Bayesian functional linear regression with sparse step functions

Paul-Marie Grollemund, Christophe Abraham, Meïli Baragatti, Pierre Pudlo

Bayesian functional linear regression with sparse step functions

Paul-Marie Grollemund[1,2], Christophe Abraham [2], Meïli Baragatti[2], and Pierre Pudlo[3]

[1]*IMAG UMR 5149, Université de Montpellier, CNRS, Place E. Bataillon, 34095 Montpellier CEDEX, France* *paul-marie.grollemund@umontpellier.fr*
[2]*MISTEA UMR 729, INRA, Montpellier SupAgro, Place Pierre Viala, 34060 Montpellier CEDEX, France* *christophe.abraham@supagro.fr meili.baragatti@supagro.fr*
[3]*I2M UMR 7373, Aix-Marseille Université, CNRS, Centrale Marseille, Rue F. Joliot Curie, 13453 Marseille CEDEX 13, France* *pierre.pudlo@univ-amu.fr*

Abstract

The functional linear regression model is a common tool to determine the relationship between a scalar outcome and a functional predictor. This paper focuses on the interpretability of the estimation of the coefficient function of this model. We propose a Bayesian functional Linear regression with Sparse Step functions (Bliss). The aim of the method is to provide interpretable estimates: the Bayesian model is based on an adaptive decomposition of the coefficient function into sparse and simple functions. A Bayes estimator is constructed with a specific loss function. The method is compared to its competitors on simulated datasets and is illustrated on a black Périgord truffle dataset.

MSC 2010 subject classifications: Primary 62F15; Secondary 62J05.

Keywords: Bayesian regression, function data, interpretability, parsimony.

1 Introduction

Linear models have been adapted to handle functional predictors and/or functional outcomes, see Ramsay and Silverman (2005). The scalar-on-function case names the framework where the outcome y is a simple scalar and the covariate $x(t)$ lies in a functional space, $L^2(\mathcal{T})$ say, where $\mathcal{T}$ is an interval of $\mathbb{R}$. In the linear framework the outcome y is predicted with

$$\widehat{y} = \mu + \int_{\mathcal{T}} \beta(t)x(t)\,dt, \tag{1}$$

where μ is the intercept and $\beta(t)$ the coefficient function. To fit the model, one can assume that the coefficient function $\beta(t)$ belongs to a Hilbert space and estimate the coefficients of an expansion of $\beta(t)$ on a truncated Hilbert basis, see for instance Cardot et al. (2003), James et al. (2009) and Zhao et al. (2012). The truncation of the expansion can be data-driven, as in Cardot et al. (1999), Yuan and Cai (2010) and Zhu et al. (2014), among others. The analysis can be conducted in Bayesian framework, see Brown et al. (2001), Crainiceanu et al. (2005) and Goldsmith et al. (2011). For more details on the scalar-on-function methods, we refer the reader to Reiss et al. (2016).

As often in regression issues, we can fit the model either to achieve the best prediction accuracy, or to recover a model that is interpretable. The latter goal has been rarely tackled in the functional regression literature with two notable exceptions (James et al., 2009; Zhou et al., 2013). Interpretability usually consists in seeking a parsimonious coefficient function $\beta(t)$ which has a simple shape and is null on some intervals of $\mathcal{T}$. Searching for intervals where $\beta(t)$ is null can be seen as a variable selection procedure. Indeed, if $\beta(t) = 0$ for all $t \in I$, then, conditionally on $\{x(t), t \notin I\}$, the output y is independent of $\{x(t), t \in I\}$. This notion of interpretability is of major interest in functional linear regression, as interpreting the coefficient function can be cumbersome.

The Flirti method of James et al. (2009) obtains an interpretable estimate by focusing on sparse functions with sparse derivatives. The authors resort to a penalty such as lasso to recover parsimony. The resulting estimate is associated to an interpretable function $\beta(t)$ with sparse derivatives. Nevertheless the Flirti method is difficult to calibrate: its numerical results depend heavily on tuning parameters. The authors propose to rely on cross-validation to set these parameters. From our experience, Flirti's estimate is so sensitive to the values of the tuning parameters that we can miss the range of good values with cross-validation. Zhou et al. (2013) propose a two-stage method to estimate the coefficient function. Preliminarily, $\beta(t)$ is expanded onto a B-spline basis to reduce the dimension of the model. The first stage is the estimation of the resulting coefficients using a lasso method to find the null intervals. Then, the second stage refines the estimation of the null intervals and estimates the magnitude of $\beta(t)$ for the rest of the support. Another approach to obtain parsimony is to rely on Fused lasso (Tibshirani et al., 2005): if we discretize both the covariate functions and the coefficient function as described in James et al. (2009), the penalization of Fused lasso induces parsimony in the coefficients.

Throughout the paper we assume that covariates $x(t)$ are functions lying in $L^2(\mathcal{T})$. We focus on the fixed design regression problem and drop the conditioning on the design from all notations. We define an interpretable estimate as a step function which is nonnull only on a few intervals. For the sake of interpretability and parsimony, we resort only to the simplest functions, namely step functions, although the method could be extended to other basis functions such as B-splines.

The Bayesian functional linear regression with sparse step functions we proposed provides good estimates in terms of interpretability. Furthermore, the prior distribution can easily include some constraints on the coefficient function, such as lying in the set of sparse step functions and permits the use of prior knowledge from experts if available. And, last but not least, we can evaluate the uncertainty on the estimate of $\beta(t)$ with a full Bayesian procedure.

In the present paper, Section 2 describes a hierarchical model and details the construction of the estimator and its implementation. We also introduce a visualization tool of the posterior distribution of the coefficient function. In Section 3, we compare the Bliss method to its competitors (James et al., 2009; Tibshirani et al., 2005) on simulated datasets and study its robustness with respect to hyperparameters. In Section 4, we illustrate the method on a black Périgord truffle dataset, and conclude with a discussion in Section 5.

2 The Bliss method

We present the hierarchical Bayesian model in Section 2.1 and two Bayes estimates in Section 2.2. The implementation and visualization details are given in Section 2.3.

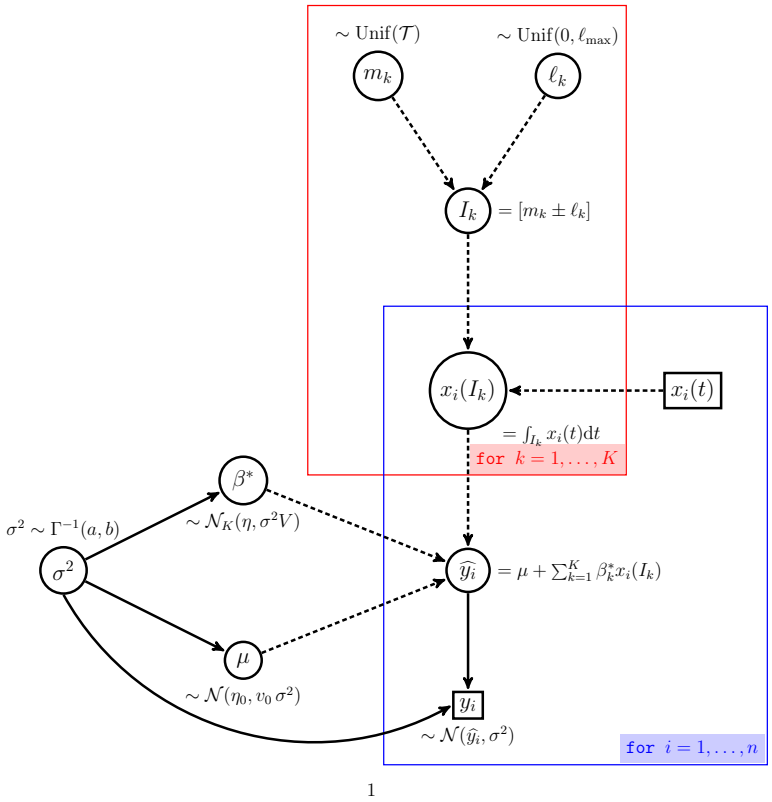

Figure 1: The full Bayesian model. *The coefficient function* $\beta(t) = \sum_{k=1}^K \beta_k^* \mathbb{1}\{t \in I_k\}$ *defines both a projection the covariate functions* $x_i(t)$ *onto* $\mathbb{R}^K$ *by integrating over each interval* I_k *and a prediction* $\widehat{y}_i$ *which depends on the vector* $\beta^* = (\beta_1^*, \ldots, \beta_K^*)$ *and the intercept* μ. *Hyperparameters of the prior are* $\eta_0, v_0, \eta, V, a, b, \ell_{max}$ *and* K.

2.1 Model

Assume we have observed n independent replicates $y_1, \ldots, y_n$ of the outcome, explained with the covariate functions $x_1(t), \ldots x_n(t)$ respectively. The dataset is $\mathcal{D} = \{(y_i, x_i(t)), i = 1, \ldots, n\}$. We resort to the Gaussian likelihood defined as

$$y_i | \mu, \beta(t), \sigma^2 \overset{\text{ind}}{\sim} \mathcal{N}\left(\mu + \int_{\mathcal{T}} \beta(t) x_i(t) \mathrm{d}t\,,\ \sigma^2\right), \qquad i = 1, \ldots, n. \tag{2}$$

For parsimony we seek the coefficient function $\beta(t)$ in the following sets of sparse step functions

$$\mathcal{E}_K = \left\{ \sum_{k=1}^K \beta_k^* \mathbb{1}\{t \in I_k\} :\ I_1, \ldots, I_K \text{ intervals } \subset \mathcal{T}, \beta_1^*, \ldots, \beta_K^* \in \mathbb{R} \right\} \tag{3}$$

where K is a hyperparameter that counts the number of intervals required to defined the function. Note that we do not make any assumptions regarding the intervals $I_1, \ldots, I_K$. First they do not form a partition of $\mathcal{T}$. As a consequence, a function $\beta(t)$ in $\mathcal{E}_K$ is piecewise constant and null outside the union of the intervals I_k, $k = 1, \ldots, K$ which ensures parsimony and interpretability. Second the intervals $I_1, \ldots, I_K$ can even overlap to ease the

parametrization of the intervals: we do not have to add constraints on the parametrization to remove possible overlaps.

Now if we pick a function $\beta(t) \in \mathcal{E}_K$ with

$$\beta(t) = \sum_{k=1}^{K} \beta_k^* \mathbb{1}\{t \in I_k\}, \tag{4}$$

the integral of the covariate functions $x_i(t)$ against $\beta(t)$ becomes a linear combination of partial integrals of the covariate functions over the intervals I_k and we predict y_i with

$$\widehat{y_i} = \mu + \sum_{k=1}^{K} \beta_k^* \, x_i(I_k), \quad \text{where } x_i(I_k) = \int_{I_k} x_i(t)\mathrm{d}t.$$

Thus, knowing the intervals $I_1, \ldots, I_K$, we face a multivariate linear model with the usual Gaussian likelihood.

It remains to set a parametrization on $\mathcal{E}_K$ and a prior distribution. Each interval I_k is parametrized with its center m_k and its half length ℓ_k:

$$I_k = [m_k - \ell_k, m_k + \ell_k] \cap \mathcal{T}. \tag{5}$$

As a result, when K is fixed, the parameter of the model is

$$\theta = (m_1, \ldots, m_K, \ell_1, \ldots, \ell_K, \mu, \beta_1^*, \ldots, \beta_K^*, \sigma^2).$$

Let Θ_K denote the range of the parameter. Note that $\dim(\Theta_K) = 3K + 2$. The prior distribution on Θ_K is

$$\begin{aligned}
\mu|\sigma^2 &\sim \mathcal{N}\left(\eta_0, v_0\sigma^2\right), \\
\beta^*|\sigma^2 &\sim \mathcal{N}_K\left(\eta, \sigma^2 V\right), \\
\sigma^2 &\sim \Gamma^{-1}(a, b) \\
m_k &\overset{i.i.d.}{\sim} \mathrm{Unif}(\mathcal{T}), \quad k = 1, \ldots, K, \\
\ell_k &\overset{i.i.d.}{\sim} \mathrm{Unif}(0, \ell_{\max}), \quad k = 1, \ldots, K,
\end{aligned} \tag{6}$$

where $\beta^* = (\beta_1^*, \ldots, \beta_K^*)$. The resulting full Bayesian modelling is given in Figure 1 and depends on hyperparameters which are $\eta_0, v_0, \eta, V, a, b, \ell_{\max}$ and K. Note that the last hyperparameter drives the number of intervals, thus the dimension of Θ_K. We denote by $\pi_K(\theta)$ and $\pi_K(\theta|\mathcal{D})$ the prior and the posterior distributions.

2.2 Loss and Bayes estimates

Bayes estimates are obtained by minimizing a loss function integrated with respect to the posterior distribution, see Robert (2007). With the default quadratic loss, the Bayes estimate is

$$\widehat{\beta}_{L^2}(\cdot) \in \underset{d(\cdot)\in L^2(\mathcal{T})}{\arg\min} \iint (\beta_\theta(t) - d(t))^2 \, \mathrm{d}t \, \pi_K(\theta|\mathcal{D})\mathrm{d}\theta \tag{7}$$

where $\beta_\theta(t)$ is coefficient function as parametrized in (4). At least heuristically $\widehat{\beta}_{L^2}(\cdot)$ is the average of $\beta_\theta(\cdot)$ over the posterior distribution $\pi_K(\theta|\mathcal{D})$, though the average of functions taking values in $L^2(\mathcal{T})$ under some probability distribution is hard to define (using either Bochner or Pettis integrals). In this simple setting we can claim the following (see Appendix A.1 for the proof).

Proposition 1. *Let $\|\cdot\|$ be the norm of $L^2(\mathcal{T})$. If $\int \|\beta_\theta(\cdot)\|\,\pi_K(\theta|\mathcal{D})\mathrm{d}\theta < \infty$, then the estimate defined by*

$$\widehat{\beta}_{L^2}(t) = \int \beta_\theta(t)\pi_K(\theta|\mathcal{D})\mathrm{d}\theta, \quad t \in \mathcal{T}, \tag{8}$$

is in $L^2(\mathcal{T})$ and solve the optimization problem (7).

Averages such as (8) belong to the closure of the convex hull of the support $\mathcal{E}_K$ of the posterior distribution. We can prove (see Proposition 3 in Appendix A.3) that the convex hull of $\mathcal{E}_K$ is the set $\mathcal{E} = \cup_{K=1}^{\infty}\mathcal{E}_K$ of step functions on $\mathcal{T}$, and the closure of $\mathcal{E}$ is $L^2(\mathcal{T})$. Hence the only guaranty we have on $\widehat{\beta}_{L^2}$ as defined in (8) is that $\widehat{\beta}_{L^2}$ lies in $L^2(\mathcal{T})$, a much larger space than the set of step functions.

To obtain an estimate lying in the set of step functions, namely $\mathcal{E}$, we can consider the projection of $\widehat{\beta}_{L^2}$ onto the set $\mathcal{E}_{K_0}$ for a suitable value of K_0 possibly different to K. However, due to the topological properties of $L^2(\mathcal{T})$ and $\mathcal{E}_{K_0}$, the projection of $\widehat{\beta}_{L^2}$ onto the set $\mathcal{E}_{K_0}$ do not always exist (see Appendix A.3). To address this problem, we introduce a subset $\mathcal{E}_{K_0}^{\varepsilon}$ of $\mathcal{E}_{K_0}$, where $\varepsilon > 0$ is a tuning parameter. Let $\mathcal{F}^{\varepsilon}$ denote the set of step functions $\beta(t) \in L^2(\mathcal{T})$ which can be written as

$$\beta(t) = \sum \beta_k^{\dagger}\mathbb{1}\{t \in J_k\}$$

where the intervals J_k's are mutually disjoint and each of lengths greater than ε. The set $\mathcal{E}_{K_0}^{\varepsilon}$ is now defined as $\mathcal{F}^{\varepsilon} \cap \mathcal{E}_{K_0}$. By considering this set, we remove from $\mathcal{E}_K$ the step functions which have intervals of very small length, and we can prove the following.

Proposition 2. *Let $K_0 \geq 1$ and $\varepsilon > 0$.*

(i) The function $d(\cdot) \mapsto \|d(\cdot) - \widehat{\beta}_{L^2}(\cdot)\|^2$ admits a minimum on $\mathcal{E}_{K_0}^{\varepsilon}$. Thus a projection of $\widehat{\beta}_{L^2}(\cdot)$ onto this set, defined by

$$\widehat{\beta}_{K_0}^{\varepsilon}(\cdot) \in \underset{d(\cdot)\in\mathcal{E}_{K_0}^{\varepsilon}}{\arg\min}\ \|d(\cdot) - \widehat{\beta}_{L^2}(\cdot)\|^2, \tag{9}$$

always exist.

(ii) The estimate $\widehat{\beta}_{K_0}^{\varepsilon}(\cdot)$ is a true Bayes estimate with loss function

$$L_{K_0}^{\varepsilon}(d(\cdot), \beta(\cdot)) = \begin{cases} \|d(\cdot) - \beta(\cdot)\|^2 = \int_{\mathcal{T}} (\beta(t) - d(t))^2\,\mathrm{d}t & \text{if } \beta \in \mathcal{E}_{K_0}^{\varepsilon}, \\ +\infty & \text{otherwise.} \end{cases} \tag{10}$$

That is to say

$$\widehat{\beta}_{K_0}^{\varepsilon}(\cdot) \in \underset{d(\cdot)\in L^2(\mathcal{T})}{\arg\min} \int L_{K_0}^{\varepsilon}(d(\cdot), \beta_\theta(\cdot))\pi_K(\theta|\mathcal{D})\mathrm{d}\theta.$$

2.3 Implementation

The posterior distribution can be written as

$$\begin{aligned}\pi_K(\theta|\mathcal{D}) \propto \left(\sigma^2\right)^{-\left(a+\frac{n+K+1}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left\|y - \mu\mathbb{1}_n - x_{\cdot}(I_{\cdot})\beta^*\right\|^2\right\} \\ \times \exp\left\{-\frac{1}{2\sigma^2}\left[2b + \frac{1}{v_0}(\mu - \eta_0)^2 + \|\beta^* - \eta\|_{V^{-1}}^2\right]\right\}\end{aligned} \tag{11}$$

where $\|\cdot\|$ is the canonical Euclidean norm, $\|\cdot\|^2_{V^{-1}}$ is the definite positive quadratic form of matrix V^{-1} and $x_\cdot(I_\cdot)$ the matrix whose (i,k)-entry is

$$x_i(I_k) = \int_{m_k-\ell_k}^{m_k+\ell_k} x_i(t)\mathrm{d}t.$$

We rely on a Monte Carlo sampler to approximate the posterior and on numerical optimization to compute the Bayes estimates. As often in hierarchical models, sampling from the posterior distribution $\pi_K(\theta|\mathcal{D})$ can be done with a Gibbs algorithm (see, e.g., Robert and Casella, 2013, Chapter 7). The details of the MCMC algorithm are given in Appendix B.1.

Now let $\theta(s)$, $s = 1,\dots,N$, denote the output of the MCMC sampler after the burn-in period. Its remains to approximate $\widehat{\beta}_{L^2}(\cdot)$ and $\widehat{\beta}^{\mathcal{E}}_{K_0}(\cdot)$ based on the MCMC sample. First, the Monte Carlo approximation of (8) is given by

$$\widehat{\beta}_{L^2}(t) \approx \frac{1}{N}\sum_{s=1}^{N} \beta_{\theta(s)}(t).$$

And the more interesting Bayes estimate $\widehat{\beta}^{\mathcal{E}}_{K_0}(\cdot)$ can be computed by minimizing

$$\left\| d(\cdot) - \frac{1}{N}\sum_{s=1}^{N} \beta_{\theta(s)}(\cdot) \right\|^2$$

over the set $\mathcal{E}^{\mathcal{E}}_{K_0}$. To this end we run a Simulated Annealing algorithm (Kirkpatrick et al., 1983), described in Appendix B.2.

We also provide a striking graphical display of the posterior distribution on the set $\mathcal{E}_K$ with a heat map. More precisely, the aim is to sketch all marginal posterior distributions $\pi^t_K(\cdot|\mathcal{D})$ of $\beta(t)$ for any value of $t \in \mathcal{T}$ in one single figure. To this end we introduce the probability measure Q on $\mathcal{T}\times\mathbb{R}$ defined as follows. Its marginal distribution over $\mathcal{T}$ is uniform, and knowing the value t of the first coordinate, the second coordinate is distributed according to the posterior distribution of $\beta(t)$. In other words,

$$(t,b) \sim Q \quad \Longleftrightarrow \quad t \sim \mathrm{Unif}(\mathcal{T}),\ b|t \sim \pi^t_K(\cdot|\mathcal{D}).$$

We can easily derive an empirical approximation of Q from the MCMC sample $\{\theta(s)\}$ of the posterior. Indeed, the first marginal distribution of Q, namely $\mathrm{Unif}(\mathcal{T})$ can be approximated by a regular grid t_i, $i = 1,\dots,M$. And, for each value of i, set $b_{is} = \beta_{\theta(s)}(t_i)$, $s = 1,\dots,N$. The resulting empirical measure is

$$\widehat{Q} = \frac{1}{M\,N}\sum_{i=1,\dots,M}\ \sum_{j=1,\dots,N} \delta_{(t_i,b_{is})},$$

where $\delta_{(t,b)}$ is the Dirac measure at (t,b). The graphical display we propose is representing $\widehat{Q}$ with a heat map on $\mathcal{T}\times\mathbb{R}$. Each small area of $\mathcal{T}\times\mathbb{R}$ is thus colored according to its $\widehat{Q}$-probability. This should be done cautiously as the marginal posterior distribution $\pi^t_K(\cdot|\mathcal{D})$ has a point mass at zero: $\pi^t_K(b = 0|\mathcal{D}) > 0$ by construction of the prior distribution. The color scale can be any monotone function of the probabilities, in particular non linear functions to handle the atom at 0. Many examples are provided in Section 3 below, for instance the first column of plots in Figure 4.

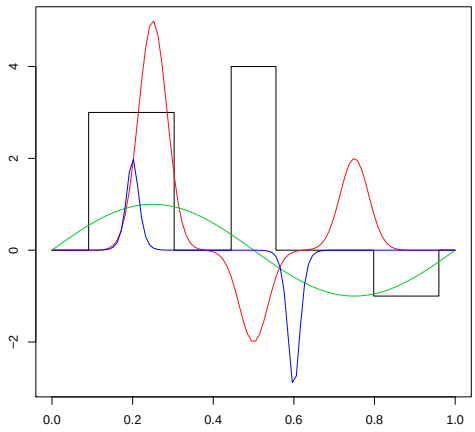

Figure 2: The simulated coefficient functions. *The black (resp. red, green and blue) curve corresponds to the coefficient function of Shape 1 (resp. 2, 3 and 4).*

3 Simulation study

In this section, the performance of Bliss is evaluated and compared to two competitors: Fused lasso (Tibshirani et al., 2005) and Flirti (James et al., 2009), using simulated datasets.

3.1 Simulation scheme

First of all, we describe how we generate different datasets to apply the methods. The support of the covariate curves x_i is $\mathcal{T} = [0, 1]$, observed on a regular grid $(t_j)_{j=1,\dots,p}$ on $\mathcal{T}$, for a given p. We simulate p-multivariate Gaussian vectors x_i, $i = 1, \dots, n$, corresponding to the values of curves x_i for the observation times $(t_j)_j$. The covariance matrix Σ of these Gaussian vectors is given by

$$\Sigma_{i,j} = \sqrt{\Sigma_{i,i}\Sigma_{j,j}} \exp\left(-\zeta^2 (t_i - t_j)^2\right), \qquad \text{for } i \text{ and } j \text{ from 1 to } p,$$

where the coefficient ζ tunes the autocorrelation of the $x_i(t)$. Four different shapes are considered for the functional coefficient β, given in Figure 2. The first one is an interpretable function (Shape 1), the second one is smooth and is null on small intervals of $\mathcal{T}$ (Shape 2), the third one is smooth without null interval (Shape 3), and the last one is nonnull only on small intervals of $\mathcal{T}$ (Shape 4).

- Shape 1: $\beta(t) = 3 \times \mathbb{1}\{t \in [0.1, 0.3]\} + 4 \times \mathbb{1}\{t \in [0.45, 0.55]\} - 1 \times \mathbb{1}\{t \in [0.8, 0.95]\}$.
- Shape 2: $\beta(t) = 5 \times e^{-20(t-0.25)^2} - 2 \times e^{-20(t-0.5)^2} + 2 \times e^{-20(t-0.75)^2}$.
- Shape 3: $\beta(t) = \sin(2\pi t)$.
- Shape 4: $\beta(t) = 8 \times (2 + e^{20-100t} + e^{100t-20})^{-1} - 12 \times (2 + e^{60-100t} + e^{100t-60})^{-1}$.

The outcomes y_i are calculated according to (1) with an additional noise following a centred Gaussian distribution with variance σ^2. The value of σ^2 is fixed such that the signal to noise

ratio is equal to a chosen value r. Datasets are simulated for $n = 50$, $\mu = 1$ and for the following different values of p, ζ and r:

- $p = 100, 200, 500,$
- $\zeta = 1, 1/3, 1/5,$
- $r = 1, 3, 5.$

3.2 Choice of parameters

For this study, we apply Bliss to these different datasets and fix the hyperparameters η_0, v_0, η, V, a, b, $\ell_{\max}$ and K so that the prior is weakly informative. The hyperparameter K, which is the number of intervals, can be fixed at 10 or 15. A high value of K makes β less interpretable. On the other hand, if K is too small then the estimate may miss nonnull intervals. In the following, the default K value is 10. For the others hyperparameters, the default values are:

- $\eta_0 = 0$ and $v_0 = 10^2$,
- $\eta = (0, \ldots, 0)$ and V is a diagonal matrix with diagonal element $v = 10^2 \times \hat{\mathbb{V}}(y)/\operatorname{var}_x$ where $\operatorname{var}_x = \min_{j=1,\ldots,p} \left(\hat{\mathbb{V}}(x(t_j))\right)$ and $\hat{\mathbb{V}}(u) = \sum_{i=1}^{n}(u_i - \bar{u})^2/(n-1)$, for any vector u of length n.
- $a = 0.1$, $b = 0.1$ and
- $\ell_{\max} = \text{length}(\mathcal{T})/5$.

For the Gibbs Sampler, the default iteration number value N is 50000 and the default burn-in B is 2000. For the Simulated Annealing algorithm, the default iteration number is 100000. The cooling schedule is deterministic and logarithmic decreasing as prescribed by Bélisle (1992).

3.3 Quality criterion to compare the different methods

We use several different measures to examine the performance of Bliss and compare it to competitors. The first one considered is the estimation error with respect to the true coefficient function β:

$$\text{Err} = \|\hat{\beta}(\cdot) - \beta(\cdot)\|^2 = \int_{\mathcal{T}} \left(\hat{\beta}(t) - \beta(t)\right)^2 \mathrm{d}t.$$

Note that this measure does not take into account the interpretability of the estimate. Furthermore, three measures depending on false null *FN*, false nonnull *FnN*, true null *TN* and true nonnull *TnN* counts are also considered to verify whether the estimate recovers the support of $\beta(t)$. An observation time t_j is considered as

- a false nonnull if $\hat{\beta}(t_j) \neq 0$ and $\beta(t_j) = 0$,
- a false null if $\hat{\beta}(t_j) = 0$ and $\beta(t_j) \neq 0$,

Table 1: Numerical results of Bliss, Flirti and Fused lasso on the Simulated Datasets.

				Estimation Error			Correct Classification Rate			Dataset
Shape	p	r	ζ	Bliss	Fused lasso	Flirti	Bliss	Fused lasso	Flirti	
	100	5	1	**0.597**	0.978	3.848	0.760	**0.800**	0.520	1
	200	5	1	**0.468**	0.845	0.554	**0.920**	0.795	0.890	2
	500	5	1	0.470	**0.418**	12.606	**0.906**	0.828	0.644	3
	100	3	1	1.783	1.659	**1.638**	**0.710**	0.480	0.480	4
1	100	1	1	2.286	**0.990**	1.901	0.530	**0.640**	0.630	5
	100	5	$\sqrt{3}$	4.795	9.373	**4.129**	0.450	**0.680**	0.570	6
	100	5	$1/5$	12.916	**12.639**	$> 10^6$	**0.530**	0.480	0.490	7
	100	5	1	**0.689**	0.781	0.980	**0.850**	0.520	0.620	8
	200	5	1	**0.706**	26.604	8.474	**0.830**	0.565	0.535	9
	500	5	1	**0.601**	1.383	1.613	**0.790**	0.644	0.644	10
	100	3	1	0.547	**0.346**	0.461	0.730	0.800	**0.810**	11
2	100	1	1	**0.792**	1.075	$> 10^6$	**0.800**	0.470	0.450	12
	100	5	$\sqrt{3}$	**1.044**	1.175	1.160	**0.560**	0.490	0.470	13
	100	5	$1/5$	**0.988**	1.629	2.088	**0.730**	0.330	0.530	14
	100	5	1	0.080	**0.060**	0.078	0.770	0.870	**0.960**	15
	200	5	1	0.070	**0.044**	0.059	0.680	**0.855**	0.780	16
	500	5	1	0.079	**0.035**	0.072	0.748	0.910	**0.912**	17
	100	3	1	**0.045**	0.049	0.056	0.720	**0.950**	0.770	18
3	100	1	1	0.344	**0.154**	0.237	0.540	**0.860**	0.570	19
	100	5	$\sqrt{3}$	0.088	**0.073**	0.085	0.640	**0.960**	**0.960**	20
	100	5	$1/5$	**0.083**	46.578	$> 10^2$	**0.860**	0.290	0.240	21
	100	5	1	0.205	0.115	**0.083**	0.870	0.840	**0.980**	22
	200	5	1	**0.140**	0.172	0.480	**0.895**	0.890	0.850	23
	500	5	1	0.165	**0.067**	0.076	0.942	0.960	**0.964**	24
	100	3	1	0.202	**0.098**	0.152	**0.800**	0.780	0.710	25
4	100	1	1	0.197	0.089	**0.035**	0.670	**0.970**	0.920	26
	100	5	$\sqrt{3}$	0.277	**0.234**	0.376	0.380	0.700	**0.840**	27
	100	5	$1/5$	0.284	**0.235**	0.346	0.540	0.770	**0.840**	28

				False Null Rate			False Nonnull Rate			Dataset
Shape	p	r	ζ	Bliss	Fused lasso	Flirti	Bliss	Fused lasso	Flirti	
	100	5	1	**0.088**	0.203	0.480	0.318	**0.195**	NaN	1
	200	5	1	**0.040**	0.194	0.170	0.118	0.217	**0.000**	2
	500	5	1	**0.040**	0.218	0.382	0.145	**0.082**	0.250	3
	100	3	1	**0.129**	NaN	NaN	**0.362**	0.520	0.520	4
1	100	1	1	0.432	0.333	**0.258**	0.492	**0.384**	0.420	5
	100	5	$\sqrt{3}$	0.538	0.371	**0.344**	0.557	**0.136**	0.464	6
	100	5	$1/5$	**0.391**	0.500	0.490	**0.493**	0.611	0.531	7
	100	5	1	0.045	**0.000**	0.095	**0.232**	0.516	0.455	8
	200	5	1	**0.204**	0.437	0.443	**0.109**	0.416	0.517	9
	500	5	1	**0.280**	0.386	0.346	**0.000**	0.223	0.370	10
	100	3	1	0.181	**0.142**	0.256	0.339	0.254	**0.000**	11
2	100	1	1	0.203	**0.000**	NaN	**0.195**	0.540	0.550	12
	100	5	$\sqrt{3}$	**0.351**	0.450	0.466	**0.492**	0.550	0.557	13
	100	5	$1/5$	**0.088**	0.676	0.454	**0.363**	0.666	0.583	14
	100	5	1	**0.880**	0.846	NaN	**0.013**	0.022	0.040	15
	200	5	1	0.888	**0.862**	0.909	**0.000**	0.023	0.025	16
	500	5	1	**0.910**	0.939	1.000	**0.010**	0.029	0.033	17
	100	3	1	**0.875**	1.000	0.880	**0.000**	0.040	0.013	18
3	100	1	1	0.977	**0.916**	0.933	0.053	0.034	**0.018**	19
	100	5	$\sqrt{3}$	**0.900**	NaN	NaN	**0.000**	0.040	0.040	20
	100	5	$1/5$	**0.857**	0.971	0.973	**0.023**	0.068	0.083	21
	100	5	1	**0.000**	**0.000**	0.011	0.448	0.500	**0.062**	22
	200	5	1	**0.026**	0.112	0.134	0.361	**0.071**	0.357	23
	500	5	1	**0.007**	0.045	0.041	0.245	**0.000**	**0.000**	24
	100	3	1	**0.000**	0.069	0.017	**0.555**	0.607	0.651	25
4	100	1	1	**0.000**	0.034	0.086	0.673	**0.000**	**0.000**	26
	100	5	$\sqrt{3}$	**0.000**	**0.000**	0.095	0.724	0.652	**0.500**	27
	100	5	$1/5$	**0.000**	**0.000**	0.160	0.741	**0.589**	NaN	28

Section 3.1 describes the simulation scheme of the datasets. Section 3.3 describes the criteria: Estimation Error, Correct Classification Rate, False Null Rate and False Nonnull Rate. The bold font indicates the best result for each criterion and each dataset.

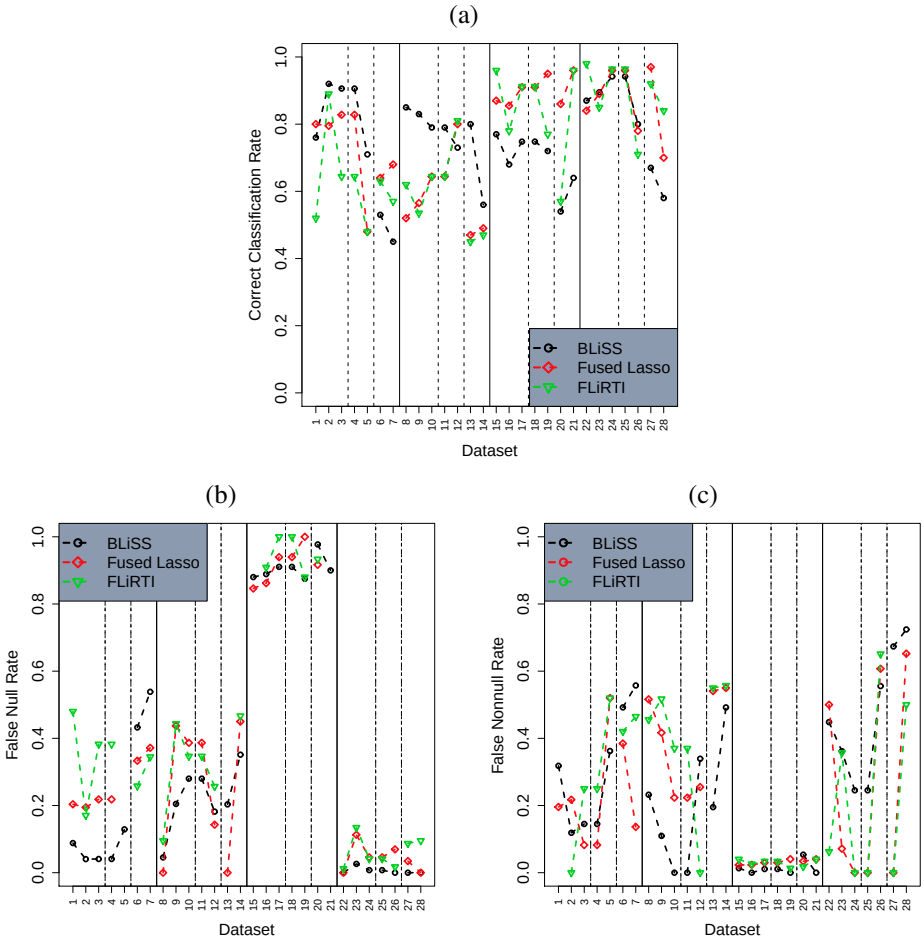

Figure 3: Comparison of the ability of Bliss, Flirti and Fused lasso to recover the support of $\beta(t)$. *Plot* (a) *(resp.* (b) *and* (c)*) gives the results of these methods in terms of Correct Classification Rate (resp. False Null Rate and False Nonnull Rate), on the different datasets described in Section 3.1. Each plot is divided into four parts by solid vertical lines; each part corresponds to a specific shape. Each part is divided into three sections by dashed vertical lines; where each section allows one parameter to vary. For example, the first section of plot* (a) *shows the CCR on Shape 1 where the parameter p varies.*

- a true nonnull if $\hat{\beta}(t_j) \neq 0$ and $\beta(t_j) \neq 0$,
- a true null if $\hat{\beta}(t_j) = 0$ and $\beta(t_j) = 0$.

To handle cases where $\beta(t)$ is almost always nonnull, we define the pseudo-support as the intervals where $|\beta(t)| > \max_{t \in \mathcal{T}} (|\beta(t)|)/20$. The considered measures are

$$CCR = \frac{TN + TnN}{TN + FN + TnN + FnN}, \; FnNR = \frac{FnN}{FnN + TnN} \text{ and } FNR = \frac{FN}{FN + TN},$$

which are the correct classification rate, the false nonnull rate and the false null rate respectively.

3.4 Comparison to competitors

Bliss and its competitors are applied on the datasets described in Section 3.1, and are compared using the previous accuracy measures. The numerical results regarding the support are presented in Table 1 and summarized in Figure 3. Some plots of the estimated coefficient functions are given in Figures 4, 5 6, 7 and 8. All first columns of these figures concern Bliss, and represent the posterior distributions of the coefficient functions as explained in Section 2.3. We have superposed two Bayes estimates, namely $\widehat{\beta}_{L^2}(t)$ of Proposition 1 which is the blue smooth function of t, and $\widehat{\beta}^{\varepsilon}_{K_0}(t)$ of Proposition 2 which is the green step function.

It appears that the numerical results of the three methods have the same order of magnitude, see Table 1. Although the three methods may have different accuracy, depending on the shape of the coefficient function that generated the dataset. For instance, Shape 1 (Figures 4 and 5) and Shape 2 (Figure 6) appear to be well adapted to Bliss while Shape 4 (Figure 8) is more adapted to Fused lasso and Flirti. This trend is corroborated by the numerical results of Table 1.

As expected we observe that when the autocorrelation increases the three methods have difficulties to estimate the coefficient function: either the range of an interval can be shifted from the true range (see, e.g., Figures 5(b) and 5(d)), or its magnitude can be misestimated (see, e.g., Figures 5(b) and 5(d)). Moreover, as the noise increases the methods have difficulties to recover the coefficient function (see, e.g., Figures 5(a) and 5(c)) and it induces a high variability of the estimators, especially for the Flirti estimate, see, e.g., Figure 6(d). Note that in this simulation study each method has its own features. For instance, Fused lasso is quite stable (the only exception being plotted in Figure 7(d)); even when the autocorrelation is high Fused lasso gives decent estimates. Flirti is rather unstable and the estimates can be erratic, see, e.g., Figures 4(d), 6(d), 7(d). In Figure 4(c), the Flirti estimate is even a null function. In contrast Bliss always gives interpretable estimates and enables a representation of the posterior distribution.

Concerning the behaviour of the methods when the autocorrelation increases, Figures 5(b) and 5(d) show that the methods can misestimate the range of intervals. Both Figures illustrate that when the magnitude of a given interval is misestimated, it is counterbalanced with an artefact on a correlated interval: if $\beta(t)$ is overestimated on an interval, it is often underestimated on a neighboring interval. Note that Fused lasso gives a narrow estimate (Figure 5(d)), which fails to detect the whole range of nonnull intervals. Indeed, only small parts of the nonnull intervals are selected, which results in an overestimation of the magnitude of the coefficient function on these intervals.

Concerning the behaviour of the methods when the noise increases, Figures 5(a) 5(c), 6(b) and 6(d) emphasize the difficulty to estimate the coefficient function.

Concerning the interpretability, Figures 6(b), 6(d), 7(a) and 7(c) show that in opposition to Flirti and Fused lasso estimates, the Bliss estimates are always interpretable in the sense that Bliss always returns null intervals. Moreover, Bliss enables a representation of the posterior distribution and of its variance, which supplies a level of credibility of the estimate. In order to give ideas of power of the estimate, we can contrast results of Figures 5(b) and 7(b) where the posterior variance is high with, e.g., Figure 4(b) where the posterior variance is much smaller.

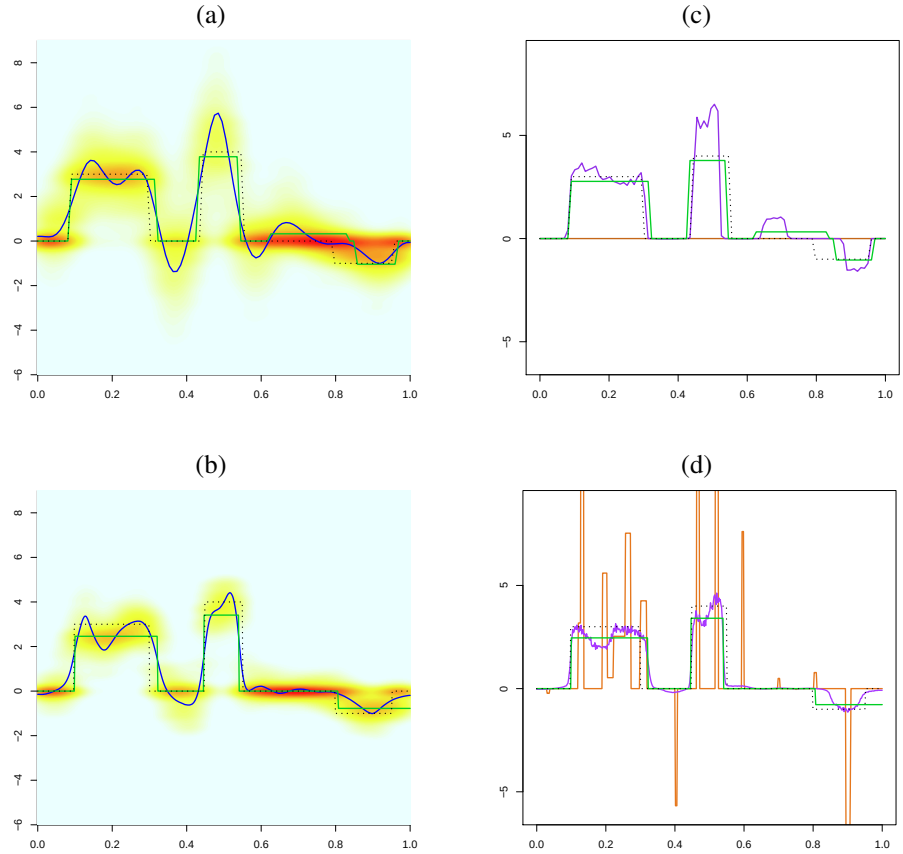

Figure 4: Bliss, Fused lasso and Flirti estimates when the true function $\beta(t)$ is interpretable (Datasets 1 and 3 generated from Shape 1). *Each row of plots corresponds to a given dataset. In every plot the dotted line ($\cdots\cdots$) is the coefficient function $\beta(t)$ used to generate the data. The first column of plots corresponds to the Bliss method. The set of marginal posterior distributions of $\beta_\theta(t)$ for any $t \in [0,1]$ is displayed by the heat map, see Section 2.3: white (resp. red) areas correspond to low (resp. high) posterior densities. The green curve —— is the Bliss estimate $\widehat{\beta}^{\varepsilon}_{K_0}(t)$ as defined in Proposition 2, and the blue curve —— is the posterior expected value $\widehat{\beta}_{L^2}(t)$ of Proposition 1. The second column of plots compares different estimators: the green line —— (resp. violet —— and orange ——) is the estimate of Bliss (resp. Fused lasso and Flirti).*

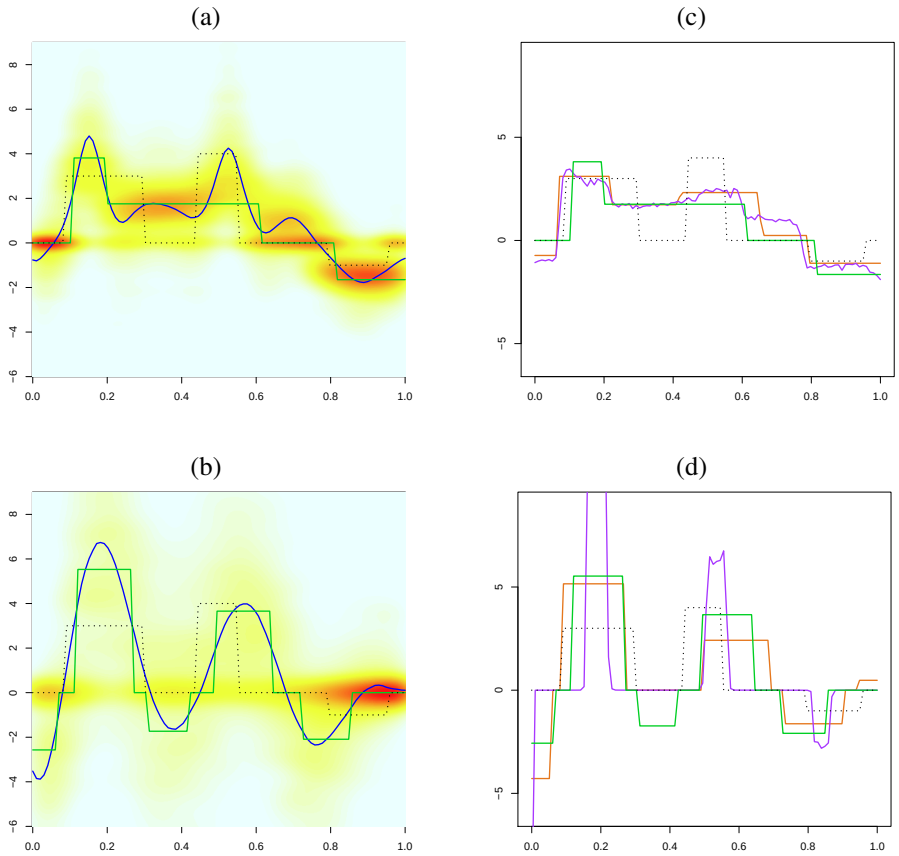

Figure 5: Bliss, Fused lasso and Flirti estimates when the true function $\beta(t)$ is interpretable (Datasets 4 and 6 generated from Shape 1). *Same legend as Figure 4.*

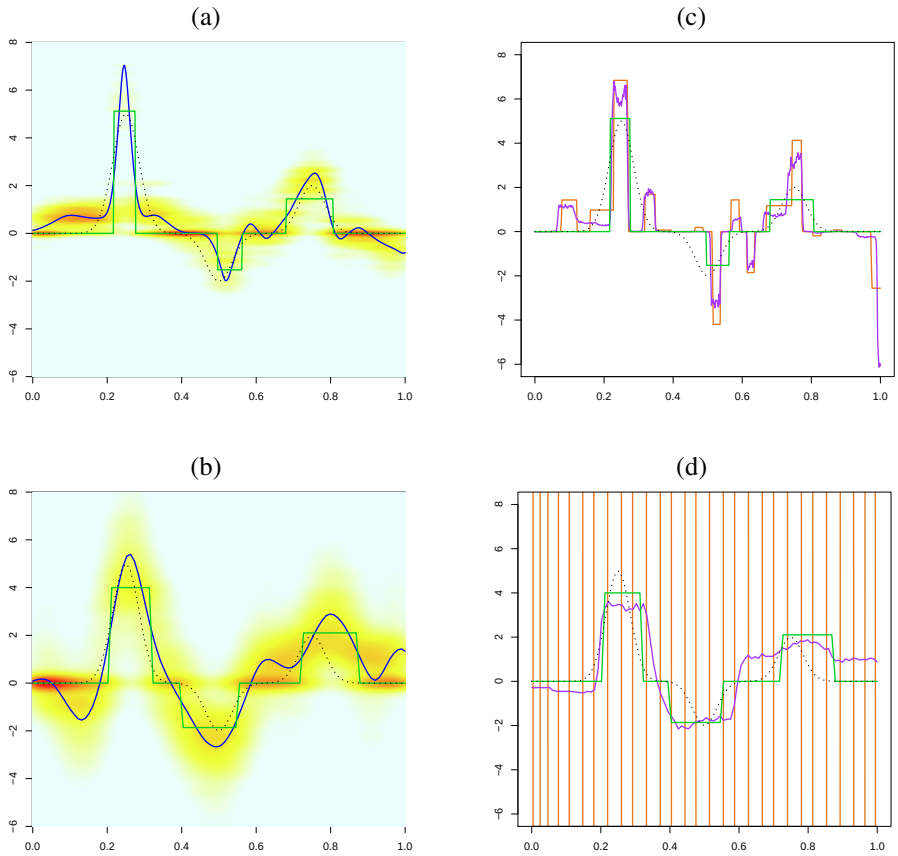

Figure 6: Bliss, Fused lasso and Flirti estimates when the true function $\beta(t)$ is smooth (Datasets 10 and 12 generated from Shape 2). *Same legend as Figure 4.*

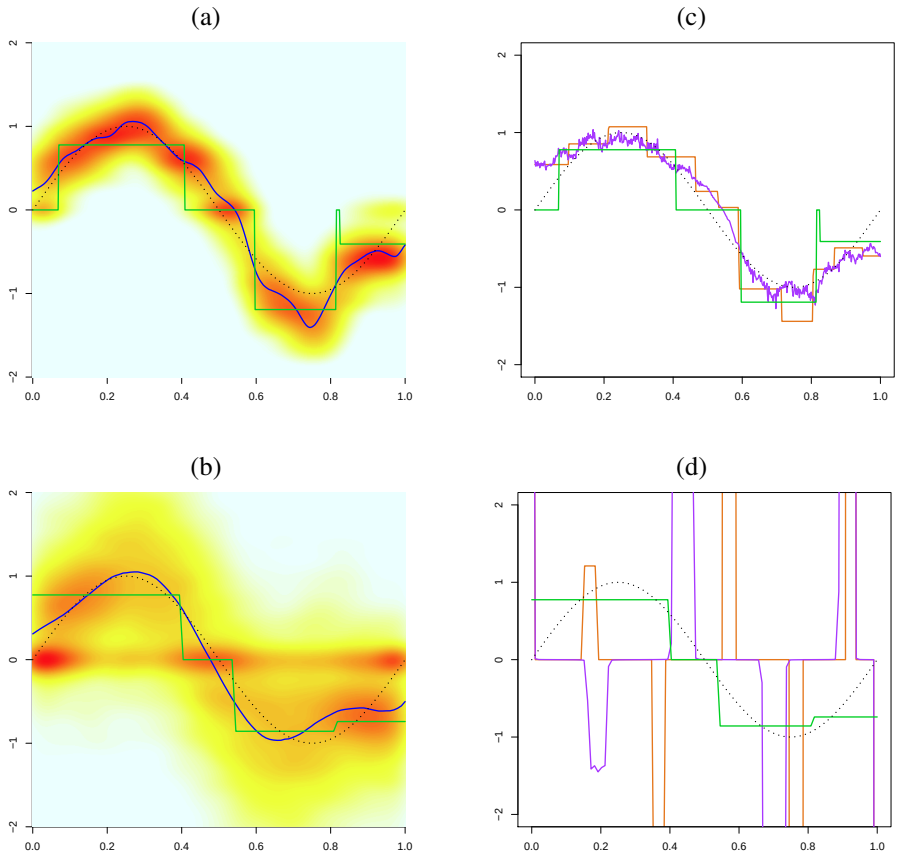

Figure 7: Bliss, Fused lasso and Flirti estimates when the true function $\beta(t)$ is sinusoidal (Datasets 17 and 21 generated from Shape 3). *Same legend as Figure 4.*

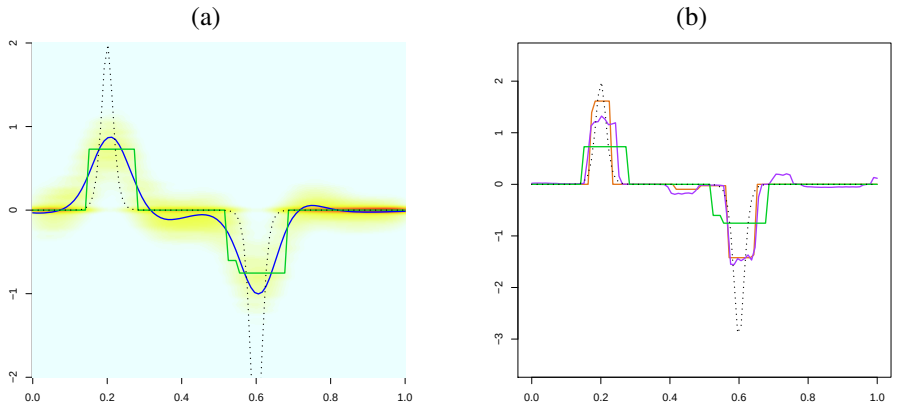

Figure 8: Bliss, Fused lasso and Flirti estimates when the true function $\beta(t)$ is smooth and nonnull on tight intervals (Dataset 22 generated from Shape 4). *Same legend as Figure 4.*

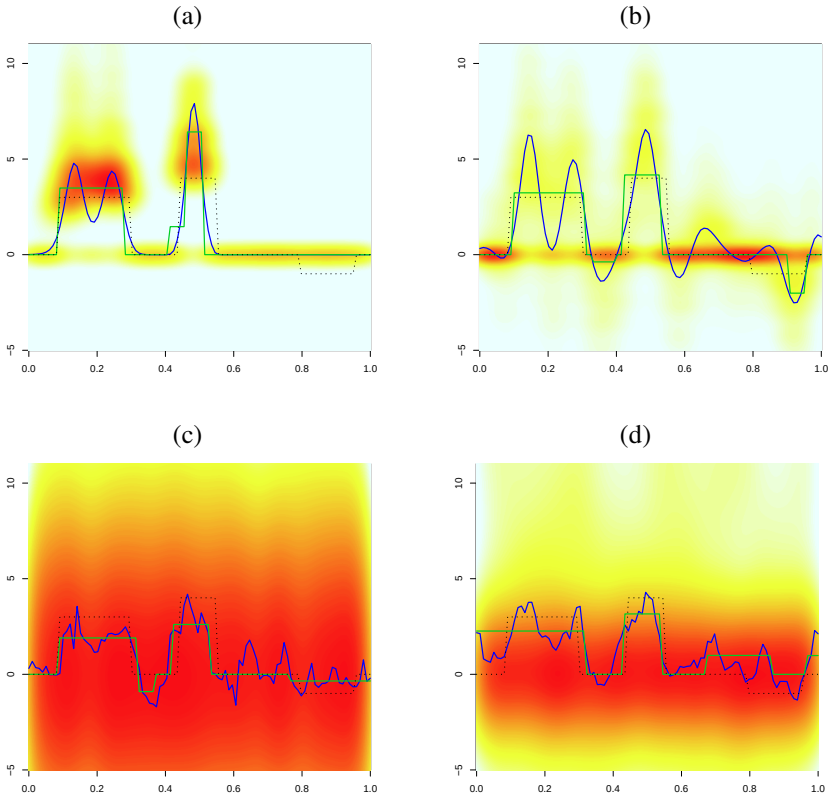

Figure 9: Graphical results of Bliss for poor choices of sensitive hyperparameters. *The heat map is a representation of an estimated posterior density of $\beta(t)$, see Section 2.3. White (resp. red) areas correspond to low (resp. high) posterior densities. Plots (a) and (b) illustrate the impact of K. Plot (c) (resp. (d)) illustrates the variability induced by a poor choice of η_0 (resp. η).*

Table 2: Performances of the Bliss estimator with respect to the hyperparameters.

	Err	CCR	FNR	FnNR	Run
$N = 2000$	1.026	0.810	0.188	0.191	2
$N = 10\,000$	1.173	0.830	0.222	0.081	3
$N = 50\,000$ ♡	1.270	0.760	0.266	0.200	1
$N = 100\,000$	1.038	0.730	0.244	0.294	4
$K = 3$	1.525	0.750	0.309	0.103	5
$K = 5$	1.427	0.770	0.301	0.037	6
$K = 10$ ♡	1.270	0.760	0.266	0.200	1
$K = 15$	1.057	0.760	0.250	0.227	7
$\eta_0 = 0$ ♡	1.270	0.760	0.266	0.200	1
$\eta_0 = 10$	1.013	0.850	0.206	0.054	8
$\eta_0 = 100$	1.084	0.680	0.314	0.326	9
$\eta_0 = 1000$	1.084	0.810	0.054	0.269	10
$v_0 = 10^1$	1.158	0.840	0.178	0.136	11
$v_0 = 10^2$ ♡	1.270	0.760	0.266	0.200	1
$v_0 = 10^3$	1.211	0.660	0.312	0.365	12
$v_0 = 10^4$	1.168	0.630	0.307	0.409	13
$\eta = 0$ ♡	1.270	0.760	0.266	0.200	1
$\eta = 10$	1.198	0.710	0.317	0.243	14
$\eta = 100$	1.514	0.620	0.294	0.424	15
$\eta = 1000$	44.864	0.390	0.615	0.606	16
$v = 10^1$	0.689	0.770	0.177	0.272	17
$v = 10^2$ ♡	1.270	0.760	0.266	0.200	1
$v = 10^3$	2.544	0.740	0.333	0.000	18
$v = 10^4$	5.661	0.720	0.333	0.142	19
a and $b = 1$	0.885	0.880	0.083	0.153	20
a and $b = 10^{-1}$ ♡	1.270	0.760	0.266	0.200	1
a and $b = 10^{-2}$	1.098	0.900	0.137	0.047	21
a and $b = 10^{-4}$	1.453	0.750	0.320	0.040	22
$\ell_{\max} = \lvert\mathcal{T}\rvert$	0.982	0.830	0.127	0.207	23
$\ell_{\max} = \lvert\mathcal{T}\rvert/2$	0.655	0.740	0.240	0.280	24
$\ell_{\max} = \lvert\mathcal{T}\rvert/5$ ♡	1.270	0.760	0.266	0.200	1
$\ell_{\max} = \lvert\mathcal{T}\rvert/8$	0.784	0.700	0.316	0.275	25

Err stands for the Estimation Error, CCR for Correct Classification Rate, FNR for False Null Rate and FnNR for False Nonnull Rate. The ♡ symbol indicates the default values.

3.5 Performance of Bliss with respect to its hyperparameters

In this section, performances of the Bliss estimator with respect to the hyperparameters are studied on Dataset 1. The following values are considered for each hyperparameter:

- η_0 and $\eta = 0, 10, 100, 1000$ (note that η is a K-vector and here each of its elements has the same value),
- v_0 and $v = 10^1, 10^2, 10^3, 10^4$,
- a and $b = 1, 10^{-1}, 10^{-2}, 10^{-4}$,
- $\ell_{\max} = |\mathcal{T}|,\ |\mathcal{T}|/2,\ |\mathcal{T}|/5,\ |\mathcal{T}|/8$ and
- $K = 3, 5, 10, 15$.

Additionally we varied the size N of the MCMC sample: $N = 2.10^3, 10^4, 5.10^4$ and 10^5. For each run of this study, only one parameter is modified and the others are chosen by default as given in Section 3.2.

Table 2 presents numerical results. Bliss is not really sensitive to these parameters, except for K, η, η_0 and v. Variations of the others parameters do not induce significant variations on the estimation error or on the classification rates.

Concerning the convergence, the results are similar for the different values of N (runs 2, 3, 1, 4). Hence, for this example 2.10^3 iterations seem enough to achieve convergence.

Concerning the influence of K (runs 5, 6, 1, 7), it appears that this hyperparameter should be chosen large enough, so that it is larger than the true K. Figures 9(a) and 9(b) show estimates for $K = 3$ or $K = 15$ while the true K is 3. For $K = 3$ (run 5), one can note that even if K is the actual K, the estimate does not find the third interval because of its small magnitude. For $K = 15$ (run 7), the estimate finds all the intervals and have only one false null interval. Indeed, intervals $I_1, \ldots, I_K$ can overlap so they gather on the relevant nonnull intervals.

Concerning the choice of the prior expectation of β^* (runs 1, 14, 15, 16), we note that a specified value of η far away from the true value $(3, 4, -1)$ induces a high variability of the posterior distribution, see Figure 9(c).

Concerning the choice of the prior expectation of μ (runs 1, 8, 9, 10), it appears that as η_0 increases, the quality criteria considered are similar. However, Figure 9(d) shows that the increase of η_0 induces higher variability of the posterior distribution of $\beta(t)$.

Finally, it appears (runs 17, 1, 18, 19) that large values for v have a negative effect on the estimation error, but only a small impact on the classification rates.

4 Application to the black Périgord truffle dataset

We apply the Bliss method on a dataset to predict the among of production of black truffles knowing the rainfall curves. The black Périgord truffle (*Tuber Melanosporum Vitt.*) is one of the most famous and valuable edible mushrooms, because of its excellent aromatic and gustatory qualities. It is the fruiting body of a hypogeous Ascomycete fungus, which grows in ectomycorrhizal symbiosis with oaks species or hazelnut trees in Mediterranean conditions. Modern truffle cultivation involves the plantation of orchards with tree seedlings inoculated with *Tuber Melanosporum*. The planted orchards could then be viewed as ecosystems that should be managed in order to favour the formation and the growth of truffles. The formation begins in late winter with the germination of haploid spores released by mature ascocarps. Tree roots are then colonised by haploid mycelium to form ectomycorrhizal symbiotic associations. Induction of the fructification (sexual reproduction) occurs in May or June (the smallest truffles have been observed in mid-June)). Then the young truffles grow during summer months and are mature between the middle of November and the middle of March (harvest season). The production of truffles should then be sensitive to climatic conditions throughout the entire year (Le Tacon et al., 2014). However, to our knowledge few studies focus on the influence of rainfall or irrigation during the entire year (Demerson and Demerson, 2014; Le Tacon et al., 2014). Our objective is then to investigate the influence of rainfall throughout the entire year on the production of black truffles. Indeed, knowing this influence could lead to a better management of the orchards, to a better understanding of the sexual reproduction, and to a better understanding of the effects of

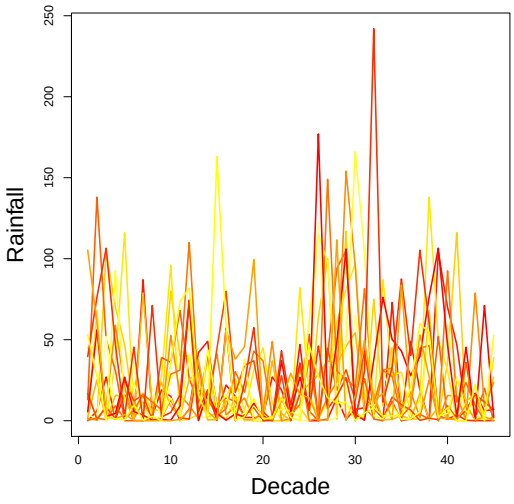

Figure 10: Covariate curves of the Truffle datsaset. *Plot shows the rainfall for each year, color-coded by their truffle yield.*

climate change. Indeed, concerning sexual reproduction Le Tacon et al. (2014, 2016) made the assumption that climatic conditions could be critical for initiation of sexual reproduction throughout the development of the mitospores expected to occur in late winter or spring. And concerning climate change, its consequences on the geographic distribution of truffles is of interest (see Splivallo et al., 2012 or Büntgen et al., 2011, among others).

The data used have been provided by J. Demerson. They consist of records of rainfall on an orchard near Uzès (France) between 1985 and 1999, and of the production of black truffles on this orchard between 1985 and 1999. In practice, to explain the production of the year n, we will take into account the rainfall between the 1st of January of the year $n-1$ and the 31st of March of the year n. Indeed, we want to take into account the whole life cycle, from the formation of new ectomycorrhizas following acospore germination during the winter preceding the harvest (year $n-1$) to the harvest of the year n. The cumulative rainfall is measured every 10 days (every decade), hence between the 1st of January of the year $n-1$ and the 31st of March of the year n we have the rainfalls associated with 45 decades, see Figure 10. This dataset can be considered as reliable, as the rainfall records have been made exactly on the orchard, and the orchard was not irrigated.

Results and biological relevance We run the Bliss method and Figure 11 shows the estimate of the coefficient function and its posterior distribution.

We observe on this estimate a positive effect of the rainfall in summer of year $n-1$ (June, July and August), and this is a well known effect. Büntgen et al. (2012), Demerson and Demerson (2014) or Le Tacon et al. (2014) all confirm the importance of the negative effect of summer hydric deficit on truffle production: they found it to be the most important factor influencing the production. Indeed, in summer the truffles need water to survive the high temperatures and to grow. Otherwise they can dry out and die.

Next, we find a positive effect of the rainfall in late winter of year $n-1$ (from the third decade of January to the end of March). This was also shown by Demerson and Demerson (2014) and Le Tacon et al. (2014). Indeed, as explained in Le Tacon et al. (2014), consistent water availability in late winter could support the formation of new mycorrhizae, thus allowing

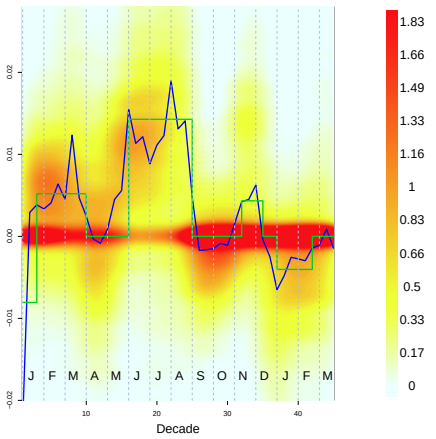

Figure 11: Bliss estimates for the truffle dataset. *The green curve is the Bliss estimate of Proposition 2, and the blue curve is the posterior expected value of Proposition 1. The heat map is a representation of the estimated posterior density, which is described in Section 2.3. White (resp. red) areas correspond to low (resp. high) posterior densities.*

a new cycle. Moreover, from results obtained by Healy et al. (2013) they made the assumption that rainfall is critical for the initiation of sexual reproduction throughout development of mitospores, which is expected to occur in late winter or spring of the year $n-1$. This is an assumption as the occurrence and the initiation of sexual reproduction is largely unknown, see Murat et al. (2013) or Le Tacon et al. (2016). Our finding of a positive effect of rainfall in late winter of year $n-1$ supports this assumption.

Moreover, we find a negative effect of the rainfall during the harvest (January and first and second decades of February of the year n). We can assume that excess water during the harvest season favor pests and truffle rot, especially for truffles close to, or at, the soil surface (see Olivier et al., 2012, page 109). It is then quite interesting to note that we find a negative effect of the rainfall in winter during the harvest, and a positive effect of rainfall in winter the year before the harvest.

We find a negative effect of the rainfall in the beginning of January of the year $n-1$ (first and second decade of January). This effect was not found by others, and we did not have a biological interpretation. We suspect some edge effect or an artefact as explained in Section 3.4.

Then, we find a positive effect of the rainfall in November and December of the year $n-1$ (second and third decades of November and the first decade of December). This was also shown by Demerson and Demerson (2014) and Le Tacon et al. (2014). Le Tacon et al. explained that rainfall in autumn allows the growth of young truffles which have survived the summer.

In contrary to Demerson and Demerson (2014), we did not find a negative effect of the rainfall in late April. Demerson and Demerson (2014) made two assumptions about this effect. The first is that excess water in the soil would reduce the oxygenation of truffle primordia. The second is that the initiation of sexual reproduction requires a hydric stress, otherwise the truffle fungus can survive using vegetative propagation of ectomycorrhizas along the roots and do not need to use sexual reproduction (hence fructification). However, as noted in Olivier et al. (2012) (page 109), the assumption of a necessary hydric stress for initiation of sexual reproduction has not been proven yet.

Furthermore, we did not find a negative effect of the rainfall in September. The assumption made was that in September the soil temperature is still high, so micro-organisms responsible for rot are quite active, while a truffle impregnated by water has its respiratory system perturbed and can not defend itself against these micro-organisms. We did not find a negative effect of rainfall for these two intervals of time, but it is interesting to look at the posterior distribution of the coefficient function in Figure 11. Indeed, on these intervals the posterior probability of a negative coefficient function is non negligible (some negative regions are in yellow and orange on these intervals). However our interpretable estimate did not detect a negative effect of rainfall on these intervals, which can be due to a lack of observations, a lack of power (only 14 years and one orchard are used in this study).

The results obtained by the Bliss method are then biologically relevant: they validate well known effects, and support recent assumptions concerning the truffle life cycle. Other assumptions are not validated by our final estimate of the coefficient function, but the estimate of the posterior distribution gives interesting insights and suggests a lack of power.

5 Conclusion and Discussion

In this paper, we provide a prior distribution and an associated loss function for estimating an interpretable function coefficient, $\beta(t)$, of the scalar-on-function linear regression model. Simulations from the posterior distribution are obtained with a Gibbs Sampler while the Bayes action is computed with a Simulated Annealing algorithm. A tailored graphical representation of the posterior distribution enables to visualize posterior credible regions for $\beta(t)$ and the intervals of the t-axis where this coefficient function is zero. Such intervals are of primary interest for the interpretability of the coefficient function. A comparison study with competitors reveals a good behaviour of our estimate: estimation errors are of the same magnitude as for competitors while our estimate remains always interpretable. Finally, an application to a real dataset on the black Périgord truffle validates some recent assumptions on the truffle life cycle and highlights the usefulness of an interpretable function coefficient in practice.

Several improvements of our method can be contemplated for future works. First, a Zellner prior on β^* (Zellner, 1986), instead of a diagonal matrix for V, would probably be suitable because of the high autocorrelations of the observed curves $x_i(t)$. It may improve the estimation of the intervals I_k and potentially reduce the artefact discussed in Section 3.4. This can be done with no additional difficulties.

Second, it would be more realistic to consider an additional term for modelling a non-interpretable part of $\beta(t)$. This suggests the following decomposition:

$$\beta(t) = \sum_{k=1}^{K} \beta_k^* \mathbb{1}\{t \in I_k\} + \xi(t).$$

If we assume a Gaussian process prior $GP(0, \Sigma)$ for the function parameter $\xi(t)$, under mild conditions on $x_i(t)$ and the covariance function $\Sigma(t, t')$, the random variable $\xi_i^* = \int \xi(t) x_i(t) \mathrm{d}t$ have a centered normal distribution with

$$\mathrm{cov}(\xi_i^*, \xi_j^*) = \iint x_i(t) x_j(t') \Sigma(t, t') \, \mathrm{d}t \mathrm{d}t'.$$

Then, it is easy to see that, given $m_i = \mu + \sum_{k=1}^{K} \beta_k^* x_i(I_k)$, σ^2 and Σ, $y = (y_1, \ldots, y_n)'$ has a multivariate normal distribution $\mathcal{N}_n(m, \Sigma_n + \sigma^2 Id_n)$ with $m = (m_1, \ldots, m_n)'$ and where

Σ_n is the $n \times n$ matrix with entries $\mathrm{cov}(\xi_i^*, \xi_j^*)$. Inference on Σ_n can be performed using a standard inverse Wishart prior or by taking into account a particular parametric form for Σ. Note that this inference can be simplified if the non-interpretable part of $\beta(t)$ is depending on i as follows:

$$\beta(t) = \sum_{k=1}^{K} \beta_k^* \mathbb{1}\{t \in I_k\} + \xi_i(t). \tag{12}$$

If we assume that ξ_i are a priori independent with a $GP(0, \Sigma)$ distribution, the random variables $\xi_i^* = \int \xi_i(t) x_i(t) \mathrm{d}t$ are independent with a $\mathcal{N}(0, \sigma_i^2)$ distribution where $\sigma_i^2 = \Sigma_n(i,i)$. The model studied in the present paper can be seen as a special case of (12) when σ_i^2 is negligible. In the same spirit, the model could be more realistic with a delay parameter τ_i for each observation:

$$y_i = \int \beta(t - \tau_i)\, x_i(t) \mathrm{d}t.$$

Third, expert opinions should be included in the prior distribution. For instance, it is now well known that an important rainfall during the summer does increase the truffle yield. It is then consistent to include such an information in the inferential process. Note that information of this kind may be vague, qualitative rather then quantitative and it may be not straightforward to construct an associated prior.

Finally, it would be of practical interest to provide some typical patterns of the whole curves $x_i(t)$ responsible for high (resp. low) yields. This entails to classify the curves in some clusters and, potentially, to consider individual coefficient functions $\beta_i(t)$, with similar values within each clusters. Several approaches can be contemplated for such a purpose and we plan to pursue this direction in future works.

Supplementary Materials

The implementation of the method is available at the following webpage:
http://www.math.univ-montp2.fr/~grollemund/Implementation/BLiSS/.

A Theoretical results

A.1 Proof of Proposition 1

Obviously, $\widehat{\beta}_{L^2}(\cdot)$ minimizes

$$\int \int_{\mathcal{T}} (\beta_\theta(t) - d(t))^2 \,\mathrm{d}t \; \pi_K(\theta|\mathcal{D}) \mathrm{d}\theta = \int_{\mathcal{T}} \int (\beta_\theta(t) - d(t))^2 \, \pi_K(\theta|\mathcal{D}) \mathrm{d}\theta \,\mathrm{d}t$$

because it does optimize $\int (\beta_\theta(t) - d(t))^2 \pi_K(\theta|\mathcal{D})\mathrm{d}\theta$ for all $t \in \mathcal{T}$. It remains to show that $\widehat{\beta}_{L^2}(\cdot) \in L^2(\mathcal{T})$. We have

$$\begin{aligned}
\|\widehat{\beta}_{L^2}(\cdot)\|^2 &= \int_{\mathcal{T}} \left(\int \beta_\theta(t)\pi_K(\theta|\mathcal{D})\mathrm{d}\theta \right)^2 \mathrm{d}t \\
&= \int_{\mathcal{T}} \iint \beta_\theta(t)\beta_{\theta'}(t)\ \pi_K(\theta|\mathcal{D})\pi_K(\theta'|\mathcal{D})\mathrm{d}\theta\mathrm{d}\theta'\mathrm{d}t \\
&= \iint \int_{\mathcal{T}} \beta_\theta(t)\beta_{\theta'}(t)\mathrm{d}t\ \pi_K(\theta|\mathcal{D})\pi_K(\theta'|\mathcal{D})\mathrm{d}\theta\mathrm{d}\theta' \\
&\leq \iint \|\beta_\theta(\cdot)\|\|\beta_{\theta'}(\cdot)\|\ \pi_K(\theta|\mathcal{D})\pi_K(\theta'|\mathcal{D})\mathrm{d}\theta\mathrm{d}\theta' \quad \text{with Cauchy-Schwarz inequality} \\
&\leq \left(\int \|\beta_\theta(\cdot)\|\pi_K(\theta|\mathcal{D})\mathrm{d}\theta \right)^2
\end{aligned}$$

And the last integral is finite because of the assumption. Hence $\widehat{\beta}_{L^2}(\cdot)$ is in $L^2(\mathcal{T})$. □

A.2 Proof of Proposition 2

First, the norm $\|d(\cdot) - \widehat{\beta}_{L^2}(\cdot)\|$ is non negative, hence the set

$$\left\{ \|d(\cdot) - \widehat{\beta}_{L^2}(\cdot)\|,\ d(\cdot) \in \mathcal{E}^{\varepsilon}_{K_0} \right\}$$

admits an infimum. Let m denote this infimum. We have to prove that m is actually a minimum of the above set, namely that there exists a function $d(\cdot) \in \mathcal{E}^{\varepsilon}_{K_0}$ such that $m = \|d(\cdot) - \widehat{\beta}_{L^2}(\cdot)\|$.

To this end, we introduce a minimizing sequence $\{d_n(\cdot)\}$ and we will show that one of its subsequence admits a limit within $\mathcal{E}^{\varepsilon}_{K_0}$. Let $d_n(\cdot)$ be such that

$$m = \inf \left\{ \|d(\cdot) - \widehat{\beta}_{L^2}(\cdot)\|,\ d(\cdot) \in \mathcal{E}^{\varepsilon}_{K_0} \right\} \leq \|d_n(\cdot) - \widehat{\beta}_{L^2}(\cdot)\| \leq m + 2^{-n}. \tag{13}$$

The step function $d_n(\cdot)$ can be written as

$$d_n(t) = \sum_{k=1}^{L} \alpha_{k,n} \mathbb{1}\{t \in (a_{k,n},\ b_{k,n})\}$$

where the $(a_{k,n},\ b_{k,n})$, $k = 1, \ldots, L$ are non overlapping intervals. Note that their number L does not depend on n because all $d_n(\cdot)$ lie in $\mathcal{E}_{K_0}$ for some fixed value of K_0, and we can always choose $L = 2K_0 - 1$. Moreover, because $d_n(t)$ is in $\mathcal{F}^{\varepsilon}$, we can assume that

$$b_{k,n} - a_{k,n} \geq \varepsilon, \quad \text{for all } k, n. \tag{14}$$

Now the sequence $\{a_{1,n}\}_n$ has its elements in the compact interval $\mathcal{T}$ hence we extract a subsequence (still denoted $\{a_{1,n}\}_n$) which converges an element $a_{1,\infty}$ of $\mathcal{T}$. Likewise, by extracting subsequences $2L$ times, we can assume that all sequences $\{a_{1,n}\}_n, \ldots, \{a_{L,n}\}_n$, $\{b_{1,n}\}_n, \ldots, \{b_{L,n}\}_n$ are convergent, and that

$$a_{k,\infty} = \lim_{n\to\infty} a_{k,n}, \quad b_{k,\infty} = \lim_{n\to\infty} b_{k,n}, \quad \text{and} \quad b_{k,\infty} - a_{k,\infty} \geq \varepsilon, \quad k = 1, \ldots, L$$

where the last inequalities come from (14).

The sequence $d_n(\cdot)$ is bounded (in L^2-norm):

$$\|d_n(\cdot)\| \leq \|\widehat{\beta}_{L^2}(\cdot)\| + \|d_n(\cdot) - \widehat{\beta}_{L^2}(\cdot)\| \leq R + \sqrt{m+1}$$

with (13), where $R = \|\widehat{\beta}_{L^2}(\cdot)\|$. Moreover

$$\|d_n(\cdot)\|^2 = \sum_{k=1}^{L} \alpha_{k,n}^2 (b_{k,n} - a_{k,n}) \geq \varepsilon \sum_{k=1}^{L} \alpha_{k,n}^2.$$

Hence, each sequence $\{\alpha_{1,n}\}_n, \ldots, \{\alpha_{L,n}\}_n$ is bounded. Thus, by further extracting subsubsequences, we can assume that, for $k = 1, \ldots, L$,

$$\lim_{n\to\infty} \alpha_{k,n} = \alpha_{k,\infty}$$

Finally, by setting

$$d_\infty(\cdot) = \sum_{k=1}^{L} \alpha_{k,\infty} \mathbb{1}\{t \in (a_{k,\infty},\ b_{k,\infty})\}$$

we can easily prove that $d_n(\cdot)$ tends to $d_\infty(\cdot)$ in L^2-norm and that $d_\infty(\cdot) \in \mathcal{E}_{K_0}^{\varepsilon}$. And, with (13)

$$m = \|d_\infty(\cdot) - \widehat{\beta}_{L^2}(\cdot)\|$$

which concludes the proof. □

A.3 Topological properties of $\mathcal{E}_K$

Proposition 3. *Let $K \geq 1$.*

(i) The convex hull of $\mathcal{E}_K$ is $\mathcal{E}$.

(ii) Under the $L^2(\mathcal{T})$-topology, the closure of $\mathcal{E}$ is $L^2(\mathcal{T})$.

Proof. The result of *(ii)* is rather classical, see, e.g., Rudin (1986). The convex hull of $\mathcal{E}_K$ contains any step function. Indeed, any step function can be written as a convex combination of simple $a\mathbb{1}\{t \in I\}$'s which all belongs to $\mathcal{E}_K$. Moreover, $\mathcal{E}$ is convex because it is a linear space. Hence claim *(i)* is proven. □

For a given K, the set of functions $\mathcal{E}_K$ is not suitable to define a projection of $\hat{\beta}_{L^2}(\cdot)$. Indeed, let $\{d_n(\cdot)\}$ be a minimizing sequence of the set $\{\|d(\cdot) - \hat{\beta}_{L^2}(\cdot)\|,\ d(\cdot) \in \mathcal{E}_K(\cdot)\}$, so

$$m = \inf\left\{\|d(\cdot) - \widehat{\beta}_{L^2}(\cdot)\|,\ d(\cdot) \in \mathcal{E}_K\right\} \leq \|d_n(\cdot) - \widehat{\beta}_{L^2}(\cdot)\| \leq m + 2^{-n}.$$

Using that $\hat{\beta}_{L^2}(\cdot)$ and $d_n(\cdot)$ belong to L^2 for all n, we have

$$d_n(.) \in \mathcal{E}_K \cap \mathcal{B}_{L^2}(R + m + 1), \qquad \text{for all } n,$$

where $\mathcal{B}_{L^2}(r)$ is the L^2-ball of radius r around the origin. Note that $\mathcal{E}_K \cap \mathcal{B}_{L^2}(R + m + 1)$ is not a compact set, for example consider $d_n(t) = \sqrt{n}\,\mathbb{1}\{t \in [0, \frac{1}{n}]\}$. Hence it is not possible to extract a subsequence of $\{d_n(\cdot)\}$ which converges to a $d_\infty(\cdot) \in \mathcal{E}_K$ such that $\|d(\cdot) - \hat{\beta}_{L^2}(\cdot)\| = m$.

B Details of the implementations

B.1 Gibbs algorithm and Full conditional distributions

The full conditional distributions for the Gibbs Sampler in Section 2.3 are the following,

$$\begin{aligned}
\beta^*|y,\mu,\sigma^2,m,\ell &\sim \mathcal{N}_{K+1}\left(\Sigma_\beta\mu_\beta,\sigma^2\Sigma_\beta\right),\\
\sigma^2|y,\mu,\beta^*,m,\ell &\sim \Gamma^{-1}\left(a+\frac{n+K+1}{2},b+\frac{1}{2}\mathrm{SSE}+\frac{1}{2}\left\|\beta^*-\eta\right\|^2_{V^{-1}}\right),\\
\Pi\left(m_k|y,\mu,\beta^*,\sigma^2,m_{-k},\ell\right) &\propto \exp\left(-\mathrm{SSE}/2\sigma^2\right)\\
\Pi\left(\ell_k|y,\mu,\beta^*,\sigma^2,m,\ell_{-k}\right) &\propto \exp\left(-\mathrm{SSE}/2\sigma^2\right)
\end{aligned}$$

where $\mathrm{SSE}=\|y-\mu\mathbb{1}-x_{\cdot}(I_{\cdot})\beta^*\|^2$, $\Sigma_\beta^{-1}=x_{\cdot}(I_{\cdot})^Tx_{\cdot}(I_{\cdot})+V^{-1}$, $\mu_\beta=x_{\cdot}(I_{\cdot})^Ty+V^{-1}\eta$ and $x_{\cdot}(I_{\cdot})$ the matrix whose (i,k)-entry is

$$x_i(I_k)=\int_{m_k-\ell_k}^{m_k+\ell_k}x_i(t)\mathrm{d}t.$$

The full conditional distributions for the hyperparameters m_k and ℓ_k are unusual distributions. As the covariate curves x_i are observed on a grid $\mathcal{T}_G=(t_j)_{j=1,\dots,p}$, we consider that m_k belongs to $\mathcal{T}_G$ and ℓ_k is such that $m_k\pm\ell_k\in\mathcal{T}_G$. Thus, the number of possible values for m_k and ℓ_k is finite and the full conditional distributions of m_k and ℓ_k are easily computable.

B.2 Simulated Annealing algorithm

We give in this section the details of the Simulated Annealing algorithm we use. Let $\tilde{\Theta}_{K_0}=\bigotimes_{K=1}^{K_0}(K,\Theta_K)$ where Θ_K is the space of all $\theta=(\beta_1^*,\dots,\beta_K^*,m_1,\dots,m_K,\ell_1,\dots,\ell_K)$ and let the function $C(d(\cdot))=\|d(\cdot)-\hat{\beta}_{L^2}(\cdot)\|^2$.

Algorithm : *Simulated Annealing*

- Initialize: a deterministic decreasing schedule of temperature $(\tau_i)_{i=1,\dots,N_{\mathrm{SANN}}}$, a value of K_0 and an initial vector $(K_{(0)},\theta_{(0)})\in\tilde{\Theta}_{K_0}$.
- Compute the function $\beta_{(0)}(t)$ from $(K_{(0)},\theta_{(0)})$.
- Repeat for i from 1 to N_{SANN} :
  - Choose randomly a move from $(K_{(i-1)},\theta_{(i-1)})$ to (K',θ') among :
    1. propose a new $\beta_k^{*\prime}$ for an arbitrary $k\leq K_{(i-1)}$,
    2. propose a new m_k' for an arbitrary $k\leq K_{(i-1)}$,
    3. propose a new ℓ_k' for an arbitrary $k\leq K_{(i-1)}$,
    4. propose to append a new interval $(\beta^{*\prime},m',\ell')$ or
    5. propose to drop out an interval (β_k^*,m_k,ℓ_k) for an arbitrary $k\leq K_{(i-1)}$.
  - Compute the function $\beta'(t)$ from the proposal (K',θ').
  - Compute the acceptance ratio

$$\alpha=\min\left\{1,\exp\left(\frac{C(\beta'(\cdot))-C(\beta_{(i)}(\cdot))}{\tau_i}\right)\right\}.$$

- Draw u from Unif$(0, 1)$.
- If $u < \alpha$, $(K_{(i)}, \theta_{(i)}) = (K', \theta')$ (move accepted),
  else $(K_{(i)}, \theta_{(i)}) = (K_{(i-1)}, \theta_{(i-1)})$ (move rejected).
- Compute the function $\beta_{(i)}(t)$ from $(K_{(i)}, \theta_{(i)})$.

- Return the iteration $(K_{(i)}, \theta_{(i)})$ minimizing the criteria $C(.)$.

For the schedule of temperature, we use by default a logarithmic schedule (see Bélisle, 1992), which is given for each iteration i by

$$\mathrm{Te}/\log\left((i-1)+e\right), \tag{15}$$

where Te is a parameter to calibrate and corresponds to the initial temperature. The result of the Simulated Annealing algorithm is sensitive to the scale of Te and it is quite difficult to find an a priori suitable value. For example, if the initial temperature is too small, almost all the proposed moves are rejected during the algorithm. On the opposite, if it is too large, they are almost all accepted. So, we run few times the algorithm and each time Te is determined with respect to the previous runs. For instance, if for a run the moves are always rejected or always accepted, the initial temperature for the next run is accordingly adjusted. Only 2 or 3 runs are sufficient to find a suitable scale of Te.

References

Bélisle, C. (1992). Convergence Theorems for a Class of Simulated Annealing Algorithms on $\mathbb{R}^d$. *Journal of Applied Probability*, 29(4).

Brown, P. J., Fearn, T., and Vannucci, M. (2001). Bayesian Wavelet Regression on Curves With Application to a Spectroscopic Calibration Problem. *Journal of the American Statistical Association*, 96(454).

Büntgen, U., Egli, S., Camarero, J., Fischer, E., Stobbe, U., Kauserud, H., Tegel, W., Sproll, L., and Stenseth, N. (2012). Drought-induced decline in Mediterranean truffle harvest. *Nature Climate Change*, 2:827–829.

Büntgen, U., Tegel, W., Egli, S., Stobbe, U., Sproll, L., and Stenseth, N. (2011). Truffles and climate change. *Frontiers in Ecology and the Environment*, 9(3):150–151.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45(1).

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 13(3).

Crainiceanu, C., Ruppert, D., and Wand, M. P. (2005). Bayesian Analysis for Penalized Spline Regression Using WinBUGS. *Journal of Statistical Software*, 14(14).

Demerson, J. and Demerson, M. (2014). *La truffe, la trufficulture, vues par les Demerson, Uzès (1989-2015)*. Les éditions de la Fenestrelle.

Goldsmith, J., Wand, M. P., and Crainiceanu, C. (2011). Functional regression via variational Bayes. *Electronic journal of statistics*, 5.

Healy, R., Smith, M., Bonito, G., Pfister, D., Ge, Z., Guevara, G., Williams, G., Stafford, K., Kumar, L., Lee, T., Hobart, C., Trappe, J., Vilgalys, R., and McLaughlin, D. (2013). High diversity and widespread occurence of mitotic spore mats in ectomycorrhizal Pezizales. *Molecular Ecology*, 22(6):1717–1732.

James, G., Wang, J., and Zhu, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics*, 37(5A).

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598).

Le Tacon, F., Marçais, B., Courvoisier, M., Murat, C., Montpied, P., and Becker, M. (2014). Climatic variations explain annual fluctuations in French Périgord black truffle wholesale markets but do not explain the decrease in black truffle production over the last 48 years. *Mycorrhiza*, 24:S115–S125.

Le Tacon, F., Rubini, A., Murat, C., Riccioni, C., Robin, C., Belfiori, B., Zeller, B., De La Varga, H., Akroume, E., Deveau, A., Martin, F., and Paolocci, F. (2016). Certainties and uncertainties about the life cycle of the Périgord black Truffle (Tuber melanosporum Vittad.). *Annals of Forest Science*, 73(1):105–117.

Murat, C., Rubini, A., Riccioni, C., De La Varga, H., Akroume, E., Belfiori, B., Guaragno, M., Le Tacon, F., Robin, C., Halkett, F., Martin, F., and Paolocci, F. (2013). Fine-scale spatial genetic structure of the black truffle (Tuber Melanosporum) investigated with neutral microsatellites and functional mting type genes. *The New Phytologist*, 199(1):176–187.

Olivier, J., Savignac, J., and Sourzat, P. (2012). *Truffe et trufficulture*. Fanlac.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer-Verlag New York.

Reiss, P., Goldsmith, J., Shang, H., and Ogden, T. R. (2016). Methods for scalar-on-function regression. *International Statistical Review*.

Robert, C. P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer-Verlag New York.

Robert, C. P. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer-Verlag New York.

Rudin, W. (1986). *Real and complex analysis*. McGraw-Hill Inc, New York, 3rd edition.

Splivallo, R., Rittersma, R., Valdez, N., Chevalier, G., Molinier, V., Wipf, D., and Karlovsky, P. (2012). Is climate change altering the geographic distribution of truffles? . *Frontiers in Ecology and the Environment*, 10(9):461–462.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B*.

Yuan, M. and Cai, T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics*, 38(6).

Zellner, A. (1986). *Bayesian inference and decision techniques - essays in honour of Bruno De Finetti*, chapter On assessing prior distributions and Bayesian regression analysis with g-prior distributions, pages 233–243. Elsevier Science Ltd, Amsterdam.

Zhao, Y., Ogden, T., and Reiss, P. (2012). Wavelet-Based LASSO in Functional Linear Regression. *Journal of Computational and Graphical Statistics*, 21(3).

Zhou, J., Wang, N.-Y., and Wang, N. (2013). Functional Linear Model with Zero-Value Coefficient Function at Sub-Regions. *Statistica Sinica*.

Zhu, H., Yao, F., and Zhang, H. (2014). Structured functional additive regression in reproducing kernel Hilbert spaces. *Journal of the Royal Statistical Society Series B*, 76(3).

Acknowledgement

We are very grateful to Jean Demerson for providing the truffle dataset and for his explanations. Pierre Pudlo carried out this work in the framework of the Labex Archimède (ANR-11-LABX-0033) and of the A*MIDEX project (ANR-11-IDEX-0001-02), funded by the "Investissements d'Avenir" French Government program managed by the French National Research Agency (ANR).