Rit (1995) we simulate data from

$$y(s,t) = h(t)^{(r_b)} \otimes \text{Sigm}(v)^{(r_b)} + \epsilon(s,t),$$

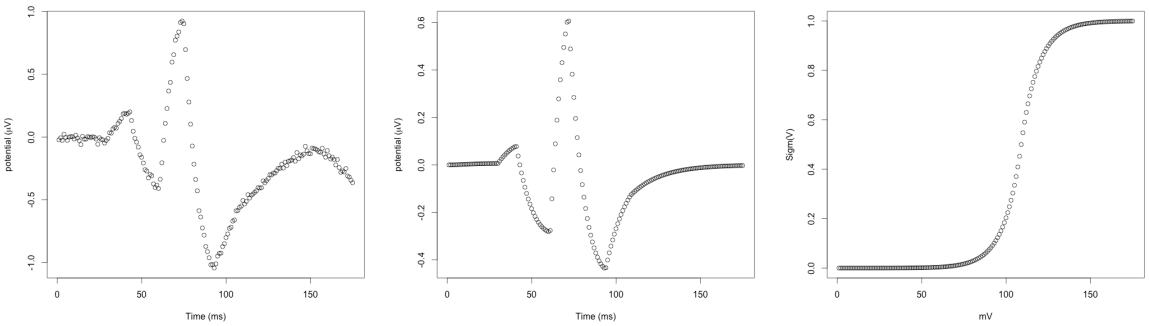
where $y(\cdot)$ is a functional ERP like response, $\otimes$ denotes convolution,

$$h(t) = \begin{cases} Aat\, e^{-at} & t >= 0 \\ -0 & t < 0 \end{cases}$$

is the impulse response, with A the maximum amplitude of the excitatory or inhibitory neuron, a the lumped time constant of passive membrane currents and other spatially distributed delays in the dendritic network,

$$\text{Sigm}(v) = \frac{2e_0}{1 + e^{\epsilon(v_0 - v)}} \tag{3.22}$$

a sigmoid function, used for the voltage-to-pulse conversion process with e_0 the maximum firing rate of the neural population, v_0 the postsynaptic membrane potential for which a 50% firing rate is achieved, and ϵ the steepness of the sigmoidal transformation and $\epsilon(s,t)$ is a random noise specific to electrode s and time $t = 1, \ldots, 175$. In Figure 3.2 are illustrated each component of the model.

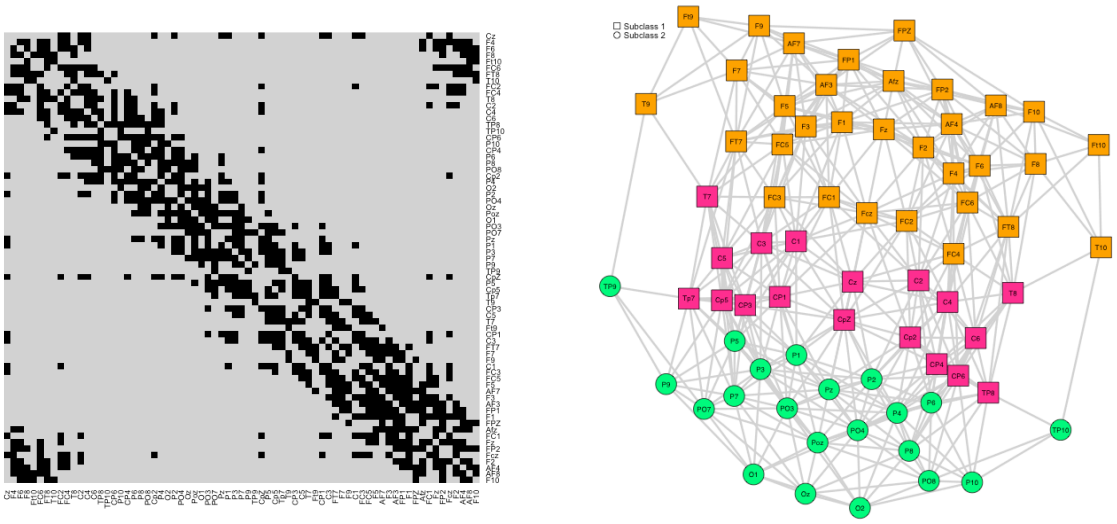


Outcome $y(1,\cdot)$ Impulse response $h(\cdot)$ Sigmoid function

Fig. 3.2 Simulated data: ERP like functional data, $h(\cdot)$ and Sigm$(\cdot)$ functions.

The adjacency matrix W in the covariance matrix of the CAR prior in Equation 3.4 is of clique 2, i.e, $i \sim j$ if and only if $j \sim i$ (see Section 2.5 for details). In Figure 3.3a is illustrated the adjacency matrix with neighborhood structure specified, as illustrated in Figure 3.3b, by choosing a threshold distance of 0.77 mm between electrodes where the range distance is 0.215–1.997 mm; this threshold was chosen by visual inspection since with this distance the electrodes seems to be reasonably connected. Additionally, the orange and pink colors

identify the frontal and parietal brain lobes within subclass 1 and the green color identify the occipital lobe within subclass 2.



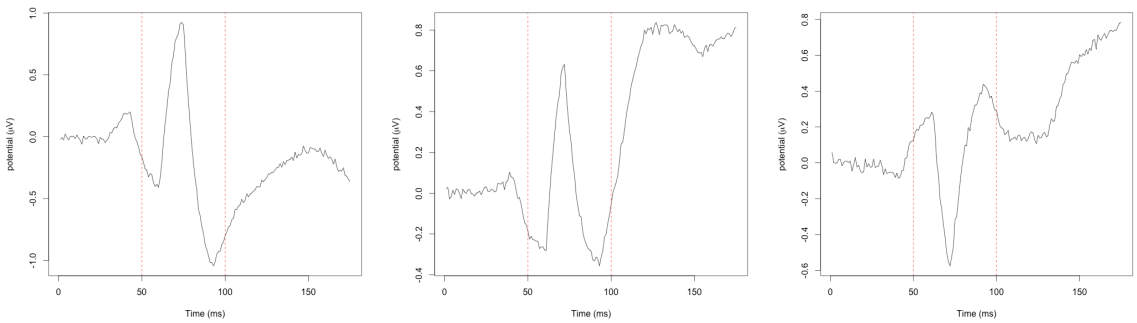
(a) Adjacency matrix (b) Neighborhood structure

Fig. 3.3 Simulated data: adjacency matrix and neighborhood structure for spatial configuration of simulated data.

The remaining parameters of the model was chosen such that $c_\tau = 1$ favor a model with small set of basis functions, $a_\tau = b_\tau = a_\nu = b_\nu = a_\alpha = b_\alpha = 1$ are non informative priors and $a_\Pi = (0.7, 0.3)^T$ informative prior to the proportion of curves falling into each subclass. The meaning of this prior is the following: 70% of the ERP waveforms in the simulated data present maximum value between the time interval 50–100 ms, see illustration in Figures 3.4a and 3.4b, in opposition to the remaining 30% of the curves that present minimum value within this interval, as illustrade in Figure 3.4c. This configuration divide the simulated data in symmetrical subclasses, see Section 2.1.

In the following we show the results after run the Gibbs sampling algorithm for 25,000 iterations after a burn-in period of 5,000. Gelman-Rubin diagnostics (Gelman and Rubin, 1992) and trace plots of the parameters showed no evidence against convergence. Binder's loss was used to classify the waveforms within regions and subclasses.

The posterior mean waveforms for electrodes belonging to each brain region is illustrated in Figure 3.5. Note that waveforms belonging to regions 1 and 2 are within subclass 1 while waveforms from region 3 are within subclass 2. The gray curves represent the simulated data and the black ones the posterior mean for each curve. The posterior mean of $\mathbf{\Pi}$ was $\widehat{\Pi}_1 = 0.722$ (ci: 0.721–0.723) and $\widehat{\Pi}_2 = 0.278$ (ci: 0.277–0.279).

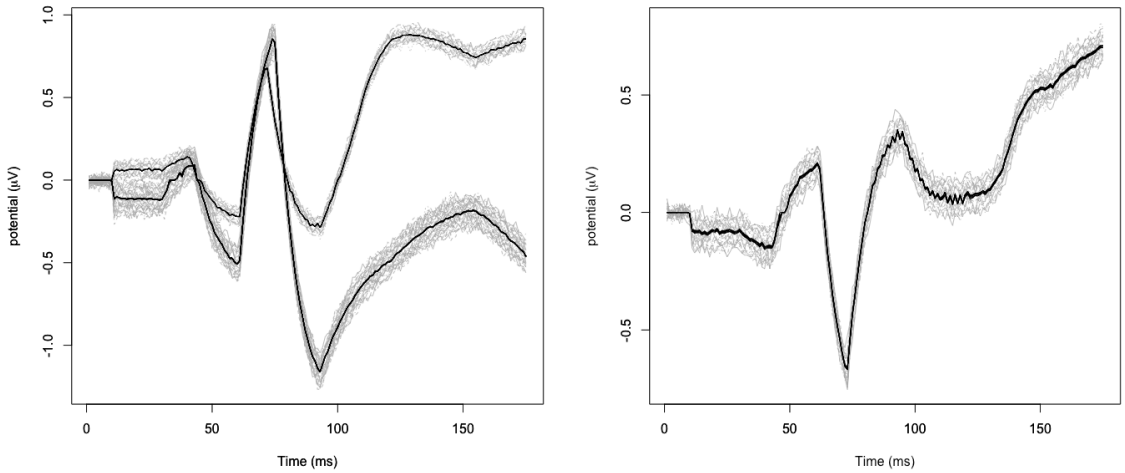


(a) Region 1 (b) Region 2 (c) Region 3

Fig. 3.4 Simulated data: Example of ERP like functional data for each region.

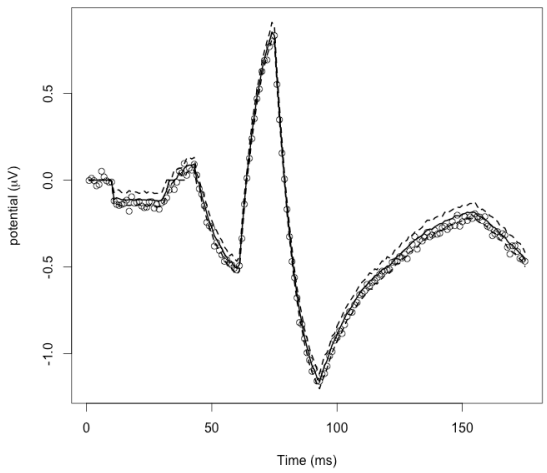
Figure 3.6 shows the posterior mean and highest posterior density interval for selected electrodes within each of the 3 regions where regions 1 and 2 fall into subclass 1 and region 3 into subclass 2.

With our approach we were able to detect the boundaries between the regions and consequently correctly group the electrodes within the frontal, parietal and occipital brain lobes as in the simulated data. Additionally, the use of feature classes enrich the interpretation of the results as the classification of regions into subclasses allows us to locate symmetries in the brain response to stimuli.

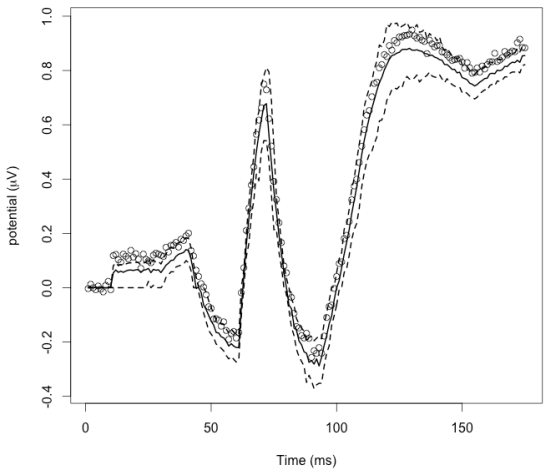


(a) Subclass 1 (b) Subclass 2

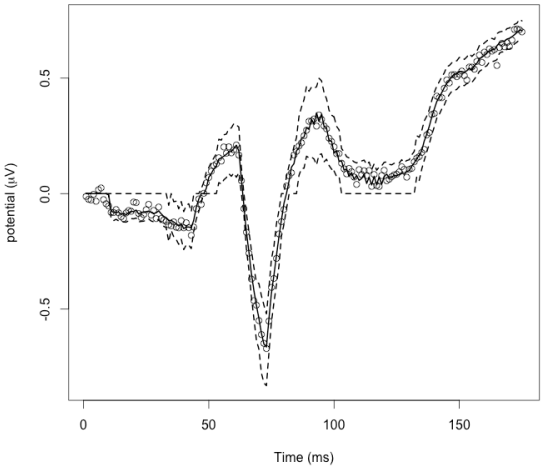
Fig. 3.5 Simulated data: posterior mean waveforms for electrodes within each subclass along with the simulated data in gray color.



(a) Subclass 1
electrode 21, region 1



(b) Subclass 1
electrode 61, region 2



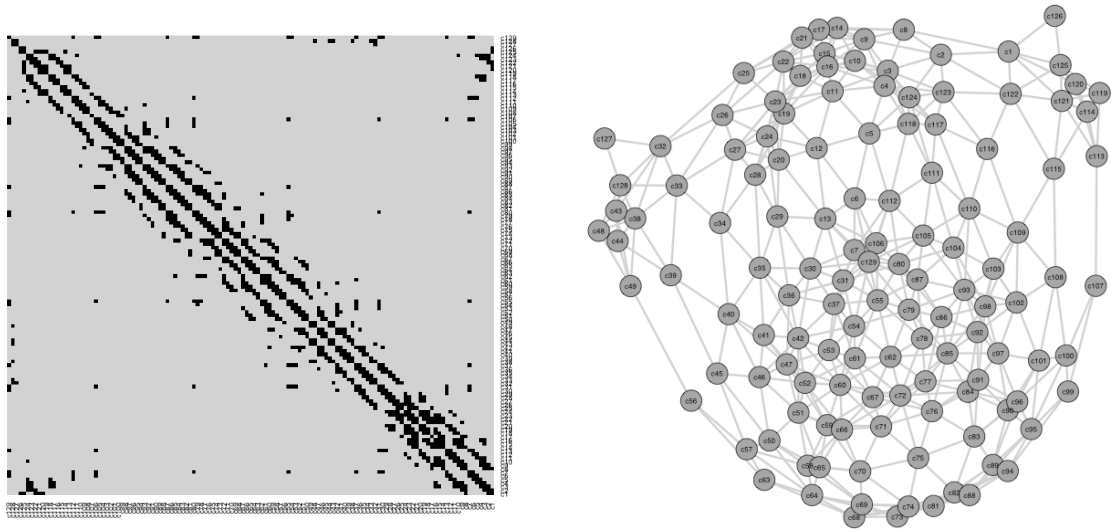
(c) Subclass 2
electrode 43, region 3

Fig. 3.6 Simulated data: posterior mean and highest posterior density.

3.4 Application: ERP data in neuroimagin

We applied our model to the ERP data in neuroimaging presented in Section 2.1. Goals in this analysis is to model the waveforms, group the electrodes and classify groups as belonging to subclass 1 or 2 depending on the curves have their maximum value within the time interval 50–100 ms or minimum value within this interval, respectively.

The adjacency matrix W in the covariance matrix of the CAR prior in Equation 3.4 is illustrated in Figure 3.7a with neighborhood structure specified, as illustrated in Figure 3.7b, by choosing a threshold distance of 0.5 mm between electrodes considering the range of distance is 0.157–2 mm; this threshold was chosen by visual inspection since with this distance the electrodes seems to be reasonably connected. The remaining parameters of the



(a) Adjacency matrix (b) Neighborhood structure

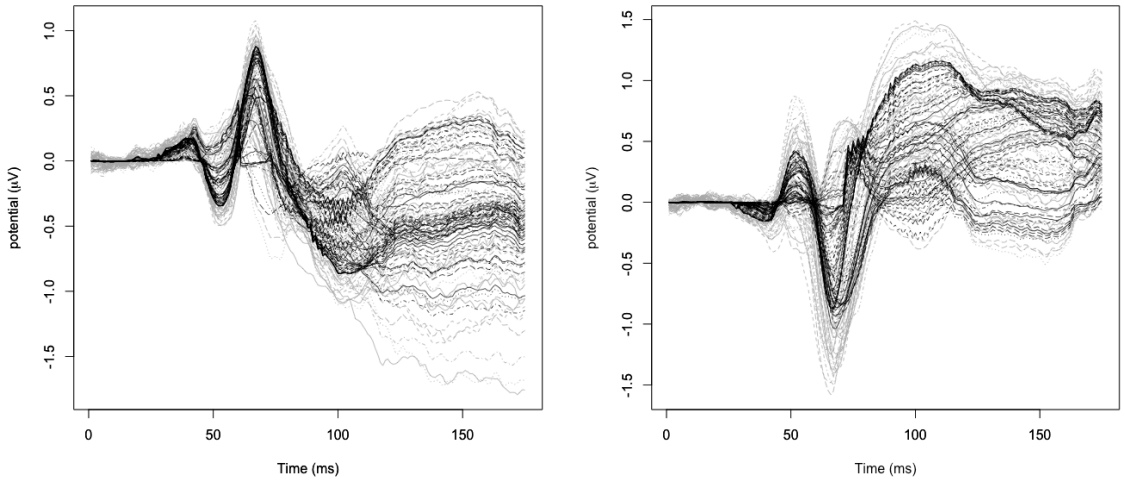
Fig. 3.7 Real data: adjacency matrix and neighborhood structure for spatial configuration of real data.

model was set in which $c_\tau = 1$ favor a model with smaller set of basis functions, $a_\tau = b_\tau = a_\nu = b_\nu = a_\alpha = b_\alpha = 1$ are non informative priors and $a_\Pi = (0.7, 0.3)^T$ informative prior to the proportion of curves falling into each subclass.

In the following we show the results after run the Gibbs sampling algorithm for 22,000 iterations after a burn-in period of 8,000. Gelman-Rubin diagnostics (Gelman and Rubin, 1992) and trace plots of the parameters showed no evidence against convergence. Binder's loss (Binder, 1978; Lau and Green, 2007) was used to classify the waveforms within regions and subclasses.

The posterior mean waveforms for electrodes belonging to each brain region is illustrated In Figure 3.8. Note that waveforms belonging to regions 1 and 2 are within subclass 1 while waveforms from region 3 are within subclass 2. The gray curves represent the simulated data and the black ones the posterior mean for each curve. The posterior mean of $\mathbf{\Pi}$ was $\widehat{\Pi}_1 = 0.5152$ (ci: 0.5147–0.5158) and $\widehat{\Pi}_2 = 0.4848$ (ci: 0.4842–0.4853). Figure 3.9 shows the

posterior mean and highest posterior density interval for selected electrodes within subclass 1 and subclass 2.

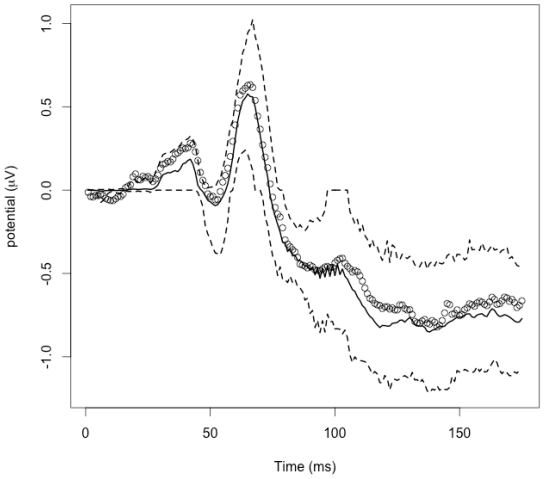


(a) Subclass 1 (b) Subclass 2

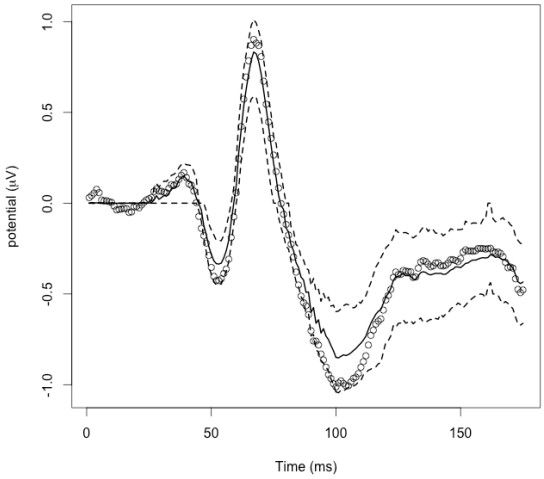
Fig. 3.8 Real data: posterior mean waveforms for electrodes within each subclass along with the real data in gray color.

With our approach we were able to detect the boundaries between the regions and consequently group the electrodes, as illustrated in Figure 3.10, within 6 regions (squares in cyan, yellow, red, green, pink and blue) into subclass 1 and 5 regions (circles in yellow, green, cyan, red and blue) into subclass 2. Additionally, the use of feature classes enrich the interpretation of the results as the classification of regions into subclasses allows us to locate symmetries in the brain response to stimuli.

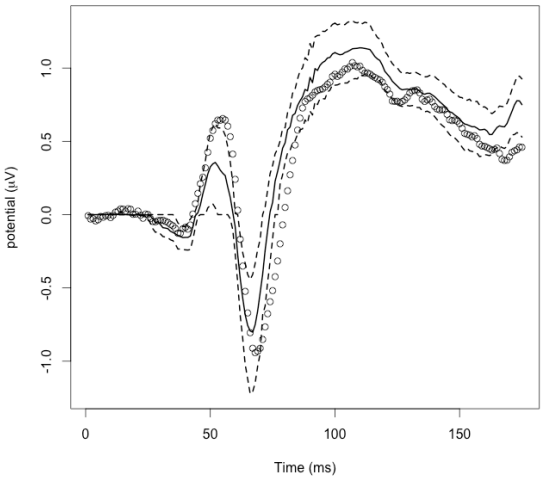
Versace et al. (2011) analyzed the same dataset by computing ERP differences among stimuli using randomization tests on time regions identified by temporal principal components analysis. Though the author's interest is not group electrodes and do not able us to compare the results obtained with our model it is worth to mention that the activated regions are in concordance with their findings as we detect boundaries between the frontal, parietal and occipital brain lobes, and additionally detect subregions within each lobe. Group electrodes is a novel approach and potentially interesting though difficult to interpret biologically each detected region.



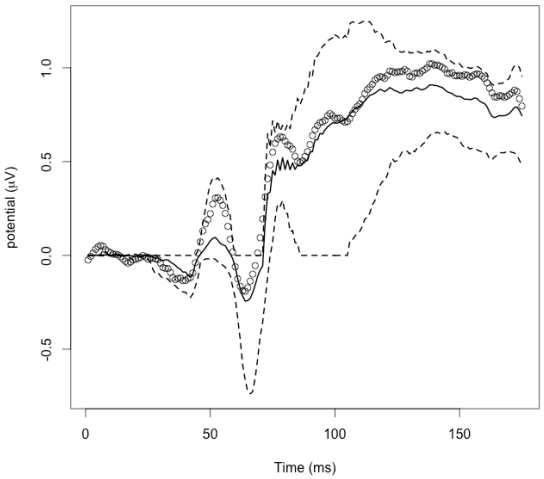
(a) Subclass 1
electrode 1



(b) Subclass 1
electrode 19



(c) Subclass 2
electrode 58



(d) Subclass 2
electrode 86

Fig. 3.9 Real data: posterior mean and highest posterior density.

3.5 Discussion

In this Chapter we proposed a Bayesian nonparametric approach that allows to group spatial correlated curves observed over areal units of a lattice. Many phenomenons occur over a lattice space with well know neighborhood structure. Include neighborhood information

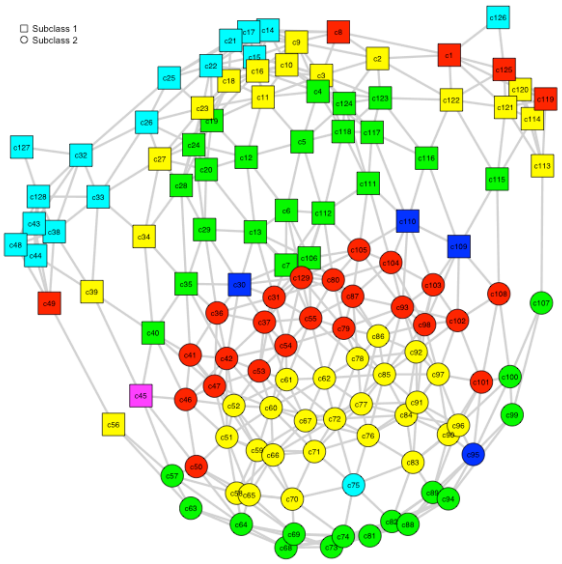


Fig. 3.10 Real data: resulting configuration of electrodes within regions and subclasses. Different colors within classes stand for regions.

into the Dirichlet Processes improve in interpretation of the underlying phenomenon since clusters can be interpreted as regions over the considered space.

In our motivating application in neuroimaging we see the configuration of the electrodes in the ERP as areal units of a lattice. Others examples of applications include phenomenons observed over discrete points in a geographical map; one could see the daily temperature collected by different weather stations as curves and there euclidean distance as an adjacency matrix to be incorporated into our proposed model.

A desirable extension of our model is allow clustering of the curves taking in consideration time-windows since in different times the ERP waveforms have different activated regions.

Chapter 4

A Bayesian Nonparametric approach for the analysis of the balancing marketing

Motivated by the the Italian natural gas balancing platform presented in Section 2.2, in this Chapter we propose a Bayesian nonparametric approach that allows to group time dependent curves within feature classes. A probit dynamic model defined over a partitioned functional sample space guide the time dependence. Feature classes allow to use different base measures for each partition of the sample space improving in interpretation of the results.

4.1 Proposed Model

For simplicity consider f a generic monotonic bounded function since both supply and demand curves boast similar features. Let $y(t,q)$ denote the supply or demand curves in the Italian Natural Gas Balancing Platform dataset observed at days $t = 1, \dots, T$ for quantities $q = 1, \dots, Q$. We assume that

$$y(t,q) \;=\; f(t,q) + \epsilon(t,q), \tag{4.1}$$

where the collection $\{f(t,q)\}_{q=1}^{Q}$ is considered to be the realization of a random function $f(t,\cdot) : \mathscr{R}^+ \to \mathscr{R}$ and $\epsilon(t,q) \sim t_\nu(\sigma^2)$ is a temporal error process with ν the degrees of freedom and σ the scalar parameter of a Student-t distribution. The functions f's are assumed to be time dependent and clustered together by an underlying Functional Dependent Dirichlet Process (FDDP), viz

$$\boldsymbol{f} \;=\; \{f_t\}_{t=1}^{T} | P_t \sim P_t \tag{4.2}$$

with

$$P_t = \sum_{k=1}^{K} p_{k(t)} \delta_{\theta_k} \tag{4.3}$$

a FDDP embedded with stochastic weights $p_{k(t)}$ appropriately chosen to guide the time dependence.

To guide the dependence of the FDDP in Equation 4.3 we make use of a DLM as defined in Section 2.6. Let $z_t \sim N(\zeta_t, \nu_t)$, $t = 1, \ldots, t$, a latent variable to the FDDP, govern the time dependence over the sample space Ω. Define,

$$z_t = \zeta_t + \nu_t, \quad \nu_t \overset{iid}{\sim} N(0, V_t) \tag{4.4}$$

$$\zeta_t = \zeta_{t-1} + \omega_t, \quad \omega_t \overset{iid}{\sim} N(0, W_t) \tag{4.5}$$

where Equations 4.4 and 4.5 are the observation and state equation, respectively. Heckman (1981) proposed the use of Dynamic probit model to study state dependence in heterogeneous system using conditional probability to past events and dummy variables to indicate the state of the subject. Dunson and Rodríguez (2011) used Dynamic probit model to propose a new class of DDP as follows.

Define,

$$Pr(\mathscr{K}_t = k) = Pr(z_{kt} < 0, \ z_{rt} > 0 \text{ for } r < k) \tag{4.6}$$

$$= \Phi(\zeta_{kt}) \prod_{r<k} \left\{1 - \Phi(\zeta_{rt})\right\} = p_{kt} \tag{4.7}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function for the standard normal distribution and $\Phi(\zeta_{Kt}) = 1$ guarantee $\sum_k^K p_{kt} = 1$ almost surely and $\mathscr{K}$s the allocation variables defined in Section 2.9. With this carachterization the FDDP can be written as

$$P_t = \sum_{k=1}^{K} \Phi(\zeta_{kt}) \prod_{r<k} \left\{1 - \Phi(\zeta_{rt})\right\} \delta_{\theta_k}, \tag{4.8}$$

where θ_k is drawn from a base measure P_0.

It is well know that the clustering allocation of the FDDP just defined above would be influenced by two sources: the time dependence and the shape of the curves. As illustration in Figure 4.1 note that the function observed at day 1 is quite different of that observed in day 2. As consequence, though they are time dependent, their difference in shape is decisive in the clustering allocation difficulting the inference; to improve the cluster configuration interpretation we enrich our model with feature classes as explained in the next section.

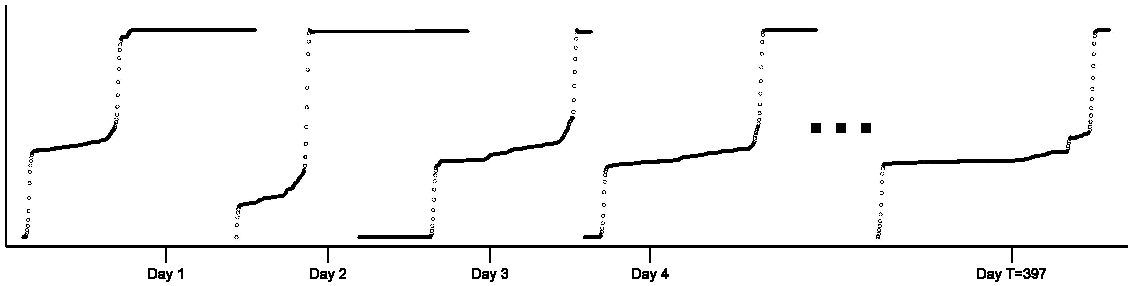


Fig. 4.1 Time dependent functional time series

4.1.1 Time Dependence Specification Over Partitioned Function Space

Recalling the feature classes introduced in Section 2.7 and allocation variables introduced in Section 2.9, define the DLM above over the partitioned function space $\Omega = \cup_{a=1}^{\sharp\mathscr{A}}\Omega_a$ as previously defined. Let $z_{a,t} \sim N(\zeta_{a,t}, \nu_{a,t})$, $t = 1, \ldots, t$, a latent variable to the process, govern the time dependence over Ω_a for each a. Define,

$$z_{a,t} = \zeta_{a,t} + \nu_{a,t}, \quad \nu_{a,t} \overset{iid}{\sim} N(0, V_{a,t}) \tag{4.9}$$

$$\zeta_{a,t} = \zeta_{a,t-1} + \omega_{a,t}, \quad \omega_{a,t} \overset{iid}{\sim} N(0, W_{a,t}) \tag{4.10}$$

where Equations 4.9 and 4.10 are the observation and state equation, respectively.

Define,

$$Pr(\mathscr{K}_t = k) = Pr(z_{a,kt} < 0,\ z_{a,rt} > 0 \text{ for } r < k \mid \Pi_a) \tag{4.11}$$

$$= \Pi_a \Phi(\zeta_{a,kt}) \prod_{r<k} \{1 - \Phi(\zeta_{a,rt})\} = p_{a,kt} \tag{4.12}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function for the standard normal distribution, Π_a is the proportion of curves falling within feature class a, $\Phi(\zeta_{a,kt}) = 1$ for $k = N_a$ for each a guarantee $\sum_k^K p_{a,kt} = 1$ almost surely.

4.1.2 Proposed Prior

Consider the model 4.1 and as before assign $\boldsymbol{f} = \{f_t\}_{t=1}^T | P_t \sim P_t$ such that

$$P_t = \sum_{a=1}^{\sharp\mathscr{A}} \Pi_a \sum_{k=1}^{N_a} p_{a,kt} \delta_{\theta_{a,k}}, \qquad \theta_{a,k} \sim P_{0,a} \tag{4.13}$$

$$p_{a,kt} = \Phi(\zeta_{a,kt}) \prod_{r<k} \left\{1 - \Phi(\zeta_{a,rt})\right\} \tag{4.14}$$

where $\zeta_{a,kt}$ is the state in the DLM defined in 4.9 and 4.10, $\Phi(\cdot)$ denotes the cumulative distribution function for the standard normal distribution and Π_a is the proportion of curves falling within feature class a. To truncate the stick-breaking prior to K terms set $\Phi(\zeta_{a,kt}) = 1$ for $k = N_a$, for each a, to guaranty, by Lemma 1 in Scarpa and Dunson (2013), $\sum_{k=1}^{K} p_k = 1$ almost surely, where $K = (1, \ldots, N_1, N_1 + 1, \ldots, N_1 + N_2, \ldots, \sum N_{a\,a=1}^{\sharp\mathscr{A}-1}, \ldots, N_{\sharp\mathscr{A}})$, with $\boldsymbol{N} = \{N_a\}_{a=1}^{\sharp\mathscr{A}}$ the truncation level of each class. The base measure $P_{0,a}$ is specified as in Section 2.8.

Moreover, if $\Pi_a = 1$ and $M = 1$ Equation 4.13 is a functional dependent dirichlet process and it is a functional extension of the probit stick breaking process by Dunson and Rodríguez (2011).

4.2 Posterior Computation

Given the cluster allocation variables $\mathscr{K}$ and $\mathscr{C}$ introduced in Section 2.9, the MCMC algorithm, after initial values are set, iterate with the following steps (see Appendix A for some calculations):

1. *update* $\boldsymbol{\Pi}$ by sampling from the full conditional posterior

$$(\boldsymbol{\Pi}|\mathscr{K}, \mathscr{C}, \boldsymbol{y}) \sim \text{Diri}\left(a_{\Pi,1} + \sum_{t=1}^{T} 1(\mathscr{C}_{\mathscr{K}_t} = 1),\ \ldots, a_{\Pi,\sharp\mathscr{A}} + \sum_{t=1}^{T} 1(\mathscr{C}_{\mathscr{K}_t} = \sharp\mathscr{A})\right), \tag{4.15}$$

2. *update the dynamic probit weights* by updating $\zeta_{a,kt}$, and $\boldsymbol{W}_t$ for $a = 1, \ldots, \sharp\mathscr{A}$ and $h = 1, \ldots, N_a$,

   (a) update $\boldsymbol{\zeta}_{a,kt}$ and $\boldsymbol{W}_t$ by FFBS algorithm;

   (b) update $p_{a,kt}$, using the updated values of $\boldsymbol{\zeta}_{a,kt}$,

$$p_{a,kt} = \Pi_a \Phi(\zeta_{a,ht}) \prod_{r<h} \left(1 - \Phi(\zeta_{a,rt})\right) \tag{4.16}$$

3. *update σ^2 parameter for $\epsilon(t,q)$.* Assuming $\sigma^{-2} = \tau \sim \text{gamma}(a_\tau, b_\tau)$, update τ by sampling from the full conditional posterior

$$(\tau|\boldsymbol{\mathscr{K}}, \boldsymbol{\mathscr{C}}, \boldsymbol{\Pi}, \boldsymbol{V}, \boldsymbol{\Theta}, \boldsymbol{t}, \boldsymbol{y}) \sim \text{gamma}\left(a_\tau + \frac{TQ}{2}, b_\tau + \frac{1}{2}\sum_{t=1}^{T}\sum_{q=1}^{Q}\{y(t,q) - f(t,q)\}^2\right), \tag{4.17}$$

4. *allocating each individual* to one of the components by sampling the index $\mathscr{K}_t$ from a closed form multinomial conditional posterior, with probabilities

$$Pr(\mathscr{K}_t = h|\boldsymbol{V}, \boldsymbol{\Pi}, \boldsymbol{\theta}, \tau, \boldsymbol{t}, \boldsymbol{y}) = \frac{p_h \prod_{q=1}^{Q} \phi\left(y(t,q); \theta_k(q_t), \tau^{-1}\right)}{\sum_{l=1}^{N_+} p_l \prod_{q=1}^{Q} \phi\left(y(t,q); \theta_r(q_t), \tau^{-1}\right)}, \tag{4.18}$$

where $N_+ = \sum_a N_a$, $\phi(z; \mu, \sigma^2)$ is a Gaussian density and $p_k = \Pi_a \Phi(\zeta_{a,ht}) \prod_{r<h}(1 - \Phi(\zeta_{a,rt}))$ for $a, l : I_h(a,l) = 1$ with $a_\tau = b_\tau = 1$.

5. *updating $\Theta_k(\cdot)$,* for $k = 1, \ldots, N_+$, by updating the basis coefficients, $\boldsymbol{\beta}_k$ by sampling from the full conditional posterior distribution

$$(\beta_{kl}|\mathscr{K}_t = k, \mathscr{C}_k, \boldsymbol{V}, \boldsymbol{\Pi}, \boldsymbol{\beta}_{k(-l)}, \tau, \boldsymbol{t}, \boldsymbol{y}) \sim \hat{\nu}_{kl}\delta_0(\cdot) + (1 - \hat{\nu}_{kl})\phi_{A_{kl}}(\cdot; \hat{\beta}_{kl}, \widehat{\sigma}^2_{\beta_{kl}}) \tag{4.19}$$

where A_{kl} denotes the subset of $\mathscr{R}$ such that the resulting function, θ_k, belongs to the feature class $\mathscr{C}_k$, $\phi_A(\cdot, \mu, \sigma^2)$ denotes the $N(\mu, \sigma^2)$ density truncated to the region A,

$$\begin{aligned}
\hat{\nu}_{kl} &= \frac{1}{1 + \frac{(1-\nu_0)}{\nu_0}\frac{\phi_{kl}(0;0,c)}{\phi_{kl}(0;\hat{\beta}_{kl},\widehat{\sigma}^2_{\beta_{kl}})}}, \\
\widehat{\sigma}^2_{\beta_{kl}} &= \left(c^{-1} + \sum_{t:\mathscr{K}_t=k}\sum_{q=1}^{Q}\tau b_l(q_t)^2\right)^{-1}, \\
\hat{\beta}_{kl} &= \widehat{\sigma}^2_{\beta_{kl}}\left(\sum_{t:\mathscr{K}_t=k}\sum_{q=1}^{Q}\tau y^{(kl)}_{tq} b_l(q_t)\right),
\end{aligned} \tag{4.20}$$

where $y(t,q)^{(kl)} = y(t,q) - \sum_{m\neq l}\beta_{km}b_m(q_t)$.

6. *update the probability* of including a basis function by sampling as follows:

$$(\nu_0|-) \sim \text{beta}\left(a_\nu + \sum_k\sum_l 1(\beta_{kl} = 0), b_\nu + \sum_k\sum_l 1(\beta_{kl} \neq 0)\right). \tag{4.21}$$

7. *update* α as in Escobar and West (1995) - Assuming gamma(a_α, b_α) hyperprior for α, with the gamma paremeterized to have mean $a_\alpha b_\alpha$ and variance $a_\alpha b_\alpha^2$ the conditional posterior is

$$\begin{aligned}(\alpha|\eta, k) \sim \pi_\eta \text{gamma}\left(a_\alpha + k, b_\alpha - \log(\eta)\right) \\ + (1 - \pi_\eta)\text{gamma}\left(a_\alpha + k - 1, b_\alpha - \log(\eta)\right),\end{aligned} \tag{4.22}$$

with $\pi_\eta/(1-\pi_\eta) = (a_\alpha + k - 1)/(K\{b_\alpha - \log(\eta)\})$ and $(\eta|\alpha, k) \sim \text{beta}(\alpha + 1, K)$.

4.3 Simulation study

We simulated data for 145 supply curves, $t = 1, \ldots, 145$, mimicking the real data curve's shapes and feature types. There are two feature types; the first one characterize curves in which the gas price has quick to moderate price increase occurring around the quantities 3.576×10^6 GJ and 5.976×10^6 GJ, the second one curves with slow price increase occurring around the quantity 8.376×10^6 GJ, i.e., there exist $a \in \mathbb{R}$ such that $f(\cdot, q) < f(\cdot, q_a) \leq 23$ Euro/GJ for all $q < q_a$. Here $a \in [0, 23]$ and $q_a \in [0, 1.2 \times 10^7 \text{GJ}]$ is the quantity in which $f(\cdot, q) = a$.

This configuration partition the function space Ω in 4 subsets, M=2 and $\sharp\mathscr{A} = 4$, where Ω_1 congregates supply curves in which the price has a quick increase around the quantity 3.576×10^6 GJ, Ω_2 curves with moderate price increase around 5.976×10^6 GJ, Ω_3 curves with fast price increase around 8.376×10^6 GJ and Ω_4 contains no subjects as there is no curves with different behavior regards to the price increase.

We use a smaller set of data in the simulation study, instead 397 as in the real data, to assess the performance of the model in cluster and classify the curves within classes when a smaller set of data is available.

Specifically, we simulate data from

$$y(t, q) = \text{Sigm}(t, q)^{(r_g)} + \epsilon(t, q),$$

with eight functional clusters configured as in Table 4.1, where $y(\cdot)$ is a functional supply curve like response, $r_g = 1, \ldots \sharp\mathscr{A} - 1$ index the subclasses,

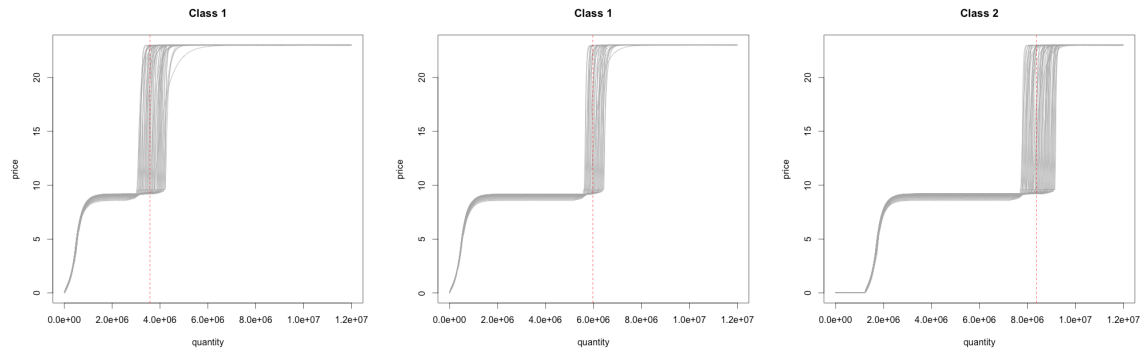
$$\text{Sigm}(t, q) = \frac{2e_0}{1 + e^{\varepsilon(q_0 - q)}} \tag{4.23}$$

a sigmoid function with e_0, q_0 and ε parameters that determines its shape according to the subclass the functions belong too, and $\epsilon(s, t)$ is a temporal error process specific to curve t

and quantity $q = 1, \ldots, 501$ with correlation between curves i and j induced by Cholesky decomposition leading to a functional time series. In Figure 4.2 is shown simulated supply curves within each class and subclass.

Table 4.1 Simulate data: configuration of functional cluster within each class and subclass and its respective probabilities

Cluster	Feature Class	Subclass	Pr
1	1	1	0.103
2	1	1	0.103
3	1	1	0.138
4	1	2	0.117
5	1	2	0.124
6	2	3	0.138
7	2	3	0.138
8	2	3	0.138



(a) Subclass 1 (b) Subclass 2 (c) Subclass 3

Fig. 4.2 Simulated data: Supply like functional data, characteristic curves within each class and subclass. Dashed red line highlight the quantity where the curves have its quick price increase.

The remaining parameters of the model was set in which $c_\tau = 1$ favor a model with smaller set of basis functions, $a_\tau = b_\tau = a_\nu = b_\nu = a_\alpha = b_\alpha = 1$ are non informative priors and $a_\Pi = (0.4, 0.3, 0.3, 0)^T$ informative prior to the proportion of curves falling into each subclass.

In the following we show the results after run the Gibbs sampling algorithm for 26,000 iterations after a burn-in period of 4,000. Gelman-Rubin diagnostics (Gelman and Rubin, 1992) and trace plots of the parameters showed no evidence against convergence. Binder's loss was used to classify the curves within classes, subclasses and groups.

In Figure 4.3 is illustrated representative trajectories for each class for some models. As showed in Table 4.2, the FDP misclassify all functions that belongs to class 2 while the FDDP misclassify all functions of cluster 2 and additionally misclassify some functions of the class 1. The E-FDP appropriately classify all functions as do our approach; though both models classify correctly the functions in the Table 4.3 is showed that our approach more accurately estimate the measurement variance getting close to the truth of 0.01 used to simulate the data. The posterior mean of $\mathbf{\Pi}$, the proportion of functions allocated to each subclass, was $\widehat{\Pi}_1 = 0.219$ (ci: 0.217–0.220), $\widehat{\Pi}_2 = 0.368$ (ci:0.367–0.369), $\widehat{\Pi}_3 = 0.413$ (ci: 0.412–0.414) and $\widehat{\Pi}_4 = 0$. In addition, the estimated mean integrated squared error in the Table 4.4 shows that our approach is closer to the truth compared to the others three models.

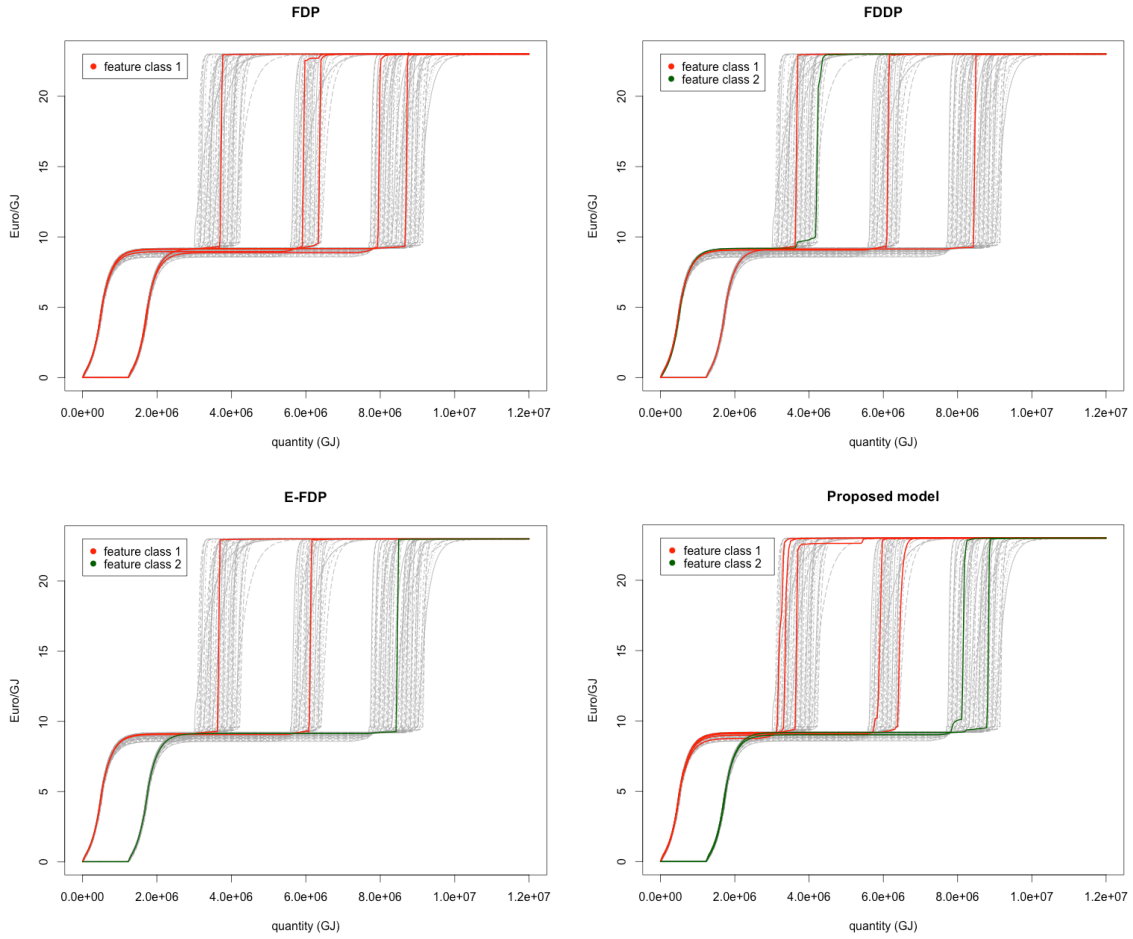
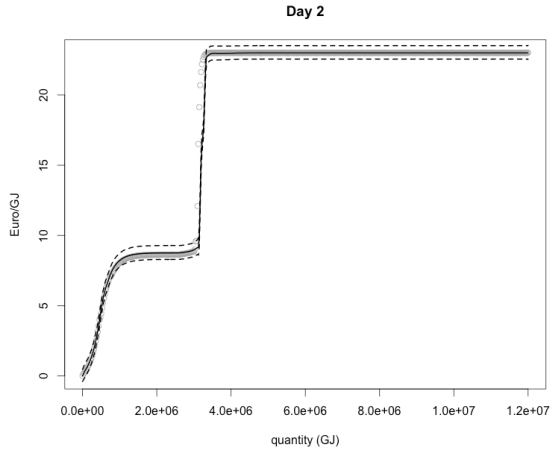


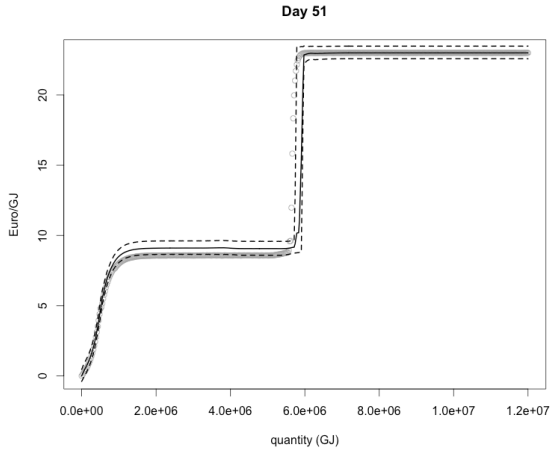
Fig. 4.3 Simulated data: variety of curves represented in each feature class for some models.

Figure 4.4 shows the posterior mean and highest posterior density interval for selected supply curves within each of the classes.

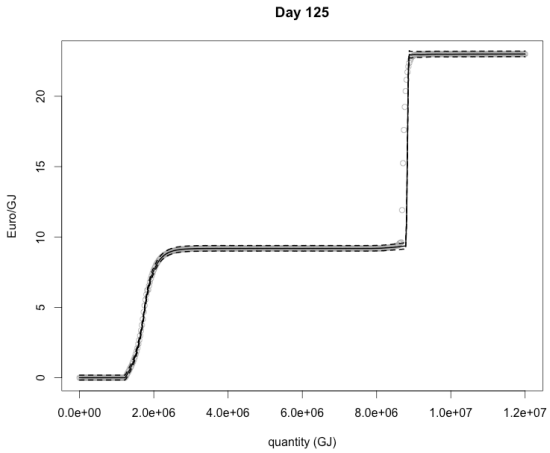
With our approach we were able to correctly group the supply curves within classes, subclasses and groups as in the simulated data. Additionally, the use of feature classes enrich the interpretation of the results as the classification of groups into classes and subclasses allows us to interpret the behavior of the market according to how fast the gas price increases.



(a) Class 1
subclass1



(b) Class 1
subclass 2



(c) Class 2
subclass 3

Fig. 4.4 Simulated data: posterior mean and highest posterior density.

4.4 Application: Italian Natural Gas Balancing Platform

We applied the proposed model to supply and demand curves from the Italian Natural Gas Balancing Platform described in Section 2.2. Some characteristics of the curves present in

Table 4.2 Simulated data: percentages of misclassified units for each model

		Models			
Actual class	Predicted class	FDP	FDDP	E-FDP	Proposed model
1	2	0	6.21%	0	0
2	1	60.00%	60.00%	0	0

Table 4.3 Simulated data: summary of the estimate of the measurement error variance for some models

	Models			
	FDP	FDDP	E-FDP	Proposed model
First quartile	0.001827	0.001725	0.001923	0.0011320
Median	0.001877	0.001774	0.001974	0.0012920
Mean	0.001929	0.001843	0.002058	0.0013070
Third quartile	0.001975	0.001894	0.002063	0.0014820
SD	0.000059	0.000070	0.000093	0.000190

Table 4.4 Simulated data: summary of the mean integrated squared error for some models

	Models			
	FDP	FDDP	E-FDP	Proposed model
First quartile	7.204	6.750	7.744	3.684
Median	7.260	6.801	7.799	3.893
Mean	7.245	6.790	7.786	3.929
Third quartile	7.329	6.878	7.872	4.081
SD	0.141	0.140	0.138	0.257

this data motivated us to develop the model presented in this Chapter. With this analysis we expect to obtain as result group of curves classified by feature classes with their time dependence influencing the cluster allocation. The prediction of exchanged price and quantity might be quite accurate as soon as the past curves are well estimated leading to good forecasted curves.

After set weakly informative priors to the parameters $a_\tau = b_\tau = a_\nu = b_\nu = 1$ and $c = 1$ and informative prior to the proportion of curves falling into each class, $\mathbf{\Pi} = (0.55, 0.35, 0.10, 0)$, i.e., the within-day price grows quickly, moderately or slowly in 55%, 35% and 10% of the days, respectively, we run he Gibbs sampling, algorithm of Section 4.2, for 24,000 iterations after a burn-in period of 6,000. Gelman-Rubin diagnostics and trace plots of the parameters showed no evidence against convergence.

In the Figures 4.5, 4.6 and 4.7 is illustrated the 2 resulting feature classes and its subclasses; the subclass 1 with supply curves with quick price increase, the subclass 2 with moderate increase and the subclass 3 with slow price increase. Note that within each class the dependence structure embedded in our model subdivided each class in distinct subclasses. In Figure 4.8 is shown the variety of feature classes representing the real data. The posterior mean of $\mathbf{\Pi}$, the proportion of supply curves within each class, was $\widehat{\Pi}_1 = 0.609$, $\widehat{\Pi}_2 = 0.270$, $\widehat{\Pi}_3 = 0.121$ and $\widehat{\Pi}_4 = 0$.

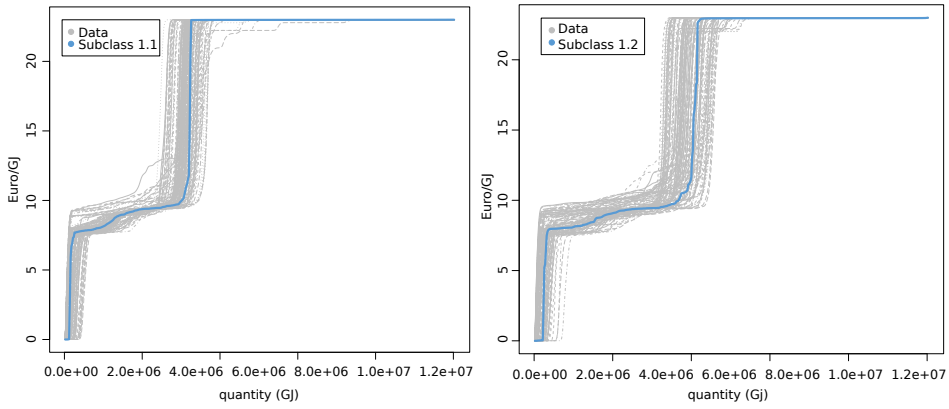


Fig. 4.5 Class 1: quick increase of supply price

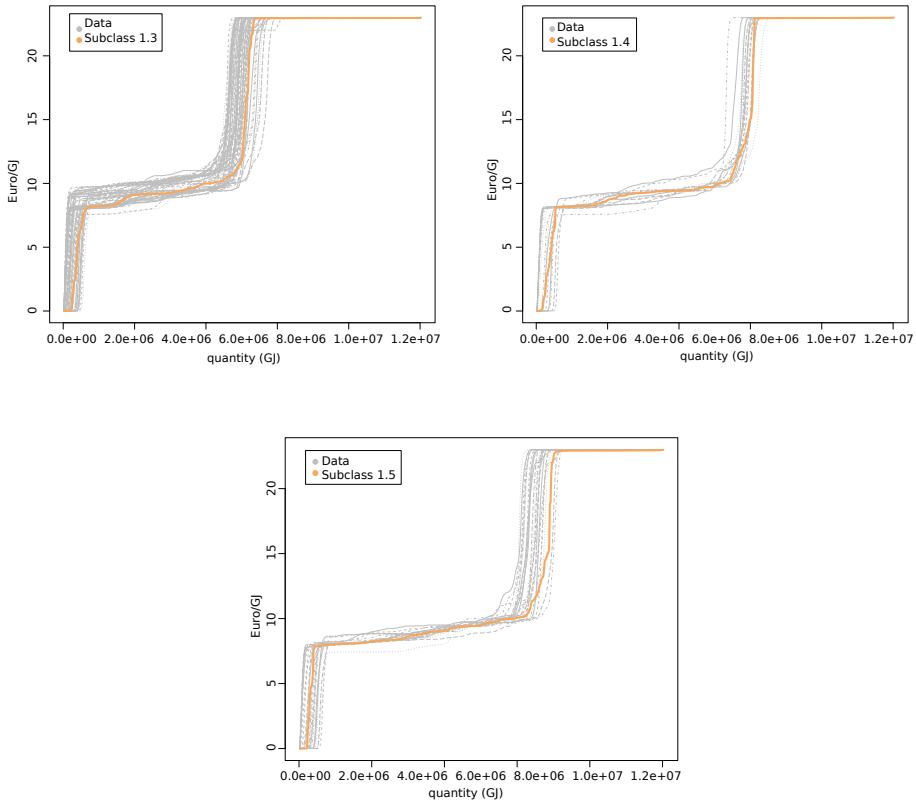


Fig. 4.6 Class 1: moderate increase of supply price

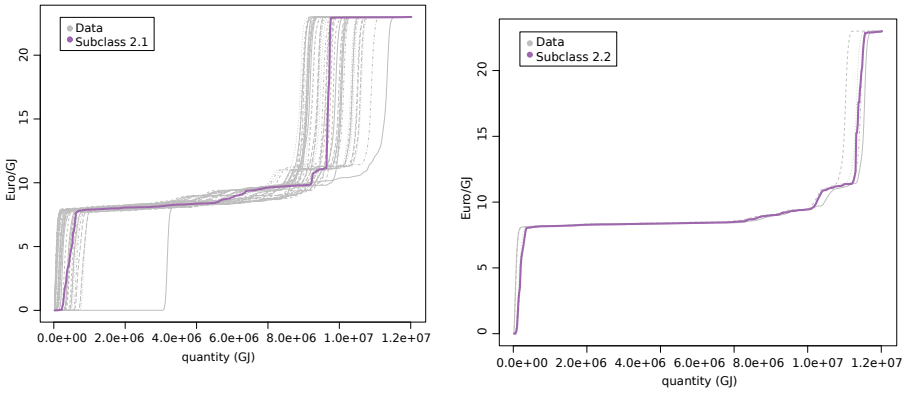


Fig. 4.7 Class 2: slowly increase of supply price

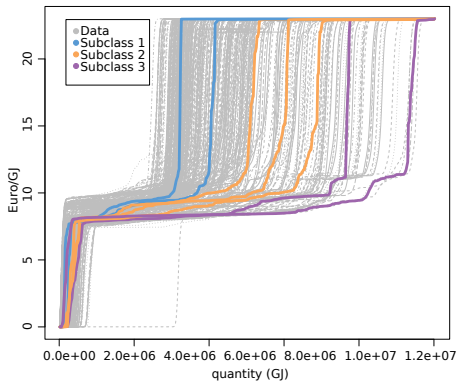


Fig. 4.8 Variety of feature classes representing the real data

Figure 4.9 shows the posterior mean and highest posterior density interval for selected supply curves within each of the classes.

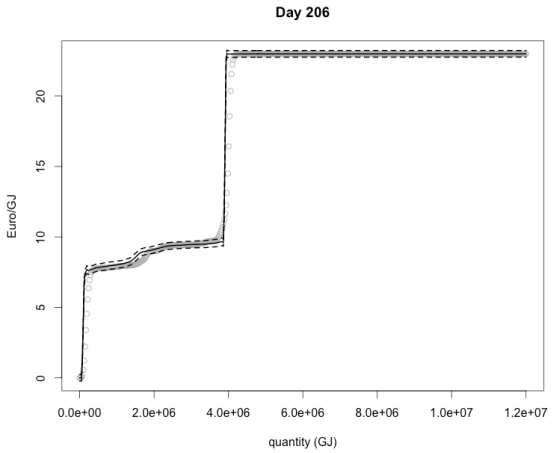
In addition, we compare our results using a simple FDP model where feature classes and time dependence are not accommodated in the model. As you may note in Figure 4.10 the FDP process cluster the subjects mainly by the shape of the curves. The are more clusters cluster compared with our model, though we loose in interpretability, i.e., the cluster behavior does not reflect the behavior of the balancing marketing.

4.5 Discussion

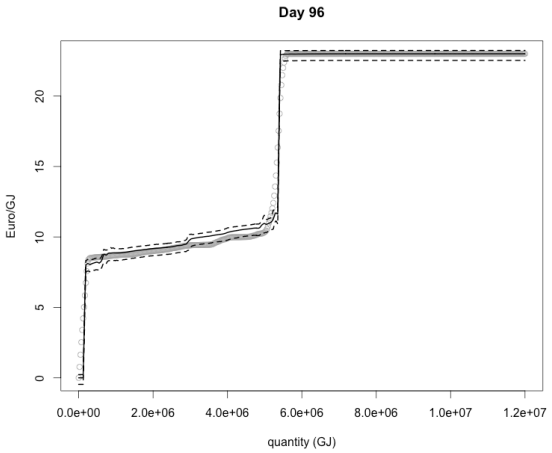
In this Chapter we proposed a Bayesian nonparametric model that allows to group functional time series within feature classes with a probit dynamic model, defined over a partitioned functional sample space, guiding the time dependence.

In our motivating application in balancing marketing feature classes were used to improve in interpretability the results by including as prior the proportion of supply curves with within day similar behavior. The use of the model is not restricted to this kind of application, e.g., the study of basal body temperature curves within menstrual cycles from different women in Scarpa and Dunson (2013) could be complemented with a study of repeated measures for each woman allowing to understand the menstrual cycle for a period of time.

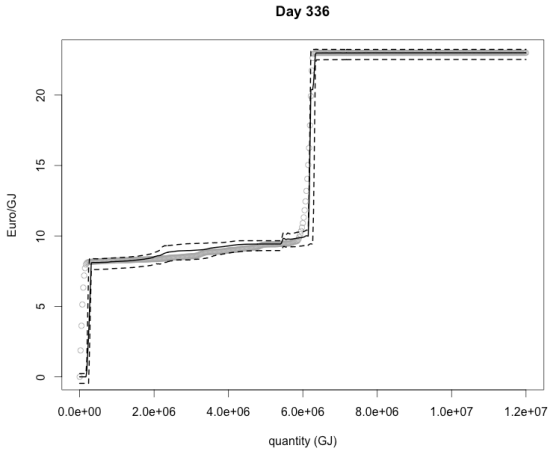
Though the B-splines basis allows the implementation of a number of constraints by restricting the draws of its coefficients to the region of interest we advert that too many restrictions can be computationally cumbersome. An alternative is use different basis functions to favor some characteristics presenting in the data avoiding extra constraints, e.g., I-spline (see Ramsay, 1988) is an alternative to B-splines to model monotonic curves. Extend our



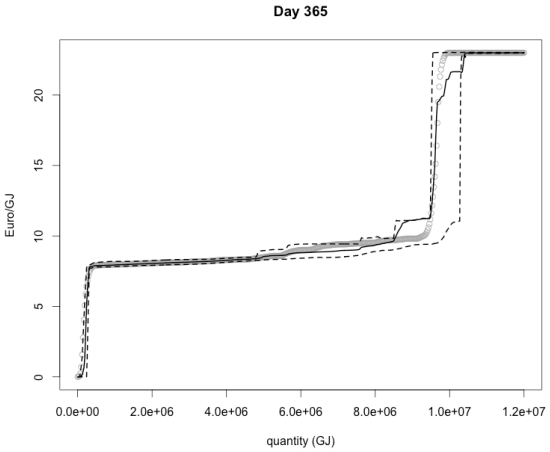
(a) Class1
subclass 1.1



(b) Class 1
subclass 1.2



(c) Calss 1
subclass 1.4



(d) Class 2
subclass 2.1

Fig. 4.9 Real data: posterior mean and highest posterior density.

model to the use of different basis functions is straightforward as it does not rely on the type of basis that is being used.

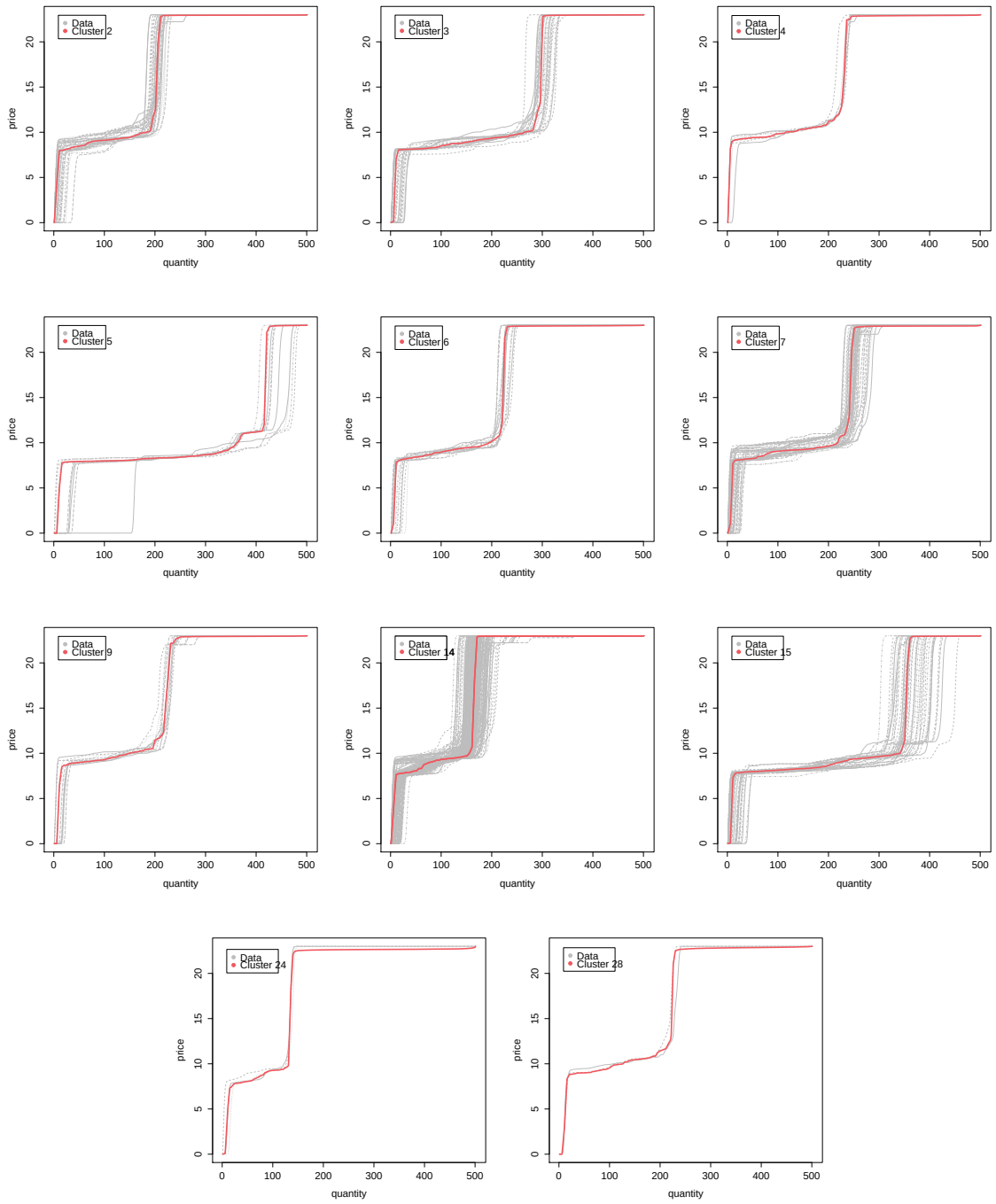


Fig. 4.10 Cluster allocation using the simple FDP; feature classes and time dependence are not embedded in the model.

References

Abraham, C. and Khadraoui, K. (2015). Bayesian regression with b-splines under combinations of shape constraints and smoothness properties. *Statistica Neerlandica*, 69(2):150–170.

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.

Berger, J. O. (1985). Prior information and subjective probability. In *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics, pages 74–117. Springer New York.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society. B*, (36):192–236.

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika*, (65):31–38.

Bloom, L. E., Potts, G. F., Evans, D. E., and Drobes, D. J. (2013). Cue reactivity in smokers: An event-related potential study. *Int J Psychophysicol*, 2(90):258–264.

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In Roussas, G., editor, *Nonparametric Functional Estimation and Related Topics*, volume 335 of *NATO ASI Series*, pages 509–529. Springer Netherlands.

Brumm, H. J. and Mccallum, B. T. (2008). Simultaneous equation econometrics: Some weak-instrument and time-series issues*.

Canale, A. and Vantini, S. (2014). Constrained functional time series: an application to demand and supply curves in the italian natural gas balancing platform. Technical Report 42, MOX, Dipartimento di Matematica "F. Brioschi", Politecnico di Milano.

Congdon, P. (2001). *Bayesian Statistical Modeling*. John Wiley & Sons, Ltd.

de Boor, C. (1978). *A Practical Guide to Splines*. Applied Mathematical Sciences, Springer.

De Boor, C. (2001). *A practical guide to splines*. Applied mathematical sciences. Springer New York.

De Iorio, M., Johnson, W., Müller, P., and Rosner, G. (2009). Bayesian nonparametric nonproportional hazards survival modelling. *Biometrics*, (65(3)):762–771.

De la Cruz-Mesia, R., Quintana, F. A., and Müller, P. (2007). Semiparametric bayesian classification with longitudinal markers. *Applied Statistics*, (56):119–37.

Dickter, C. L. and Kieffaber, P. D. (2013). *EEG Methods for the Psychological Sciences*. SAGE Knowledge.

Duan, J., Guindani, M., and Gelfand, A. E. (2007). Generalized spatial dirichlet process models. *Biometrika*, (94):809–25.

Dunson, D. B. and Park, J. H. (2008). Kernel stick-breaking processes. *Biometrika*, (95):307–323.

Dunson, D. B. and Rodríguez, A. (2011). Nonparametric bayesian models through probit stick-breaking processes. *Bayesian Analysis*, 6:145–178.

Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series. Clarendon Press.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430).

Ferguson, T. S. (1973a). A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(209-230).

Ferguson, T. S. (1973b). Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2(615-629).

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Fritsch, A. and Ickstadt, K. (2009). Improved criteria for clustering based on the posterior similarity matrix. *Bayesian Analysis*, 4(2):367–392.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatialmodeling with dirichlet process mixing. *Journal of the American Statistical Association*, (100):1021–35.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, (7):457–511.

Griffin, J. E. and Steel, M. J. (2006). Order-based dependent dirichlet process. *Journal of the American Statistical Association*, (101):179–94.

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review. *Psychophysiology*, (48):1711–1725.

Hammersley, J. M. and Clifford, P. (1971). Markov fields on finite graphs and lattices (unpublished).

Heckman, J. (1981). Heterogeneity and state dependence. In *Studies in Labor Markets*, pages 91–140. National Bureau of Economic Research, Inc.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer-Verlag New York.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. 96(453).

Ishwaran, H. and Zarepour, M. (2000a). Dirchlet priors sieves in finite normal mixtures. *unpublished manuscript*.

Ishwaran, H. and Zarepour, M. (2000b). Exact and approximate sum-represenations for the dirichlet process. *unpublished manuscript*.

Ishwaran, H. and Zarepour, M. (2000c). Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87(371-390).

Jackson, A. F. and Bolger, D. J. (2014). The neurophysiological bases of eeg and eeg measurement: A review for the rest of us. *Psychophysiology*, 51:1061–1071.

Jansen, B. H. and Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics. Springer-Verlag*.

Junghöfer, M., Peyk, P., Flaisch, T., and Schupp, H. T. (2006). Neuroimaging methods in affective neuroscience: Selected methodological issues. *Progress in Brain Research*, 156:123–143.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

Kenemans, J. L. and Kähkönen, S. (2001). How human electrophysiology informs psychopharmacology: from bottom-up driven processing to top-down control. *Neuropsychopharmacology*, 36(1):26–51.

Lang, P. J. and Bradley, M. M. (2010). Emotion and the motivational brain. *Biological Psychology*, 84(3):437–450.

Lau, J. W. and Green, P. J. (2007). Bayesian model-based clustering procedures. *Journal of Computational and Graphical Statistics*, (16):526–558.

Li, P., Banerjee, S., Hanson, T. A., and McBean, A. M. (2014). Bayesian models for detecting difference boundaries in areal data. *Statistica Sinica: Preprint*.

Liu, J. S. (1996). Nonparametric hierarchical bayes via sequential imputations. *The Annals of Statistics*, 24(911-930).

Luck, S., Mathalon, D. H., O'Donnell, B., Hämäläinen, M., Spencer, K., Javitt, D., and Uhlhaas, P. (2011). A roadmap for the development and validation of event-related potential biomarkers in schizophrenia research. *Biological Psychiatry*, 70(1):28–34.

Luu, P. and Ferree, T. (2005). Determination of the hydrocel geodesic sensor net's average electrode positions and their 10-10 international equivalents. *Technical Note. Electrical Geodesics, Inc.*

MacEachern, S. N. (2000). Dependent dirichlet processes. *Technical Report, Ohio State University, Department of Statistics*.

Maris, E. (2004). Randomization tests for erp topographies and whole spatiotemporal data matrices. *Psychophysiology*, (41):142–151.

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, (49):549–565.

Meyer, M. K., Coull, B. A., Versace, F., Cinciripini, P., and Morris J, S. (2015). Bayesian function-on-function regression for multilevel functional data. *Biometrics*, 3(71):563–74.

Muliere, P. and Secchi, P. (2001). A note on a proper bayesian bootstrap. *Technical Report 18, Universitá degli Studi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi*.

Nunez, P. L. and Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG. 2nd ed.* Oxford University Press; New York.

Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic Linear Models with R*. Use R! Springer New York.

Petrone, S., Guindani, M., and Gelfand, A. E. (2009). Hybrid dirichlet mixture models for functional data. *Journal of the Royal Statistical Society. B*, 4(71):755–782.

Pitman, J. and Yor, M. (1997). The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900.

Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer Series in Statistics, Springer.

Ramsay, J. and Silverman, B. (2006). *Functional Data Analysis*. Springer Series in Statistics, Springer.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statist. Sci.*, 3(4):425–441.

Reich, B. J. and Fuentes, M. (2007). A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, 1(1):249–264.

Scarpa, B. and Dunson, D. B. (2013). Enriched stick breaking processes for functional data. *Journal of the American Statistical Association*.

Versace, F., Cho, Y. L., Engelmann, J. M., Robinson, J. D., Minnix, J. A., Brown, V. L., and Cinciripini, P. M. (2015). Beyond cue reactivity: Blunted brain responses to pleasant stimuli predict long-term smoking abstinence. *Addiction biology*, 6(17):991–1000.

Versace, F., Minnix, J. A., Robinson, J. D., Cho, Y. L., Brown, V. L., and Cinciripini, P. M. (2011). Brain reactivity to emotional, neutral, cigarette-related stimuli in smokers. *Addiction biology*, 2(16):296–307.

Appendix A

Consider the model

$$y(s,t) \sim N\left(\sum_{l=1}^{L} \beta_{kl} b_l(t), \psi_{st}^{-1}\tau_{\epsilon}^{-1}\right)$$

and the priors

$$\begin{aligned}
\beta_{kl} &\sim \nu_0\delta_0(\cdot) + (1-\nu_0)N(0,c) \\
\nu_0 &\sim \beta(a,b) \\
\gamma_{kl} &\sim \text{Bernoulli}(1-\nu_0)
\end{aligned}$$

where the model index $\gamma = (\gamma_{k1}, \ldots, \gamma_{kL}) \in \Gamma$ with $\gamma_{kl} = 1$ denoting that basis function b_l should be included and $\gamma_{kl} = 0$ otherwise.

The full joint posterior distribution is conjugate with the posterior model probabilities available analytically as

$$Pr(\gamma|y) = \frac{\nu_0^{c-p_\gamma}(1-\nu_0)^{p_\gamma} f(y|\gamma)}{\sum_{\gamma^*\in\Gamma} \nu_0^{c-p^{\gamma^*}}(1-\nu_0)^{p^{\gamma^*}} f(y|\gamma^*)}, \text{ for all } \gamma \in \Gamma$$

where $p_\gamma = \sum_k \gamma_k$ the number of basis functions in model γ and $p(y|\gamma)$ is the marginal likelihood of the data under model γ, viz

$$f(y|\gamma) = \int \prod_{s=1}^{S} N(y(s,t)|b_{l,\gamma}(t), \beta_\gamma) N(\beta_\gamma|c) d\beta_\gamma \quad \text{with } b_{l,\gamma}(t) = (b_l(t) : \gamma_{kl} = 1).$$

Due to the possible enormous number (2^k) of different models to sum in the denominator another possibility is to use Gibbs sampling to update γ_k from it s Bernoulli full conditional

posterior distribution given $\gamma_{-k} = (\gamma_l, l \neq k)$, with

$$Pr(\gamma_k = 1|\gamma_{(-k)}, \nu_0) = \left(1 + \frac{\nu_0}{1-\nu_0}\frac{f(y|\gamma_k = 0, \gamma_{(-k)})}{p(y|\gamma_k = 1, \gamma_{(-k)})}\right).$$

To understand (A.1) consider that for γ two cases are possible:

1. $Pr(\gamma_{kl} = 0|y) = \nu_0 \times f(y|\gamma_{kl} = 0, \gamma_{k(-l)})$

2. $Pr(\gamma_{kl} = 1|y) = (1-\nu_0) \times f(y|\gamma_{kl} = 1, \gamma_{k(-l)})$

Since $Pr(\gamma_{kl} = 0|y) + Pr(\gamma_{kl} = 1|y) = 1$ and given $\gamma_{k(-l)} = \gamma_{km}, m \neq l$,

$$\begin{aligned}
Pr(\gamma_{kl} = 1|y) &= 1 - Pr(\gamma_{kl} = 0|y) \\
&= 1 - \nu_0 \times f(y|\gamma_{kl} = 0, \gamma_{k(-l)}) \\
&= 1 - \frac{\nu_0 \times f(y|\gamma_{kl} = 0, \gamma_{k(-l)})}{\nu_0 \times f(y|\gamma_{kl} = 0, \gamma_{k(-l)}) + (1-\nu_0) \times f(y|\gamma_{kl} = 1, \gamma_{k(-l)})} \\
&= \left(1 + \frac{\nu_0 \times f(y|\gamma_{kl} = 0, \gamma_{k(-l)})}{(1-\nu_0) \times f(y|\gamma_{kl} = 1, \gamma_{k(-l)})}\right)^{-1}
\end{aligned}$$

in the calculus above use the hint: $\frac{A+B}{A+B} - \frac{B}{A+B} = \frac{A}{A+B} = (\frac{A+B}{A})^{-1} = (1 + \frac{B}{A})^{-1}$.

Regards to ν_0, we have

$$\begin{aligned}
Pr(\nu_0|\gamma) &\propto \text{Bernoulli}(1-\nu_0) \times \beta(a, b) \\
&= (1-\nu_0)^{\sum_k \gamma_k} \nu_0^{\sum_k (1-\gamma_k)} \times \nu_0^{a-1}(1-\nu_0)^{b-1} \\
&= \nu_0^{(a+\sum_k(1-\gamma_k))-1}(1-\nu_0)^{(b+\sum_k \gamma_k)-1}
\end{aligned}$$

Therefore

$$\nu_0 \sim \text{Beta}\left(a + \sum_k (1-\gamma_k), b + \sum_k \gamma_k\right) \tag{A.1}$$

To calculate the full conditional for the coefficients,

$$
\begin{aligned}
& Pr(\beta_{kl}|\boldsymbol{\beta}_{k(-l)}, y) = \frac{f(y|\beta_{kl}, \boldsymbol{\beta}_{k(-l)})Pr(\beta_{kl})}{f(y(s,t))} \\
\propto\ & f(y|\beta_{kl}, \boldsymbol{\beta}_{k(-l)})Pr(\beta_{kl}) \\
=\ & \exp\left\{-\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t) - \sum_{l=1}^{L}\beta_{kl}b_l(t)\right)^2\right\}\exp\left\{-\frac{1}{2c}\beta_{kl}^2\right\} \\
=\ & \exp\left\{-\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t) - \left[\sum_{m\neq l}\beta_{km}b_m(t) + \beta_{kl}b_l(t)\right]\right)^2\right\}\exp\left\{-\frac{1}{2c}\beta_{kl}^2\right\} \\
=\ & \exp\left\{-\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t)^2 - 2y(s,t)\left[\sum_{m\neq l}\beta_{km}b_m(t) + \beta_{kl}b_l(t)\right] + \left[\sum_{m\neq l}\beta_{km}b_m(t) + \beta_{kl}b_l(t)\right]^2\right)\right. \\
-\ & \left.\frac{1}{2c}\beta_{kl}^2\right\} \\
=\ & \exp\left\{-\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon y(s,t)^2 + \sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon y(s,t)\sum_{m\neq l}\beta_{km}b_m(t) + \sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon y(s,t)\beta_{kl}b_l(t)\right. \\
-\ & \frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left[\sum_{m\neq l}\beta_{km}b_m(t)\right]^2 - \sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\sum_{m\neq l}\beta_{km}b_m(t)\beta_{kl}b_l(t) - \frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left[\beta_{kl}b_l(t)\right]^2 \\
-\ & \left.\frac{1}{2c}\beta_{kl}^2\right\} \\
=\ & \exp\left\{\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon y(s,t)\beta_{kl}b_l(t) - \sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\sum_{m\neq l}\beta_{km}b_m(t)\beta_{kl}b_l(t) - \frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left[\beta_{kl}b_l(t)\right]^2\right. \\
-\ & \left.\frac{1}{2c}\beta_{kl}^2\right\} \\
=\ & \exp\left\{\beta_{kl}^2\left(-\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon b_l(t)^2 + c^{-1}\right) + \beta_{kl}\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon y(s,t)b_l(t) - \sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\sum_{m\neq l}\beta_{km}b_m(t)\right)\right\} \\
=\ & \exp\left\{\beta_{kl}^2\left(-\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon b_l(t)^2 + c^{-1}\right) + \beta_{kl}\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t) - \sum_{m\neq l}\beta_{km}b_m(t)\right)b_l(t)\right)\right\} \\
=\ & \exp\left\{\beta_{kl}^2 + \beta_{kl}\frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t) - \sum_{m\neq l}\beta_{km}b_m(t)\right)b_l(t)}{-\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon b_l(t)^2 + c^{-1}}\right\} \\
=\ & \exp\left\{\beta_{kl}^2 - 2\beta_{kl}\frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t) - \sum_{m\neq l}\beta_{km}b_m(t)\right)b_l(t)}{\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon b_l(t)^2 + c^{-1}}\right. \\
+\ & \left.\left(\frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t) - \sum_{m\neq l}\beta_{km}b_m(t)\right)b_l(t)}{\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon b_l(t)^2 + c^{-1}}\right)^2\right\} \\
=\ & \exp\left\{\frac{1}{\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon b_l(t)^2 + c^{-1}}\left[\beta_{kl}^2\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon b_l(t)^2 + c^{-1}\right)\right.\right. \\
-\ & 2\beta_{kl}\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t) - \sum_{m\neq l}\beta_{km}b_m(t)\right)b_l(t)\right) + \frac{1}{\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon b_l(t)^2 + c^{-1}} \\
& \left.\left.\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_\epsilon\left(y(s,t) - \sum_{m\neq l}\beta_{km}b_m(t)\right)b_l(t)\right)\right]\right\}
\end{aligned}
$$

Therefore

$$(\beta_{kl}|-) \sim N\left(\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_{\epsilon}b_l(t)^2 + c^{-1}\right)^{-1}\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_{\epsilon}\left(y(s,t) - \sum_{m\neq l}\beta_{km}b_m(t)\right)b_l(t),\ \left(\sum_{s=1}^{S}\sum_{t=1}^{T}\psi_{st}\tau_{\epsilon}b_l(t)^2 + c^{-1}\right)^{-1}\right)$$

updating the $\tau_\eta|\Theta, \boldsymbol{V}, \boldsymbol{y}, \boldsymbol{\eta}$ the spatial dispersion parameter

Considering

$$\begin{aligned}
Pr(\tau_\eta|\Theta, \boldsymbol{V}, \boldsymbol{y}, \boldsymbol{\eta}) &= Pr(\boldsymbol{\eta}|\tau_\eta, \Theta, \boldsymbol{V}, \boldsymbol{y}) \times P(\tau_\eta) \\
&= N_n(\mathbf{0}, \boldsymbol{\Sigma}) \times \text{gamma}(c, d) \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}\right\} \times \tau_\eta^{c-1}\exp(-d\tau_\eta) \\
&= |\tau_\eta^{-1}(D-\rho W)^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\eta}'[\tau_\eta^{-1}(D-\rho W)^{-1}]^{-1}\boldsymbol{\eta}\right\} \times \tau_\eta^{c-1}\exp(-d\tau_\eta) \\
&= (\tau_\eta^{-n}|(D-\rho W)^{-1}|)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\eta}'\tau_\eta(D-\rho W)\boldsymbol{\eta}\right\} \times \tau_\eta^{c-1}\exp(-d\tau_\eta) \\
&\propto \tau_\eta^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\eta}'\tau_\eta(D-\rho W)\boldsymbol{\eta} - d\tau_\eta\right\} \times \tau_\eta^{c-1} \\
&= \tau_\eta^{\frac{n}{2}+c-1} \exp\left\{-\left(\frac{1}{2}\boldsymbol{\eta}'(D-\rho W)\boldsymbol{\eta} + d\right)\tau_\eta\right\}
\end{aligned}$$

sampling from gamma $\left(\frac{n}{2}+c, \frac{1}{2}\boldsymbol{\eta}'(D-\rho W)\boldsymbol{\eta} + d\right)$.

Ronaldo Rouvher Guedes Silva

Affiliation
Department of Statistical Science
University of Padua
Via Cesare Battisti, 241 - 35121 Padua, Italy

Research Interests

Theory and Methods: Bayesian nonparametric statistics; Functional data; Time series; Dynamic models; Spatial statistics; Mixed models.

Applications: Neuroimaging, Economy; Energy balancing market; Pharmacokinetics.

Education

University of Padua, Italy — Ph.D. — Statistical Sciences — 2013 - 2015
Advisor: Bruno Scarpa & Michele Guindani
Thesis: Nonparametric models for dependent functional data
Expected thesis submission on January 2016

São Paulo State University (Unesp), Brazil — M.S. — Biometry — 2012
Advisor: Luzia Trinca
Thesis: Optimum designs for nonlinear models: fixed and random effects

São Paulo State University (Unesp), Brazil — B.S. — Mathematics — 2010

Visiting Positions

Rice University, USA — 10/2014 - 11/2015
MD Anderson Cancer Center, The University of Texas , USA
Collaborative work with Michele Guindani, Marina Vannucci & Francesco Versace
Project: Nonparametric Bayesian models for detecting boundaries in event-related potentials data

University of São Paulo (USP), Brazil — 08-09/2013
Project: Multivariate Data Analysis

Manuscripts in Preparation

1. Silva, R. R.G., Guindani, M., Vannucci, M., Versace, F. A Bayesian Nonparametric approach for the analysis of functional ERP data in neuroimaging.

2. Silva, R. R.G., Canale, A. & Scarpa, B. Enriched functional probit stick-breaking process.

Selected Presentations

Silva, R. R.G., Scarpa, B. & Canale, A. On the way to Bayesian nonparametric approach for dependent functional data. University of Padua, Italy. 2014

Silva, R. R.G., Canale, A. & Scarpa, B. Enriched functional probit stick-breaking process, Rice University, USA.

Silva, R. R.G., Guindani, Vannucci, M, Versace, F. Nonparametric Bayesian models for detecting boundaries in event-related potentials data, Rice University, USA.

Silva, R. R. G. and Trinca, L. A. Criterion value approximations in the construction of pseudo-Bayesian optimal designs. Brazilian Region of International Biometric Society. 2012

Ferreira, I. E. P., Silva, R. R. G., Ferreira, C. P. and Trinca, L. A. A hybrid algorithm genetic-exchange for searching exact optimal designs. Brazilian Region of International Biometric Society, Maringá State University.

Silva, R. R. G. and Trinca, L. A. Pseudo-Bayesian optimum designs for nonlinear models using standard functions in R. National Symposium on Probability and Statistics, Joao Pessoa, Brazil.

Academic Awards

National Council for Scientific and Technological Development scholarship, Brazil. 2013 - 2015

Coordination for the Improvement of Higher Education Personnel scholarship, Brazil. *2011-2012*

Scholarship for Disadvantaged Students. São Paulo State University (Unesp), Brazil. 2005-2010

Teaching Experience

Forward filtering backward sampling algorithm (2 hours). Rice University, USA & University of Padua, Italy. 2014

Short-course on Design of experiments (5 hours). XI Workshop of postgraduate at São Paulo State University (Unesp), Botucatu, Brazil. 2012

Biostatistics (36 hours). Discipline of graduate course in Medical Physics. Teaching training at São Paulo State University (Unesp), Botucatu, Brazil. 2011

Languages

English:	Fluent
Italian:	Fluent
Portuguese:	Mothertongue
Spanish:	Fluence in reading and listening with basic writing and speaking competence.
German:	Basic competence

Membership

International Society for Bayesian Analysis (ISBA)

Referees

Bruno Scarpa

Assistant Professor
Departament of Statistical Science. University of Padova. Via Cesare Battisti, 241. 35121 Padova, Italy.

Michele Guindani

Assistant Professor
Departament of Biostatistics. MD Anderson Cancer Center. Pickens Academic Tower - 1400. Pressler Dr. Houston TX 77030, USA.

Marina Vannucci

Professor of Statistics
Departament of Statistics MS 138. Ricer University. 6100 Mains Street. Houston TX 77005, USA.