Gaussian graphical modeling for spectrometric data analysis

Laura Codazzi[a], Alessandro Colombi[a], Matteo Gianella[a], Raffaele Argiento[b], Lucia Paci[b], Alessia Pini[b]

[a]*Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, 20133, ITALY*
[b]*Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Largo Agostino Gemelli 1, Milan, 20123, ITALY*

Abstract

Motivated by the analysis of spectrometric data, we introduce a Gaussian graphical model for learning the dependence structure among frequency bands of the infrared absorbance spectrum. The spectra are modeled as continuous functional data through a B-spline basis expansion and a Gaussian graphical model is assumed as a prior specification for the smoothing coefficients to induce sparsity in their precision matrix. Bayesian inference is carried out to simultaneously smooth the curves and to estimate the conditional independence structure between portions of the functional domain. The proposed model is applied to the analysis of infrared absorbance spectra of strawberry purees.

Keywords: Bayesian Inference, Birth-death process, Functional Data Analysis, Model Selection, Spectrum Analysis

1. Introduction

The analysis of the interaction of infrared radiation with matter by absorption, emission, or reflection is accomplished by *infrared spectroscopy.* This technique is used to study and identify chemical substances in solid, liquid, or gaseous forms. For instance, mid-infrared spectroscopy coupled with chemometrics or statistical techniques have been used in the literature to study the substance composition and detect the presence of adulterants in food (Kemsley et al., 1996; Downey et al., 1997; Holland et al., 1998; Meza-Márquez et al., 2010).

From a mathematical point of view, a spectrum is a continuous function of the wavelength. The dependence structure between the signal at different wavelength bands is important to understand which bands are related to the different components: if two different bands of the spectrum are associated, we can conclude that they refer to the same components, or to closely related components. Our goal in this paper is then to investigate the structure of conditional dependence among different portions of an absorbance spectrum.

Since infrared spectra are continuous functions of the wavelength, we embed this problem in the framework of functional data analysis (see e.g., Lee et al., 2002; Ramsay and Silverman, 2005; Xiao et al., 2016). In the Bayesian setting, Yang et al. (2016) proposed a hierarchical model with Gaussian-Wishart processes for simultaneously smoothing multiple functional observations and estimating mean-covariance functions. However, like any Gaussian process based approach, their model suffers serious computational burden when functional data are observed on high-dimensional grids. To address this computational issue, Yang et al. (2017) proposed to approximate the underlying true functional data with basis functions, and derive the induced Bayesian hierarchical model for the associated smoothing coefficients based on a Gaussian-Wishart prior.

Following the latter approach, we interpret the analysis of spectrometric data as a smoothing problem of functional data, followed by inference on the smoothing coefficients. Specifically, we use a B-spline basis expansion to represent the functional data: indeed, a generic B-spline basis function has a compact support, meaning that it is different from zero only in a portion of the domain, i.e., on a defined portion of the spectrum. If the number of basis functions is large, this portion is very small compared to the whole domain, and identifies a very precise band of the spectrum. Through the basis expansion, a coefficient is associated to every basis function, and the whole function can be represented through the vector of associated coefficients $\boldsymbol{\beta}$. As a consequence, we can detect a hidden association between different bands of the spectrum through an association between the corresponding smoothing coefficients.

Inference on B-spline basis coefficients has been used in the functional data analysis literature for assessing the mean structure of functional data along the domain (see e.g., Pini and Vantini, 2016). Here, instead, we focus on the dependence structure of the smoothing coefficients. With a Bayesian perspective, this concerns the specification of a prior distribution for the coefficients that encodes a dependence structure. Usually, a random walk

prior is assumed to induce smoothness on the $\boldsymbol{\beta}$'s (Lang and Brezger, 2004; Crainiceanu and Goldsmith, 2010; Telesca and Inoue, 2008). First and second order random priors are typically used, that translate into a banded precision penalization matrix that is fixed and depends only on a smoothing parameter. However, in our application, assuming that the dependence structure is known and restricted to the structure induced by a random walk prior appears inadequate. Indeed, coefficients associated with closer bands of the spectrum are not necessarily expected to be similar since they may relate to different components. Rather, long range interactions may also exist between far apart portions of the spectrum. Hence, a more flexible modeling framework is advocated.

In this paper, we move a step forward and develop a method to learn the dependence structure among portions of the absorbance spectrum based on graphical modeling. Graphical models describe a mapping between a graph and a family of multivariate probability models (Lauritzen, 1996). Namely, they embody conditional independence relationships among a set of random variables which can be read off from the graph. Usually, the structure of the underlying graph is unknown and needs to be estimated on the basis of the available data: this is referred to as graph structural learning.

Graphical models have been widely applied to infer several types of networks, particularly under the Bayesian framework, see among others Dobra et al. (2004); Peterson et al. (2015); Tan et al. (2017); Ni et al. (2017); Cremaschi et al. (2019); Paci and Consonni (2020). All the cited works assume a graphical model as a model for the data, i.e., the graph represents the conditional independence relationships among the observed variables. Here, instead, we introduce a graphical model as a prior specification for the smoothing coefficients, such that the graph interprets the dependence structure of the portions of the absorbance spectrum. Graph-based priors for regression coefficients have been studied in variable selection setting by, among others, Liu et al. (2014), Chakraborty and Lozano (2019) and Cai et al. (2020), who proposed Bayesian regularization priors to favor sparsity and clustering based on the graph Laplacian. However, these approaches are not suitable to deal with functional data where curve-specific coefficients are needed to simultaneously smooth the curves.

In addition, recent works focused on graphical modeling for multivariate functional data analysis (Zhu et al., 2016; Li and Solea, 2018; Qiao et al., 2019, 2020). These approaches, known as functional graphical models, aim at depicting the conditional dependence structure among multiple random func-

3

tions observed over a set of individuals, i.e., they assume a graphical model where the vertices of the underlying graph are the multiple functions and the edges are estimated from the conditional covariance function extended to the functional domain. Rather, in our setting, we work with univariate functional data modeled through a B-spline basis expansion and we assume a graphical model where the nodes of the graph are the smoothing coefficients and the edges reflect the conditional dependence structure among the portions of the functional domain associated to such coefficients.

Summarizing, our contribution is to introduce a Bayesian hierarchical model for simultaneously smoothing functional data and learning the independence structure along the domain of the curves. Absorbance spectra are modeled as continuous functional data through a cubic B-spline basis expansion such that the dependence between two bands of the spectrum is reflected in the relationship among the corresponding smoothing coefficients. A Gaussian graphical model is then assumed as a prior for basis expansion coefficients to enable structural learning of frequency bands of the spectrum. To the best of our knowledge, this is the first work where a Gaussian graphical model is assumed as a prior model for the smoothing coefficients in the analysis of functional data. On the computational side, we design an efficient sampling strategy to approximate the joint posterior distribution of the graph and model parameters. We illustrate our method on simulated datasets as well as on a real data set, studying the infrared absorbance spectra of strawberry purees.

The remainder of this paper is organized as follows. Section 2 introduces the Bayesian graphical model for functional data analysis, discusses the smoothing procedure as well as the covariance selection model. In Section 3 we describe the sampling strategy implemented to sample from the joint posterior distribution of all model parameters, including the graph, and discuss how to summarize the output. Section 4 presents a simulation study, while Section 5 illustrates the results of a real data analysis carried out to study the infrared absorbance spectra of strawberry purees. Finally we conclude with a brief discussion in Section 6.

2. Bayesian graphical model for smoothing functional data

2.1. Smoothing procedure

As we have anticipated in the Introduction, we reinterpret our task as a traditional smoothing procedure of functional data in a Bayesian frame-

work. Let n be the number of curves, r the number of grid points, i.e., the wavelengths at which absorbance is measured, and p be the number of basis functions chosen for the smoothing. Let $y_i(s)$ be the absorbance at wavelength $s \in [l, u]$ for unit i, $i = 1, \ldots, n$, where $[l, u] \subset \mathbb{R}$ is the common domain of all curves. The whole curve $y_i(s)$ is then the absorbance spectrum of unit i. According to the usual smoothing technique, the model we assume for the i-th spectrum is:

$$y_i(s) = \sum_{j=1}^{p} \beta_{ij} \varphi_j(s) + \varepsilon_i(s),$$

where $\varphi_1, \ldots, \varphi_p$ are suitable cubic B-spline basis functions, $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ip})^\top$ is the spline coefficients' vector specific to i-th curve and $\varepsilon_i(s) \overset{\text{iid}}{\sim} N(0, \tau_\varepsilon^2)$. We introduce the cubic B-spline design matrix $\boldsymbol{\Phi} \in \mathbb{R}^{r \times p}$ as:

$$\boldsymbol{\Phi} = \begin{bmatrix} \varphi_1(s_1) & \varphi_2(s_1) & \cdots & \varphi_p(s_1) \\ \varphi_1(s_2) & \varphi_2(s_2) & \cdots & \varphi_p(s_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(s_r) & \varphi_2(s_r) & \cdots & \varphi_p(s_r) \end{bmatrix}. \tag{1}$$

So, the element lj-th of $\boldsymbol{\Phi}$ is j-th basis function $\varphi_j(\cdot)$ evaluated at the l-th grid point s_l. Note that, the smoothing model can be also used in the case where curves are measured at different wavelengths, by modifying $\boldsymbol{\Phi}$ accordingly.

The right panel of Figure 1 shows an example of a 10-dimensional cubic B-spline basis for the curve in the left panel. Note that the support of each basis function is a small portion of the domain, which decreases its size when the number of basis functions increases. For instance, in Figure 1, the support of each basis function overlaps with only four other function' supports. To facilitate the interpretation of the results of our model, we will consider each B-spline basis coefficient as representative of the central part of each B-spline basis function only.

Let $\boldsymbol{Y}_i = (y_i(s_1), \ldots, y_i(s_r))^\top$ be the absorbance spectrum at all observed wavelengths of curve i. The smoothing model assumes the observation to follow a conditional normal sampling distribution, that is

$$\boldsymbol{Y}_i \mid \boldsymbol{\beta}_i, \ \tau_\varepsilon^2 \overset{\text{ind}}{\sim} N_r \left(\boldsymbol{\Phi}\boldsymbol{\beta}_i, \ \tau_\varepsilon^2 \boldsymbol{I}_r \right), \tag{2}$$

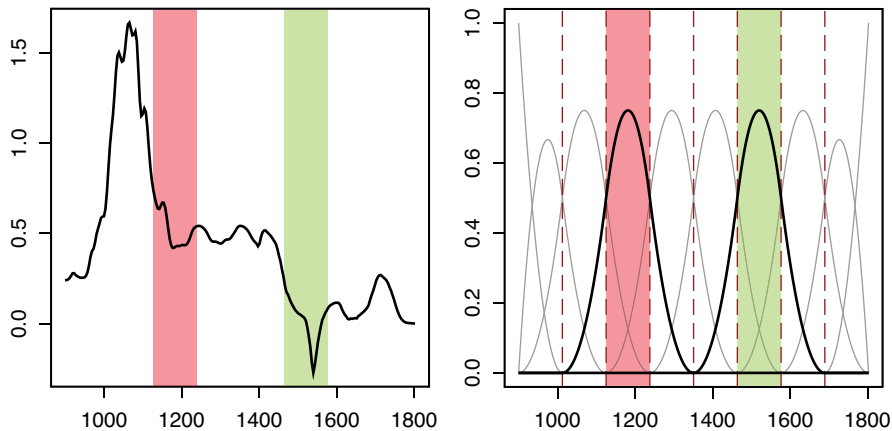where τ_ϵ^2 is the error variance and $\boldsymbol{I}_r$ is the $r \times r$ identity matrix.

Figure 1: Example of a curve (left panel) and the corresponding 10-dimensional cubic B-spline basis (right panel).

In this work, we rely on a hierarchical Bayesian approach to smooth all functional observations simultaneously, enabling the borrowing of strength across all curves. To accomplish that, we place a prior distribution for the $\boldsymbol{\beta}_i$ coefficients given by

$$\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n \mid \boldsymbol{\mu}, \boldsymbol{\Omega} \overset{\text{iid}}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}), \tag{3}$$

where $\boldsymbol{\mu}$ is the prior mean vector and $\boldsymbol{\Omega}$ is the precision matrix. Yang et al. (2017) derived the conjugate prior distribution for $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ from the assumption that the underlying true functional data come from a Gaussian process, and thus resulting in a Normal-Wishart distribution that depends on the basis functions. In this work, instead, we frame the prior model for the $\boldsymbol{\beta}$'s into the Gaussian graphical modeling setting. Indeed, given the peculiar form of each B-spline function, the dependence between two bands of the spectrum is reflected in the relationship among the corresponding smoothing regression parameters. Hence, our goal boils down to the study of the conditional dependence structure of the $\boldsymbol{\beta}$'s encoded in the precision matrix $\boldsymbol{\Omega}$.

2.2. Graph structural learning

Let $G = (V, E)$ be an undirected graph defined by the node set $V = \{1, \ldots, p\}$ representing the coefficients and by the undirected edge set $E \subset V \times V$. We assume the precision matrix $\boldsymbol{\Omega}$ to be Markov with respect to G. Namely, dropping the subscript i for simplicity in the exposition, coefficients

β_j and β_k are conditionally independent given all the remaining variables, $\boldsymbol{\beta}_{\setminus\{j,k\}}$, whenever $\{j,k\} \notin E$, if and only if the corresponding entry in matrix $\boldsymbol{\Omega}$ is zero, i.e., $\beta_j \perp \beta_k \mid \boldsymbol{\beta}_{\setminus\{j,k\}} \Leftrightarrow \boldsymbol{\Omega}_{jk} = 0$; see Figure 2 for an example. The graph G summarizes the conditional independence structure of the co-
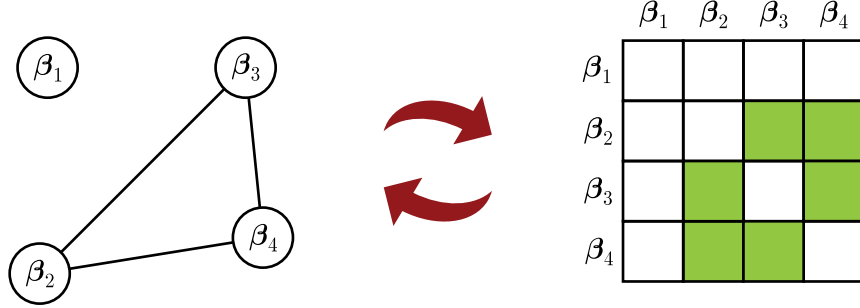

Figure 2: Example of a Gaussian graphical model with four nodes. On the left, the graph with four nodes and, on the right, the corresponding precision matrix where green cells represent nonzero entries.

efficients and represents the main goal of our inference. Indeed, the graph is unknown and must be estimated. Note that this is different from estimating the conditional network among the predictors of a regression model as in Peterson et al. (2016). Unlike measurement-error models in which predictors and response are both measured with error, spline basis expansion are fixed and known. Hence, we infer a sparse network on the associated smoothing coefficients, which provides information on which basis have similar slope.

Given $\mathbb{P}_G$ the space of $p \times p$ positive definite matrices that are Markov with respect to graph G, we assume the precision matrix $\boldsymbol{\Omega} \in \mathbb{P}_G$ to be distributed, a priori, as a G-Wishart$(d, \boldsymbol{D})$ (Roverato, 2002; Letac and Massam, 2007) with density

$$P\left(\boldsymbol{\Omega} \mid G, d, \boldsymbol{D}\right) = I_G\left(d, \boldsymbol{D}\right)^{-1} |\boldsymbol{\Omega}|^{(d-2)/2} \exp\left\{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Omega}\boldsymbol{D}\right)\right\}, \qquad (4)$$

where $d > 2$ is the shape parameter, $\boldsymbol{D}$ is the $p \times p$ positive definite inverse scale matrix, $\text{tr}(\cdot)$ denotes the trace operator and $I_G(d, \boldsymbol{D})$ is the normalizing constant,

$$I_G(d, \boldsymbol{D}) = \int_{\mathbb{P}_G} |\boldsymbol{\Omega}|^{(d-2)/2} \exp\left\{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Omega}\boldsymbol{D}\right)\right\} d\boldsymbol{\Omega}. \qquad (5)$$

7

The main challenge associated with density in (4) arises from the normalizing constant in equation (5); although an analytic form does exist (Uhler et al., 2018), the expression is mathematically complex and very difficult to compute in practice. The sampling strategy described in Section 3 is then designed to avoid the calculation of such intractable normalizing constant.

The last ingredient to fully specify the model is the prior distribution on the graph G. Several alternatives have been proposed in the literature, see among others Dobra et al. (2004); Jones et al. (2005); Scott and Carvalho (2008); Paci and Consonni (2020). In this work, we consider a discrete uniform distribution over the space of all possible graphs $\mathcal{G}$, that is

$$\pi(G) = \frac{1}{|\mathcal{G}|}, \text{ for each } G \in \mathcal{G}. \tag{6}$$

Alternatively, it is possible to induce a prior on the graph by assigning to each possible edge $f \in E$ an independent Bernoulli prior with parameter $\theta_f \in (0, 1)$, that yields

$$\pi(G) \ \propto \ \prod_{f \in E} \theta_f \prod_{f \in \bar{E}} (1 - \theta_f), \tag{7}$$

where $\bar{E}$ is the complement of E. In a general setting, the Bernoulli parameters θ_f can be different from edge to edge. Conversely, when $\theta_f = \theta$, $\forall f \in E$, we get $\pi(G) \propto \theta^{|E|}(1 - \theta)^{\binom{p}{2} - |E|}$, where $|E|$ is the cardinality of E. Notice that, when $\theta_f = 0.5$, $\forall f \in E$, prior (7) collapses to the uniform prior in equation (6).

For the remaining parameters, we choose independent semi-conjugate vague priors. Summing up, our Bayesian hierarchical model is defined as follows:

$$
\begin{aligned}
\boldsymbol{Y}_i \mid \boldsymbol{\beta}_i, \ \tau_\varepsilon^2 &\overset{\text{ind}}{\sim} \mathrm{N}_r \left(\boldsymbol{\Phi}\boldsymbol{\beta}_i, \tau_\varepsilon^2 \mathbf{I}_r \right) \\
\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n \mid \boldsymbol{\mu}, \ \boldsymbol{\Omega} &\overset{\text{iid}}{\sim} \mathrm{N}_p \left(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1} \right) \\
\boldsymbol{\Omega} \mid G &\sim \text{G-Wishart} \left(d, \boldsymbol{D} \right) \\
G &\sim \pi \left(G \right) \\
\boldsymbol{\mu} &\sim \mathrm{N}_p \left(\boldsymbol{0}, \sigma_\mu^2 \mathbf{I}_p \right) \\
\tau_\varepsilon^2 &\sim \mathrm{IG} \left(a, b \right),
\end{aligned}
\tag{8}
$$

where $\mathrm{IG}(a, b)$ denotes the Inverse Gamma distribution with shape parameter a and rate parameter b. The Bayesian model in equation (8) enables joint

modeling of all the curves with borrowing strength of information across functional data as well as to fully account for uncertainty over the graph.

3. Sampling strategy

To deliver Bayesian posterior inference we need to approximate the joint posterior distribution of the $\boldsymbol{\beta}$'s, τ_ε^2, $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$ and G. Here, one of the main challenge arises from the graph G that belongs to a huge discrete space endowed with a complex topology. Indeed, there are $2^{p(p-1)/2}$ possible graphs representing the conditional independence structure of a p-dimensional vector of coefficients. For instance, with $p = 7$ B-spline basis functions only, we end up with more than 2 million graphs. As consequence, any simulation algorithm struggles to explore the space of all possible graphs.

A customary approach to explore such space is by means of a Monte Carlo Markov chain (MCMC) algorithm. In this framework, a reversible Markov chain is built such that its limiting distribution is the joint posterior of the graph and the precision matrix. Usually, the algorithm proceeds by proposing at each step a new graph which differs from the current one for only one edge, i.e., local moves. This procedure is called add-delete Metropolis Hastings, see among others Giudici and Castelo (2003); Bhadra and Mallick (2013). Then, the Markov chain moves over the graph space by comparing different graphs, in order to find the one having the highest marginal posterior. However, this procedure has some drawbacks. For instance, Jones et al. (2005) showed that the add-delete Metropolis Hastings methods suffer from lack of convergence in high-dimensional settings.

As an alternative, Mohammadi and Wit (2015) proposed a birth and death MCMC algorithm (BDMCMC) based on a continuous time Markov process where jumps between birth and death events are taken to be random variables with specific rates. The authors showed that the BDMCMC outperforms alternative Bayesian approaches in terms of convergence and mixing in the graph space as well as computing time.

We build a sampling strategy that blends Gibbs-sampler steps for updating the smoothing components of the Bayesian model with a BDMCMC step for updating the precision matrix $\boldsymbol{\Omega}$ and the graph G, as describe in Algorithm 1. As concerns the Gibbs-sampler steps, we sample the $\boldsymbol{\beta}$ coefficients from their full conditional distribution given by

$$\boldsymbol{\beta}_i \mid \text{rest} \sim N_p \left(\boldsymbol{b}_{n_i}, \boldsymbol{B}_n \right), \tag{9}$$

where $\boldsymbol{B}_n = \left(\boldsymbol{\Phi}^T \boldsymbol{\Phi}/\tau_\varepsilon^2 + \boldsymbol{\Omega} \right)^{-1}$ and $\boldsymbol{b}_{n_i} = \boldsymbol{B}_n \left(\boldsymbol{\Phi}^T \boldsymbol{Y}_i/\tau_\varepsilon^2 + \boldsymbol{\Omega}\boldsymbol{\mu} \right)$ for all $i = 1 \ldots n$. The full conditional distribution of $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu} \mid \text{rest} \sim N_p \left(\boldsymbol{m}, \boldsymbol{M} \right), \tag{10}$$

where $\boldsymbol{M} = \boldsymbol{I}_p/\sigma_\mu^2 + n\boldsymbol{\Omega}$ and $\boldsymbol{m} = \boldsymbol{M}^{-1}\boldsymbol{\Omega}\sum_{i=1}^n \boldsymbol{\beta}_i$. For the τ_ε^2, we sample from

$$\tau_\varepsilon^2 \mid \text{rest} \sim \text{IG} \left(\frac{nr + a}{2}, \frac{1}{2} \left(b + \sum_{i=1}^n \left(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{\beta}_i \right)^T \left(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{\beta}_i \right) \right) \right). \tag{11}$$

Finally, since only $\boldsymbol{\beta}$'s are directly influenced by the graphical component, we can rewrite the joint full conditional of $(G, \boldsymbol{\Omega})$ as

$$P(G, \boldsymbol{\Omega}|\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n, \tau_\varepsilon^2, \boldsymbol{\mu}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_n) = P\left(G, \boldsymbol{\Omega} \mid \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_n, \boldsymbol{\mu} \right). \tag{12}$$

As introduced in Section 2, even when the graph is known, sampling from a G-Wishart distribution poses computational issues. As a consequence, any method based on the computation of the marginal full conditional of the graph,

$$P(G \mid \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_n) \propto \pi(G) \frac{I_G \left(d + n, D + U \right)}{I_G \left(d, D \right)}, \tag{13}$$

where $U = \sum_{i=1}^n (\boldsymbol{\beta}_i - \boldsymbol{\mu})(\boldsymbol{\beta}_i - \boldsymbol{\mu})^T$, is not efficient because requires the computation of both the prior and posterior normalizing constant of density (4).

To sample from (12) while avoiding computation of (13), we employ the BDMCMC approach developed by Mohammadi and Wit (2015) for Gaussian graphical model selection. The BDMCMC algorithm is based on a continuous time Markov process, where edges are added or removed via birth or death events and the time between jumps is taken to be an exponential random variable with specific rates. Birth and death events are characterized as follows:

- Death: Each edge $e = (j, k)$ between nodes j and k, where $j < k$, dies independently of the others as a Poisson process with rate $\delta_e(\boldsymbol{\Omega})$. The overall death rate is then $\delta(\boldsymbol{\Omega}) = \sum_{e \in E} \delta_e(\boldsymbol{\Omega})$. When the death of an edge e occurs, then the process jumps to a new state $(G^{-e}, \boldsymbol{\Omega}^{-e})$, where $G^{-e} = (V, E\backslash\{e\})$ and $\boldsymbol{\Omega}^{-e} \in P_{G^{-e}}$.

- Birth: A new edge $e = (j, k)$ between nodes j and k, where $j < k$, is born independently of the others as a Poisson process with birth rate $\lambda_e(\boldsymbol{\Omega})$. The overall birth rate is then $\lambda(\boldsymbol{\Omega}) = \sum_{e \in \bar{E}} \lambda_e(\boldsymbol{\Omega})$. If the birth of an edge $e \in \bar{E}$ occurs, then the process jumps to a new state $(G^{+e}, \boldsymbol{\Omega}^{+e})$, where $G^{+e} = (V, E \cup \{e\})$ and $\boldsymbol{\Omega}^{+e} \in P_{G^{+e}}$.

Birth and death processes are assumed to be independent. Hence, the intensity of the jump process is given by $w(\boldsymbol{\Omega}) = 1/\left(\lambda(\boldsymbol{\Omega}) + \delta(\boldsymbol{\Omega})\right)$. We recall that this implies that the time between two consecutive jumps is exponentially distributed with mean $w(\boldsymbol{\Omega})$ that we refer to as the weight of the state. The choice of birth and death rates is made such that the balance conditions given by Preston (1977) hold, so that the stationary distribution of the birth-death process is the full conditional distribution in (12). The last step is then updating the precision matrix $\boldsymbol{\Omega}$ using the exact sampler described in Mohammadi and Wit (2015) and originally developed by Lenkoski (2013). The BDMCMC for customary Gaussian graphical model selection is implemented by the R package BDgraph (Mohammadi and Wit, 2019).

Algorithm 1: MCMC sampler.

For each iteration:

Step 1. Sample the regression parameters

    1.1 Sample the parameters $\boldsymbol{\beta}$'s from their full conditional in (9).

    1.2 Sample the parameter μ from its full conditional in (10).

    1.3 Sample the parameter τ_ε^2 from its full conditional in (11).

Step 2. Sample $(G, \boldsymbol{\Sigma}^{-1})$ from their full conditional using the BDMCMC algorithm and save the expected holding time $w(\boldsymbol{\Omega})$.

To sum up, our sampling strategy is given in Algorithm 1. As far as Step 1 is concerned, it consists in sampling from conjugate full conditionals described above. Then we exploit the BDgraph function to complete Step 2. We call the function for only one iteration using as argument the parameters updated at Step 1. In other words, we run the BDgraph function as if data were $\boldsymbol{\beta_1} - \boldsymbol{\mu}, ..., \boldsymbol{\beta_n} - \boldsymbol{\mu}$. We also create an auxiliary structure to store all the explored graphs and its associated expected holding time $w(\boldsymbol{\Omega})$. The R code implementing Algorithm 1 is available under request.

3.1. Posterior graph

Given a MCMC output of size T, posterior summaries of the precision matrix and the graph need to take into account the fact that Step 2 of Algorithm 1 involves sampling from a continuous time Markov process. In particular, the posterior estimate of $\mathbf{\Omega}$ is based on the Rao-Blackwell theorem (Cappé et al., 2003), yielding to a posterior estimate that is

$$\hat{\mathbf{\Omega}} = \frac{\sum_{t=1}^{T} w_t(\mathbf{\Omega}) \mathbf{\Omega}_t}{\sum_{t=1}^{T} w_t(\mathbf{\Omega})}. \tag{14}$$

As far as the posterior graph is concerned, one approach for selecting the graph is to use the maximum a posteriori estimate, i.e., the highest posterior probability graph. However, this approach is not generally reliable since the space of possible graphs is quite large and any particular graph may be encountered only a few times in the course of the MCMC sampling. A more practical solution is instead to estimate the marginal posterior edge inclusion probabilities. Namely, given the MCMC output, the posterior inclusion probabilities are estimated as

$$\hat{p}_{jk} = \frac{\sum_{t=1}^{T} \mathbb{1}\left((j,k) \in E_t\right) w(\mathbf{\Omega}_t)}{\sum_{t=1}^{T} w(\mathbf{\Omega}_t)}, \tag{15}$$

where $\mathbb{1}\left((j,k) \in E_t\right)$ is the indicator function representing the inclusion of the edge linking nodes j and k and drawn at iteration t. Here, we select the graph containing all edges whose posterior inclusion probability in (15) exceeds a given threshold s. In particular, we compare two different thresholds. The first one is $s = 0.5$, in analogy with the median probability model of Barbieri and Berger (2004), originally proposed in the linear regression setting. The second is based on the Bayesian False Discovery rate (BFDR; Müller et al. 2007; Peterson et al. 2015)

$$\text{BFDR} = \frac{\sum_{j<k}(1 - \hat{p}_{jk})\mathbb{1}\left(\hat{p}_{jk} \geq s\right)}{\sum_{j<k} \mathbb{1}\left(\hat{p}_{jk} \geq s\right)}, \tag{16}$$

where s is selected so that BFDR in equation (16) is below 0.05.

4. Simulation study

We carried out a simulation study to evaluate the ability of our methodology to recover the structure of the generating graph as well as to compare

it with the state of the art. In particular, we analyzed three different experiments: in Section 4.1 we focus on the ability of our technique to learn the structure of the underlying graph, simulating data from two different structures. Further, Section 4.2 shows a numerical comparison with Yang et al. (2017), in terms of both the ability of the two techniques to estimate the covariance matrix and on the smoothing performance.

4.1. Structural learning

We consider a graph with $p = 40$ nodes to mimic the dimension of real data, see Section 5; then we sample a precision matrix $\boldsymbol{\Omega}$ from a G-Wishart($d = 3, \boldsymbol{D} = \boldsymbol{I}_p$) constrained by the graph and draw the $\boldsymbol{\beta}$ parameters from (3). Finally, we simulate 50 datasets from the sampling distribution in (2), where $n = 200$ and $\tau_\epsilon^2 = 0.01$.

We fit model (8) on the simulated datasets and all the results are based on the output of Algorithm 1 with 60,000 iterations and a burn-in of 10,000. As far as the choice of the hyperparameters is concerned, we set the prior variance of μ equal to $\sigma_\mu^2 = 100$ and the parameters of the prior distribution of variance τ_ε^2 equal to $a = 10$ and $b = 0.001$, i.e., standard vague priors. We computed the edge posterior inclusion probabilities and provide a posterior estimate of the graph using the two thresholds described in Section 3.1.

We evaluate the ability of recovering the underlying graph, using the Structural Hamming Distance (SHD, Tsamardinos et al. 2006), which represents the number of edge insertions and deletions needed to transform the estimated graph into the true graph. Clearly, lower values of SHD correspond to better performances. For sake of comparison the SHD has been standardized over the maximum number of edges.

Experiment 1: non-structured graph. In the first experiment we consider a non-structured graph with sparsity level of 0.3, i.e., each edge has a probability equal to 0.3 of being in the graph, as shown in Figure 3.

We fit model (8) assuming different hyperparameters for the graph prior in (7) with prior probability 0.2, 0.3, 0.4, 0.5 and 0.6. Figure 4 shows the boxplot of the standardized SHD values over the 50 replicates. From the latter we note that the best posterior estimate of the graph is obtained when the 0.5 threshold is employed and the graph prior probability is 0.3, i.e., the value equal to the true sparsity level. However, the median graph is quite sensitive to prior choices while the BFDR yields to more stable results.
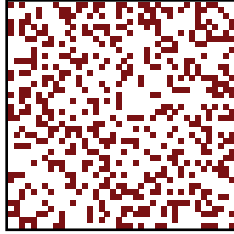
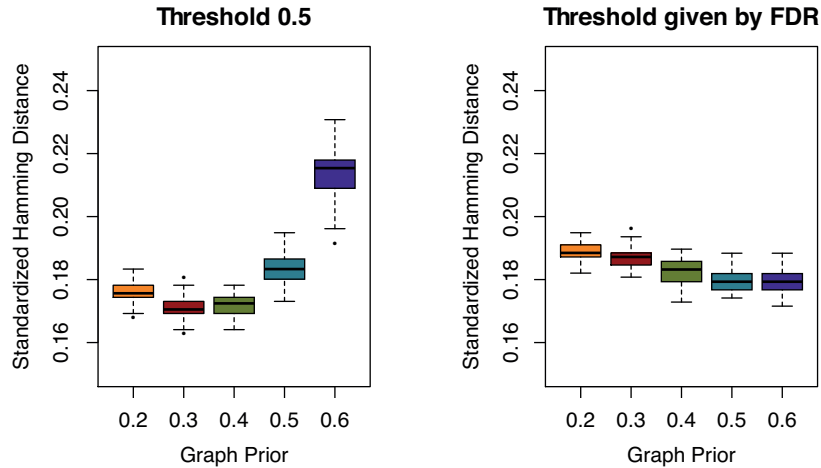Figure 3: The true non-structured graph of simulation Experiment 1.

Figure 4: Boxplot of the standardized structural Hamming distances over the 50 replicates for Experiment 1.

Experiment 2: clustered graph. The second experiment considers a block structured graph with different sparsity levels, namely 0.1, 0.2 and 0.3, as shown in Figure 5. We fit model (8) assuming the uniform distribution in (6) over the graph space. The results are summarized in Figure 6. Overall, the performance of recovering the true graph improves as the sparsity level increases, as expected. The posterior estimate of the graph based on the BFDR outperforms the posterior median graph, i.e., the one obtained with the 0.5 threshold.

4.2. Comparison with an existing method

Here, we offer a comparison with the results obtained using the Bayesian model proposed by Yang et al. (2017) for simultaneously smoothing multiple functional curves, referred to as the Bayesian framework with approximations by basis functions (BABF). As mentioned in the Introduction, Yang
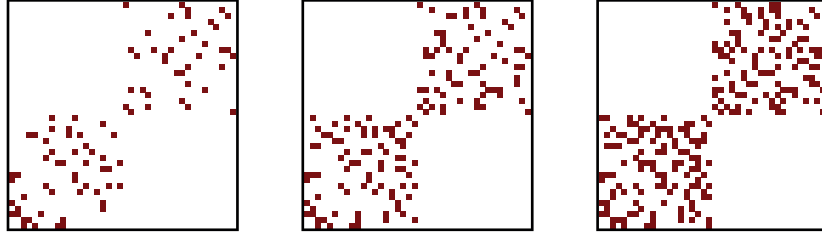
14

Figure 5: The true clustered graphs of Experiment 2 with sparsity levels of 0.1, 0.2, and 0.3 in the left, middle and right panel, respectively.
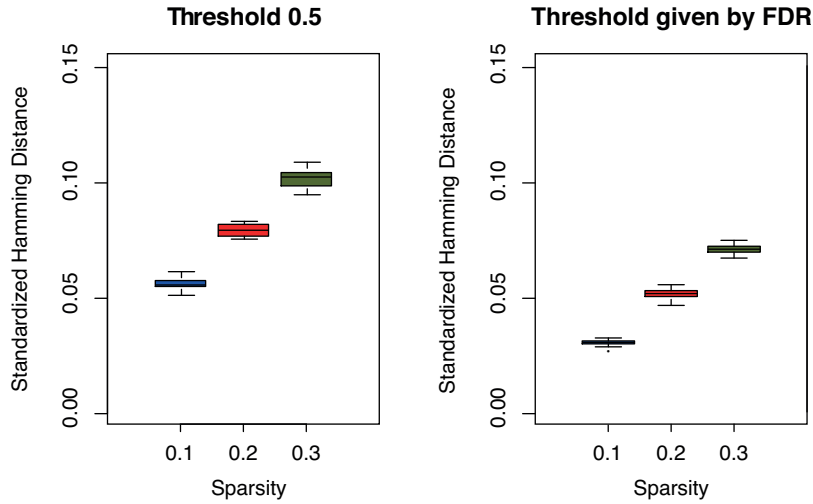

Figure 6: Boxplot of the standardized structural Hamming distances over the 50 replicates for Experiment 2.

et al. (2017) assumed that all functional observations come from Gaussian processes that are approximated through basis functions. Thus, the induced Bayesian hierarchical model for the basis expansion coefficients has the same likelihood of model (8). Moreover, the authors adopted a Gaussian-Wishart prior for basis expansion coefficients. In particular, their covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}$ is assumed to follow, a priori, an Inverse-Wishart prior centered on a Matérn covariance function, i.e., $\boldsymbol{\Sigma} \sim \mathrm{IW}(d, \sigma^2 \boldsymbol{A})$, where $\boldsymbol{A}$ is a Matérn correlation matrix. We refer the reader to Lindgren et al. (2011) for a detailed discussion on the link between Gaussian processes in the Matérn class and the corresponding Gaussian Markov random fields with a sparse precision matrix.

Following Yang et al. (2017), we generate 200 functional curves from a Gaussian process with mean $\mu(t) = 3\sin(4t)$ and covariance $\mathbf{\Sigma}(s,t) = 5\operatorname{Mat\acute{e}rn}(|s-t|, \rho = 0.5, \nu = 0.5)$ where $s, t \in [0, \pi/2]$ and $\operatorname{Mat\acute{e}rn}(d; \rho, \nu) = (\Gamma(\nu)2^{\nu-1})^{-1}(\sqrt{2\nu}d/\rho)^{\nu}K_{\nu}(\sqrt{2\nu}d/\rho)$, with $\Gamma(\cdot)$ the gamma function and $K_{\nu}(\cdot)$ the modified Bessel function of the second kind. We fit both model (8) and BABF to the simulated curves using $p = 20$ basis functions. Easy implementation of BABF model is available in the `BDFA Matlab` toolbox (Yang and Ren, 2019). To provide a fair comparison we match prior hyperparameters by setting $d = 5$ and $\mathbf{D} = 5\mathbf{I}_p$ in (4). Again, we place a uniform prior over the graph space. We repeated the analysis on 50 simulated datasets. In the following, we report the average indices as well as their standard deviations (sd).

As far as concerns the smoothing performances, both methods are effective and comparable. Indeed, the root mean square error index is, on average, equal to 1.096 (sd 0.019) and 1.187 (sd 0.013) for our approach and BABF, respectively. Figure 7 shows the true covariance matrix $\mathbf{\Sigma}_{\text{true}}$ (left panel) and the corresponding estimate $\hat{\mathbf{\Sigma}}$ obtained from BABF (middle panel) and from model (8) (right panel). As in Mohammadi and Wit (2015) and Wang (2010), we use the Kulbach-Leibler divergence (Kullback and Leibler, 1951) as an index to asses the distance between $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}_{\text{true}}$, that is

$$\mathrm{KL}(\mathbf{\Sigma}_{\text{true}}, \hat{\mathbf{\Sigma}}) = \frac{1}{2}\left[\operatorname{tr}\left(\mathbf{\Sigma}_{\text{true}}^{-1}\hat{\mathbf{\Sigma}}\right) - p - \log\left(\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{\Sigma}_{\text{true}}|}\right)\right]. \tag{17}$$

The average score in equation (17) achieved by our method is 2.046 (sd 0.9487), which is much lower than the one obtained using BABF model, that is equal to 12.407 (sd 1.3762). Note that the prior distribution of $\mathbf{\Sigma}$ under BABF is centered on a Matérn structure which strongly characterizes the posterior estimate of the covariance matrix. Rather, our approach does not provide any prior information about the underlying structure of the covariance matrix, but allows to properly learn it from the data.

5. Analysis of fruit purees data

Food safety and authenticity is extremely important to guarantee access to good-quality and healthy products. According to the Food Authenticity Assurance Organisation, "food authenticity is the process of irrefutably proving that a food or food ingredient is in its original, genuine, verifiable
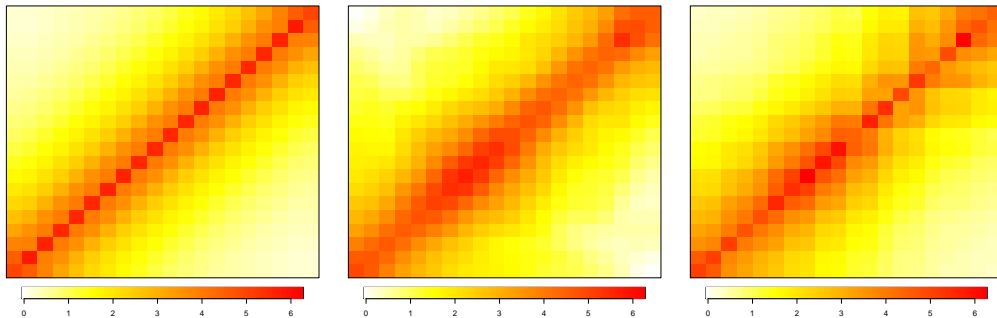
Figure 7: True covariance matrix $\mathbf{\Sigma}_{\text{true}}$ (left panel) and posterior mean of $\mathbf{\Sigma}$ obtained from BABF (middle panel) and from model (8) (right panel).

and intended form as declared and represented". Although authenticity is not a novel concept, it has recently received particular attention due to the increasing number of food frauds and adulteration. As an example, the UK National Food Crime Unit has increased from 796 cases in 2015 to $1, 193$ in 2019, with 364 notices in the first three months of 2019[1].

Traditionally, all strategies and techniques developed to detect food adulterants and, more in general, to study the composition of substances, were based on chemistry. The amount of marker compounds in a test material was determined and then compared with the values obtained for previously documented authentic material. The resulting test is often time consuming and expensive. Moreover, this approach is prone towards errors, since food adulterers are becoming increasingly complex to detect. As a consequence, the pressing demand for rapid and inexpensive tests lead to the adoption of novel techniques. Nowadays, strategies based on mid-infrared spectroscopy coupled with chemometrics or statistical techniques have been used to investigate the substance composition and detect, for instance, the presence of adulterants in coffe (Downey et al., 1997), fruit purees (Kemsley et al., 1996; Holland et al., 1998) or minced beef (Meza-Márquez et al., 2010).

In this work, we analyze the spectrum of absorbance of 351 fruit purees prepared using exclusively fresh whole strawberries (without the addition of adulterants) measured on an equally-spaced grid of 235 wavelengths and then normalized with respect to the area under the curve; see Figure 8. Data are publicly available at https://data.mendeley.com/datasets/frrv2yd9rg.

_____

[1]National Food Crime Unit of the Food Standards Agency.

The shape of the spectra is very similar for all the curves: they are all defined
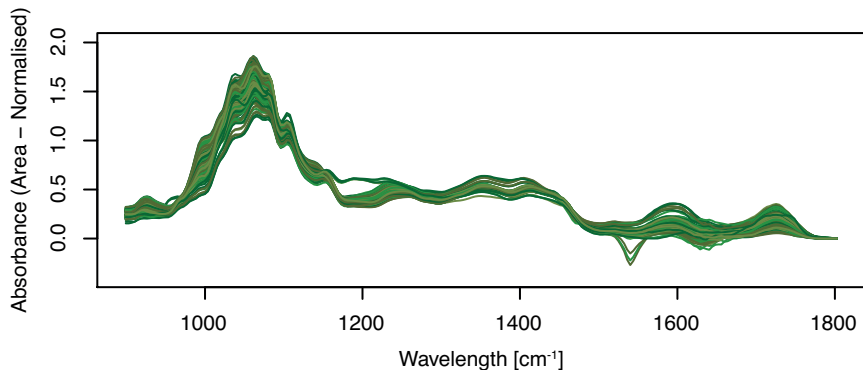

Figure 8: Plot of 351 spectra of absorbance of pure fruit purees, measured at 235 different wavelengths of the middle-infrared spectra.

by a recognizable peak at wavelengths 1000-1200 cm^{-1} and few secondary others, like the ones around 1600 cm^{-1} and 1700 cm^{-1}. A further description of the data can be found in Holland et al. (1998).

The chemical analysis of such a complex spectrum is not trivial. Indeed we are not dealing with a pure substance that would absorb at a specific wavelength, but those purees have a more heterogeneous composition which leads to overlapping effects. The goal of the analysis is to provide useful insights about which wavelength bands are related to the different components of the purees, and more specifically about the conditional dependence structure between wavelength bands.

We fit model (8) by assuming $p = 40$ cubic B-spline basis functions. Different values of p were explored, obtaining similar results with respect to the ones described in this section, so they are not reported here for brevity. The final choice of $p = 40$ faced the trade off between good fitting of the smoothed curves and limited computational burden. Figure 9 presents an example of two smoothed curves (dotted lines) compared to the original ones (solid lines). Notice that the smoothed curves follow the shape of the original ones, smoothing away some pointwise variability. The left panel of Figure 10 shows the posterior mean of the coefficients while the right panel displays the estimated precision matrix Ω.

The top panel of Figure 11 displays the posterior estimate of the graph obtained using the BFDR criterion. Specifically, the top-right panel shows the estimated graph with nodes numbered and colored according to different
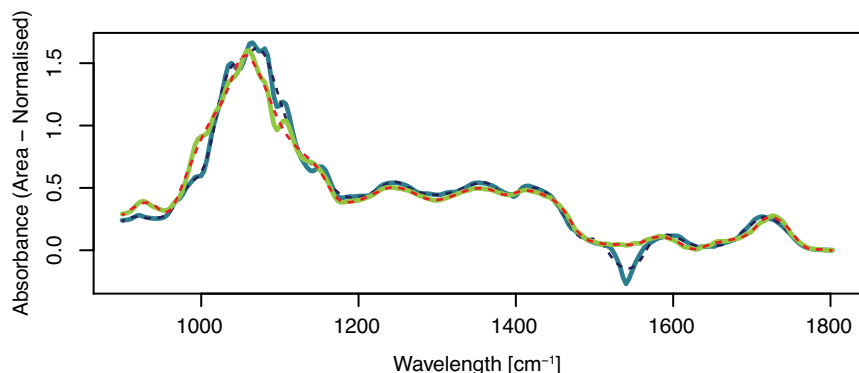
18

Figure 9: An example of two smoothed curves through model (8); solid lines represent original curves while dotted lines are the smoothed curves.

portions of the spectra as highlighted in the bottom panel of the figure. The different colors are associated to different portions of the domain that correspond to the peaks of the signal. The graph offers an interpretation of the dependence structure among the wavelength bands and so the components of the purees. The estimated graph is quite sparse and characterized by a block structure. The blocks are located mainly close to the diagonal, with some extra-diagonal exceptions. The prevalence of blocks close to the diagonal is expected, since close wavelength bands are likely to be associated to similar components. The blocks far away from the diagonal are of interest, since they suggest that some components associated to very different wavelengths might be related.

The large block in the range 899.3 - 1493.6 cm^{-1} corresponds to the main peak of the curves. The big size of this block suggests that the interactions between the corresponding bands are not only due to the overlap of spline supports, rather they may also depend on the absorbance effects of several species overlay. We also notice two small off-diagonal blocks that correspond to the interactions between the bands at the highest peak and the others corresponding to lower peaks at around 1564.9 cm^{-1} and 1700 cm^{-1}, respectively. Indeed with the infrared spectrometer, the same species can exhibit peaks at different wavelengths and our analysis identifies which peaks should be analyzed together. Finally, looking at the graph representation, we note that the isolated nodes in the graph are associated to intervals where the curves are rather flat (e.g., nodes 17-18-19), which means that the concentration of the corresponding substances is negligible. On the other hand,
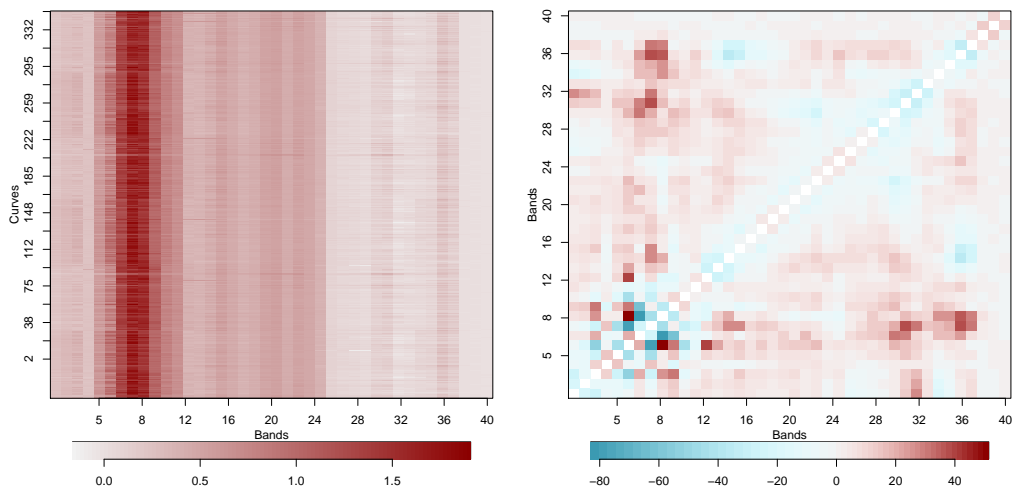
Figure 10: Posterior mean of all $\boldsymbol{\beta}$ coefficients (left panel) and the estimated precision matrix (right panel); in the latter plot, the main diagonal has been omitted.

most of the identified links connects different peaks. In particular the main peak, which represents the mostly present substance, is the center of the network, all other peaks are connected to this one.

6. Summary and future works

In this paper, we have investigated the hidden inter-dependencies among different bands of the absorbance spectrum of a given substance. To accomplish this task, we have reinterpreted the problem as a smoothing procedure of functional data by means of suitable B-spline basis functions. Our Bayesian smoothing model allows sharing of information across curves, meanwhile it induces sparsity on the shared precision matrix of the smoothing coefficients through the underlying graph structure.

An efficient sampling strategy has been designed to provide posterior inference as well as graphical model selection. In our simulation study, the selection of the graph based on the BFDR outperformed the one based on the median summary in recovering the underlying graph. Also, we have compared the results obtained from our approach and the ones from the model proposed by Yang et al. (2017). While the two methods are comparable in terms of smoothing, our approach outperforms the competitor in recovering the dependence structure.
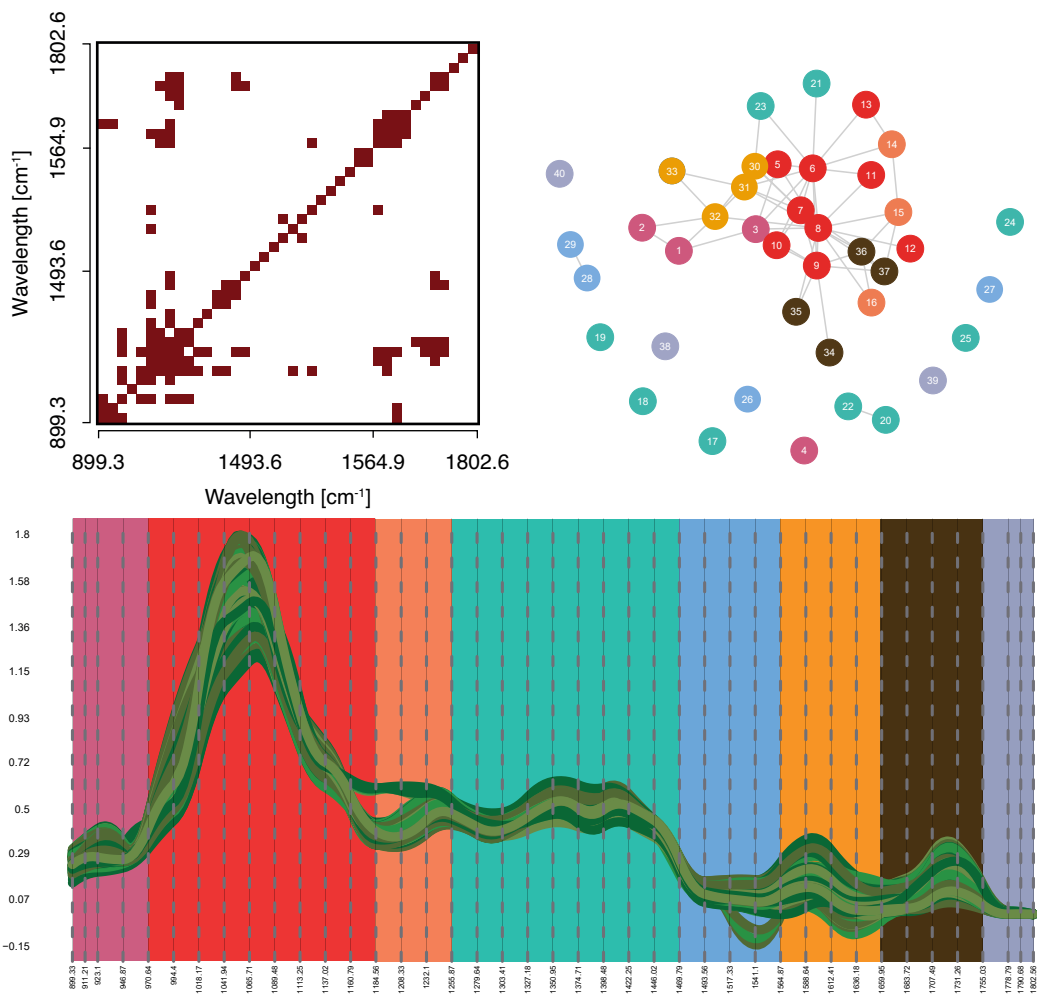
Figure 11: Top panels: posterior estimate of the graph obtained using the BFDR criterion. Nodes are colored according to different portions of the spectra highlighted in the bottom panel.

We have illustrated our methodology on a real dataset, studying the independence structure among the bands of absorbance spectrum of fruit purees. The estimated graph is sparse and equipped with a block structure; the physical interpretation of such result provides useful insights towards the understanding of interactions of the components of the substance.

Future research will address three main challenges. First, since species can exhibit peaks at different wavelengths, a clustered graph might be ex-

pected a priori. Therefore, future works will find us in investigating a prior distribution on the graph space that induces a block structure. On the computational side, local moves represent a limitation of the sampling strategy, indeed the birth and death process adds or removes only one edge at each step. Hence, it would be interesting to explore how to introduce also global moves for the Markov chain. Finally, the model can be generalized to functional discriminant analysis in order to compare different substances and to detect adulterated ones.

Acknowledgements

References

E. K. Kemsley, J. K. Holland, M. Defernez, R. H. Wilson, Detection of adulteration of raspberry purees using infrared spectroscopy and chemometrics, Journal of Agricultural and Food Chemistry 44 (1996) 3864–3870.

G. Downey, R. Briandet, R. H. Wilson, E. K. Kemsley, Near- and mid-infrared spectroscopies in food authentication: Coffee varietal identification, Journal of Agricultural and Food Chemistry 45 (1997) 4357–4361.

J. K. Holland, E. K. Kemsley, R. H. Wilson, Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purées, Journal of the Science of Food and Agriculture 76 (1998) 263–269.

O. G. Meza-Márquez, T. Gallardo-Velázquez, G. Osorio-Revilla, Application of mid-infrared spectroscopy with multivariate analysis and soft independent modeling of class analogies (SIMCA) for the detection of adulterants in minced beef, Meat Science 86 (2010) 511–519.

S.-Y. Lee, W. Zhang, X.-Y. Song, Estimating the covariance function with functional data, British Journal of Mathematical and Statistical Psychology 55 (2002) 247–261.

J. O. Ramsay, B. W. Silverman, Functional data analysis, Springer, 2005.

L. Xiao, V. Zipunnikov, D. Ruppert, C. Crainiceanu, Fast covariance estimation for high-dimensional functional data, Statistics and Computing 26 (2016) 409–421.

J. Yang, H. Zhu, T. Choi, D. D. Cox, Smoothing and mean–covariance estimation of functional data with a Bayesian hierarchical model, Bayesian Analysis 11 (2016) 649–670. doi:10.1214/15-BA967.

J. Yang, D. D. Cox, J. S. Lee, P. Ren, T. Choi, Efficient Bayesian hierarchical functional data analysis with basis function approximations using Gaussian–Wishart processes, Biometrics 73 (2017) 1082–1091.

A. Pini, S. Vantini, The interval testing procedure: a general framework for inference in functional data analysis, Biometrics 72 (2016) 835–845.

S. Lang, A. Brezger, Bayesian P-splines, Journal of Computational and Graphical Statistics 13 (2004) 183–212.

C. Crainiceanu, A. Goldsmith, Bayesian functional data analysis using WinBUGS, Journal of Statistical Software 32 (2010) 1–33.

D. Telesca, L. Y. T. Inoue, Bayesian hierarchical curve registration, Journal of the American Statistical Association 103 (2008) 328–339.

S. L. Lauritzen, Graphical models, Oxford University Press, Oxford, 1996.

A. Dobra, C. Hans, B. Jones, J. R. Nevins, G. Yao, M. West, Sparse graphical models for exploring gene expression data, Journal of Multivariate Analysis 90 (2004) 196–212.

C. Peterson, F. C. Stingo, M. Vannucci, Bayesian inference of multiple Gaussian graphical models, Journal of the American Statistical Association 110 (2015) 159–174.

L. S. Tan, A. Jasra, M. De Iorio, T. M. Ebbels, Bayesian inference for multiple Gaussian graphical models with application to metabolic association networks, The Annals of Applied Statistics 11 (2017) 2222–2251.

Y. Ni, F. C. Stingo, V. Baladandayuthapani, Sparse multi-dimensional graphical models: A unified Bayesian framework, Journal of the American Statistical Association 112 (2017) 779–793.

A. Cremaschi, R. Argiento, K. Shoemaker, C. Peterson, M. Vannucci, Hierarchical normalized completely random measures for robust graphical modeling, Bayesian Analysis 14 (2019) 1271–1301.

L. Paci, G. Consonni, Structural learning of contemporaneous dependencies in graphical VAR models, Computational Statistics & Data Analysis 144 (2020) 106880.

F. Liu, S. Chakraborty, F. Li, Y. Liu, A. C. Lozano, Bayesian regularization via graph Laplacian, Bayesian Analysis 9 (2014) 449–474.

S. Chakraborty, A. C. Lozano, A graph Laplacian prior for Bayesian variable selection and grouping, Computational Statistics & Data Analysis 136 (2019) 72–91.

Q. Cai, J. Kang, T. Yu, Bayesian network marker selection via the thresholded graph Laplacian Gaussian prior, Bayesian Analysis 15 (2020) 79 – 102.

H. Zhu, N. Strawn, D. B. Dunson, Bayesian graphical models for multivariate functional data, Journal of Machine Learning Research 17 (2016) 1–27.

B. Li, E. Solea, A nonparametric graphical model for functional data with application to brain networks based on fMRI, Journal of the American Statistical Association 113 (2018) 1637–1655.

X. Qiao, S. Guo, G. M. James, Functional graphical models, Journal of the American Statistical Association 114 (2019) 211–222.

X. Qiao, C. Qian, G. M. James, S. Guo, Doubly functional graphical models in high dimensions, Biometrika 107 (2020) 415–431.

C. B. Peterson, F. C. Stingo, M. Vannucci, Joint Bayesian variable and graph selection for regression models with network-structured predictors, Statistics in Medicine 35 (2016) 1017–1031.

A. Roverato, Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models, Scandinavian Journal of Statistics 29 (2002) 391 – 411.

G. Letac, H. Massam, Wishart distributions for decomposable graphs, Annals of Statistics 35 (2007) 1278–1323.

C. Uhler, A. Lenkoski, D. Richards, Exact formulas for the normalizing constants of Wishart distributions for graphical models, The Annals of Statistics 46 (2018) 90 – 118. URL: https://doi.org/10.1214/17-AOS1543. doi:10.1214/17-AOS1543.

B. Jones, C. Carvalho, A. Dobra, C. Hans, C. Carter, M. West, Experiments in stochastic computation for high-dimensional graphical models, Statistical Science 20 (2005) 388–400.

J. Scott, C. Carvalho, Feature-inclusion stochastic search for Gaussian graphical models, Journal of Computational and Graphical Statistics 17 (2008) 790–808.

P. Giudici, R. Castelo, Improving Markov chain Monte Carlo model search for data mining, Machine learning 50 (2003) 127–158.

A. Bhadra, B. K. Mallick, Joint high-dimensional Bayesian variable and covariance selection with an application to eQTL analysis, Biometrics 69 (2013) 447–457.

A. Mohammadi, E. C. Wit, Bayesian structure learning in sparse Gaussian graphical models, Bayesian Analysis 10 (2015) 109–138.

C. Preston, Spatial birth and death processes, Bulletin of the International Statistical Institute 46 (1977) 371–391.

A. Lenkoski, A direct sampler for G-Wishart variates, Stat 2 (2013) 119–128.

R. Mohammadi, E. Wit, BDgraph: An R package for Bayesian structure learning in graphical models, Journal of Statistical Software 89 (2019) 1–30.

O. Cappé, C. P. Robert, T. Rydén, Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers, Journal

of the Royal Statistical Society: Series B (Statistical Methodology) 65 (2003) 679–700.

M. M. Barbieri, J. O. Berger, Optimal predictive model selection, Annals of Statistics 32 (2004) 870–897.

P. Müller, G. Parmigiani, K. Rice, FDR and Bayesian multiple comparisons rules, in: J. M. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West (Eds.), Bayesian Statistics 8, Oxford University Press, Oxford, 2007, pp. 1–19.

I. Tsamardinos, L. E. Brown, C. F. Aliferis, A. W. Moore, The max-min hill-climbing Bayesian network structure learning algorithm, Machine Learning 65 (2006) 31–78.

F. Lindgren, H. Rue, J. Lindström, An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73 (2011) 423–498.

J. Yang, P. Ren, BFDA: A MATLAB toolbox for Bayesian functional data analysis, Journal of Statistical Software, Articles 89 (2019) 1–21.

H. Wang, Sparse seemingly unrelated regression modelling: Applications in finance and econometrics, Computational Statistics & Data Analysis 54 (2010) 2866–2877.

S. Kullback, R. A. Leibler, On information and sufficiency, The annals of mathematical statistics 22 (1951) 79–86.