Bayesian regression with B-splines under combinations of shape constraints and smoothness properties

Christophe Abraham

Montpellier SupAgro - INRA, 2, place Pierre Viala

34060 Montpellier Cedex 2, France

Khader Khadraoui

Laval University, Department of Mathematics and statistics

Québec city G1V 0A6, Canada

Abstract

In this paper we approach the problem of shape constrained regression from a Bayesian perspective. A B-splines basis is used to model the regression function. The smoothness of the regression function is controlled by the order of the B-splines and the shape is controlled by the shape of an associated control polygon. Controlling the shape of the control polygon reduces to some inequality constraints on the spline coefficients. Our approach enables us to take into account combinations of shape constraints and to localize each shape constraint on a given interval. The performances of our method is investigated through a simulation study. An applications to a real data sets in food industry and Global Warming are provided.

Keywords: Bayesian regression; Shape constraints; B-splines; Control polygon; MCMC algorithms.

1 Introduction

Estimation of a regression function under shape and smoothness constraints is of considerable interest in many applications. Typical examples include, among others, the (monotone) dose-response curves in medicine, the (concave) utility functions of a risk averse decision maker in economics, the (increasings) growth curves of children's height through time. Another example, which motivates the present paper, concerns the reconstruction of monotone decreasing acidification curves in food industry.

The literature on regression under shape restrictions usually focuses on a single shape constraint on the whole definition domain of the independent variable. Important efforts have been made in order to incorporate some smoothness conditions in addition to the shape restriction. A review for the frequentist literature can be found in Delecroix & Thomas-Agnan (2000) or in Mammen et al. (2001). Much of the work in constrained nonparametric regression has been done in the context of splines. As smoothing splines are defined as minimizer of a penalized sum of squares, shape constraints can easily be incorporated, at least from a theoretical point of view, by minimizing over a restricted set (Turlach, 2005; Mammen & Thomas-Agnan, 1999). Ramsay (1988) and Meyer (2008) provide special basis of regression splines. The shape constraints are then incorporated by restricting the coefficients to be nonnegative.

Bayesian approaches on regression under shape restrictions are rather rare. The main shape constraint considered in most of the papers is monotonicity although some works with unimodality and concavity constraints also appear in the literature. In some papers, the inference on the regression is localized only at a finite set of points of the independent variable as, for instance, in Gunn & Dunson (2005). In some other papers, the regression function is piecewise constant as in Lavine & Mockus (1995) and Holmes & Heard (2003). Neelon & Dunson (2004) and Abraham (2012) use a piecewise linear regression function whose smoothness is controlled by specifying the covariances between the slopes. Models with smooth regression functions are given in Shively et al. (2009) where two approaches are considered. The first approach uses a characterization of smooth monotone functions given by Ramsay (1998) and the second one uses a regression spline model with the truncated power basis. Jones et al. (2009) uses a B-splines basis to ensure a horizontal upper asymptote. To our knowledge, this is the only example of a constrained regression with B-splines in the regression splines framework. Meyer et al. (2011) uses the special basis of regression splines provided in Meyer (2008) to propose a Bayesian estimation for generalized partial linear models under shape constraints.

It is worth noting that, in the literature on constrained regression, a regression splines approach requires the construction of a special splines basis for each type of constraint: Ramsay (1988) constructs the I-splines basis for monotone regression and Meyer (2008) constructs the C-splines basis for convex regression. But it is not clear what basis should be constructed for a regression function which is, say, first increasing (but not necessarily convex) and then convex (but not necessarily increasing), or for a unimodal regression function with unknown mode. Moreover, if it is known that, for instance, the function is greater than 2 on a given interval, can we provide a basis that enables to consider this constraint in addition to some others shape constraints ? In this paper, it is shown that the B-splines basis can be used for monotone or convex regression and also for many combinations of constraints including the examples above. Shape constraints are controlled by using the control polygon associated with a spline. Although the control polygon is used in Computer Aided Design, to our knowledge, it had never been used in shape restricted regression. We show that the constraint on the shape reduces to some constraints on the B-splines coefficients. Contrary to the truncated power basis, the constraints on the coefficients are very easy to derive for all usual shape constraints such as monotonicity, unimodality, convexity or concavity and for splines of any order (note that, in Shively et al., 2009, Section 3.1, quadratic splines are used, instead of spline with higher orders, for tractability reasons). Furthermore, thanks to the local support of B-splines, combinations of constraints can be considered and each constraint can be localized on a given interval of the x-axis if needed. More generally, the local support of B-splines is an interesting robustness property since a lack of data (or a lack of prior information) in a small region of the x-axis does not affect the inference on the whole domain of definition of the independent variable. Furthermore, B-splines are very easy to use and allow to control easily the degree of smoothness of the regression function. The Bayesian framework enables us to estimate the B-spline coefficients even when there is no data on the support of a basis function. This is particularly important for the application which motivates this work as some part of the data have been removed because of some errors during the recording process. The regression function is estimated by the posterior mode as it necessarily fulfills the shape constraint (note that this not necessarily the case for the posterior mean). The posterior mode is computed with a simulated annealing algorithm. Simulations from the posterior distribution are obtained by a Metropolis-Hastings within Gibbs algorithm.

The paper is organized as follows. In Section 2, we introduce the control polygon. It enables us to show that the shape of a spline can be controlled by imposing some constraints on the coefficients in the B-spline basis. The regression model is introduced in Section 3 where the posterior distribution and the regression function estimator are derived. Section 4 is devoted to the numerical computation. The performances of our method are investigated through a simulation study and an applications with real data in cheese-making process and Global Warming. Section 5 is devoted to a discussion.

2 Spline shape and the control polygon

Let us recall the definition and some properties of B-splines of order k on $[a_0, b_0] \subset \mathbb{R}$. Let $m \geq k \geq 1$ and $\mathbf{t} := (t_j)_1^{m+k}$ be a nondecreasing sequence of knots such that $t_j < t_{j+k}$ for all j and $t_k = a_0$ and $b_0 = t_{m+1}$. The B-spline functions of order 1 are given by $B_{j,1,\mathbf{t}}(x) := \mathbf{1}_{[t_j,t_{j+1}[}$ where $\mathbf{1}_A$ denotes the indicator function of a set A. From these first-order B-splines, we obtain higher-order B-splines by recurrence:

$$B_{j,k,\mathbf{t}}(x) := \omega_{j,k,\mathbf{t}}(x)\, B_{j,k-1,\mathbf{t}}(x) + \Big(1 - \omega_{j+1,k,\mathbf{t}}(x)\Big)\, B_{j+1,k-1,\mathbf{t}}(x), \tag{2.1}$$

with

$$\omega_{j,k,\mathbf{t}}(x) := (x - t_j)/(t_{j+k-1} - t_j)\mathbf{1}_{\{t_j < t_{j+k-1}\}}. \tag{2.2}$$

From now on, we fix the order $k > 1$ and the knot sequence $\mathbf{t}$ and, for simplicity of notation, we write B_j instead of $B_{j,k,\mathbf{t}}$. A spline of order k with knot sequence $\mathbf{t}$ is, by definition, a linear combination of the B-splines $B_1, \dots, B_m$, i.e. a function of the form:

$$s = \sum_{j=1}^{m} \beta_j B_j. \tag{2.3}$$

It can be deduced from the definition that B_j vanishes outside the interval $[t_j, t_{j+k}[$. It is also well known that s is a piecewise polynomial of degree $< k$

with breakpoints t_j which are $k-1-\#t_j$ times continuously differentiable at t_j where $\#t_j$ denotes the cardinal of $\{t_l : t_l = t_j\}$. A thorough presentation of splines is given in de Boor (2001).

Suppose that it is intended to control the shape of the spline s, defined by (2.3), on an interval $[a,b] \subseteq [a_0,b_0]$. For all $x \in [a_0,b_0]$, denote by t_{j_x} the smallest knot greater than x. From now on, it will be convenient to assume that $t_j < t_{j+1}$ for all j as it enables us to write that $t_{j_x-1} < x \leq t_{j_x}$ for all $x \in [a_0,b_0]$. Note that $B_{j_a-k},\dots,B_{j_b-1}$ are the B-splines whose support intersects with $[a,b]$, that is $[t_j,t_{j+k}] \cap [a,b] \neq \emptyset$ for $j \in \{j_a-k,\dots,j_b-1\}$ and $[t_j,t_{j+k}] \cap [a,b] = \emptyset$ for $j \notin \{j_a-k,\dots,j_b-1\}$. Thus, we have:

$$s(x) = \sum_{j=j_a-k}^{j_b-1} \beta_j B_j(x), \qquad \text{for all } x \in [a,b]. \tag{2.4}$$

We define the restricted control polygon on $[a,b]$ as the control polygon associated to the spline (2.4), i.e. the piecewise linear function which interpolates the vertexes $P_j := (t_j^*, \beta_j)$, $j = j_a-k,\dots,j_b-1$, with

$$t_j^* := (t_{j+1} + \cdots + t_{j+k-1})/(k-1). \tag{2.5}$$

(Recall that $k > 1$.) We note by $C_{[a,b]}$ this restricted control polygon. Note that $t_j^* < t_{j+1}^*$ since it is assumed that $t_j < t_{j+1}$ for all j.

For all integer $\kappa \geq 1$, a function f is said to be κ-increasing on $[a,b]$ if there exists $\kappa-1$ different numbers $x_1',\dots,x_{\kappa-1}'$ in $]a,b[$ such that f is increasing and non constant on $[a,x_1']$ and $[x_l',x_{l+1}']$ with $l \in \{2,4,\dots\}$ and decreasing and non constant on $[x_l',x_{l+1}']$ with $l \in \{1,3,\dots\}$. Remark that a 1-increasing function is simply an increasing function and that a 2-increasing function is an unimodal function. We define a κ-strictly increasing function by replacing *increasing* by *strictly increasing* in the above definition (note that a strictly increasing function is necessarily non constant on any interval). We define similarly a κ-decreasing and κ-strictly decreasing function by replacing *increasing* by *decreasing* in the definitions above. We say that f is κ-monotone if it κ-increasing or κ-decreasing. We define similarly a κ-convex function by replacing *increasing* and *constant* by *convex* and *linear* respectively in the definition of a κ-increasing function. The definition of κ-strictly convex, κ-concave and κ-strictly concave functions are similar. We say that f is κ-cc if it is κ-convex or κ-concave. Finally, we say that a function f is κ-greater than a constant c if there exists $\kappa - 1$ different numbers $x'_1, \ldots, x'_{\kappa-1}$ in $]a, b[$ such that $f \geq c$ and $f \neq c$ on $[a, x'_1[$ and $]x'_l, x'_{l+1}[$ with $l \in \{2, 4, \ldots\}$ and $f \leq c$ and $f \neq c$ on $]x'_l, x'_{l+1}[$ with $l \in \{1, 3, \ldots\}$. The definitions for κ-lower, κ-strictly greater and κ-strictly lower than c are similar. We say that f is κ-gl than c is it κ-greater than c or κ-lower than c. The proof of Proposition 1 below will be given in the Appendix.

Proposition 1 *Assume that $k > 2$.*
a) If $C_{[a,b]}$ is greater than c then s is greater than c on $[a, b]$. The result still holds if greater than c *is replaced by* strictly greater than c, lower than c, strictly lower than c, increasing, strictly increasing, decreasing, strictly decreasing, convex, strictly convex, concave *and* strictly concave.
b) If $C_{[a,b]}$ is κ-gl than c then s is κ'-gl than c with $1 \leq \kappa' \leq \kappa$. The results still holds if gl *is replaced by* monotone *or by* cc.
c) Denote by $|\mathbf{t}| = \max_j |t_j - t_{j+1}|$ and assume that s is continuously differentiable. For all $|\mathbf{t}|$ small enough, if $C_{[a,b]}$ is κ-greater than c then s is κ-greater than c. The result still holds if greater than c *is replaced by* lower than c, increasing, decreasing, convex *or* concave.

The condition $k > 2$ of Proposition 1 is needed only for convexity constraints; for others types of constraints, Proposition 1 is still valid with $k > 1$. Note that if $C_{[a,b]}$ is unimodal, than s is unimodal or monotone on $[a, b]$ but it is possible to force s to be unimodal by increasing the number of knots.

It is worth noting that the shape-preserving properties of Proposition 1 are localized on an arbitrary interval. As a consequence, Proposition 1 enables us to take into account combinations of shape constraints. For example, it is possible to impose a monotonicity constraint on $[2, 4]$ and and convexity constraint on $[3, 5]$ and no constraint elsewhere. Clearly, controlling the shape of the control polygon reduces to controlling the shape of the sequence of control points P_j. Thus, for a combination of constraints, it is straightforward to construct a set $S \subset \mathbb{R}^m$ such that s necessarily fulfills the shape constraints for all $\beta \in S$. For example, for a concavity constraint on $[a, b]$, the broken line with vertexes P_{l-1}, P_l, P_{l+1} must be concave for $l \in \{j_a - k + 1, \ldots, j_b - 2\}$ and we obtain:

$$S = \cap_{l=j_a-k+1}^{j_b-2} \{\beta \in \mathbb{R}^m, (\beta_l - \beta_{l-1})(t^*_{l+1} - t^*_l) \geq (\beta_{l+1} - \beta_l)(t^*_l - t^*_{l-1})\}.$$

Similarly, for an increasing constraint to $[a,b]$, we have:

$$S = \{\beta \in \mathbb{R}^m, \beta_{j_a-k} \leq \cdots \leq \beta_{j_b-1}\}.$$

If, furthermore, we need to impose, for instance, that s is greater or equal to c on $[a_0,b_0]$, then, it is sufficient to add that $\beta_j \geq c$ for all $j \in \{1,\ldots,m\}$.

Denote by $\mathcal{S}$ the set of functions on $[a_0,b_0]$ for which a given shape constraint is fulfilled on $[a,b]$. It is of practical interest to know whether any given function $f \in \mathcal{S}$ can be approximated by a spline associated with a control polygon in $\mathcal{S}$. From de Boor (2001, page 142), if $f \in \mathcal{S}$ is twice continuously differentiable, there exists a constant $\mathrm{const_{f,k}}$ (that may depend on k and f) such that:

$$\sup_{t\in[a_0,b_0]} |f(x) - Vf(x)| \leq \mathrm{const_{f,k}}\ |\mathbf{t}|^2, \tag{2.6}$$

where Vf is the Schoenberg's spline approximation defined by

$$Vf = \sum_{j=j_a-k}^{j_b-1} f(t_j^*)B_j.$$

By construction, the control polygon associated with Vf is the broken line with vertexes $\big(t_j^*, f(t_j^*)\big)$. Then, since $f \in \mathcal{S}$, Vf necessarily belongs to $\mathcal{S}$ if $|\mathbf{t}|$ is small enough. In other words, any twice continuously differentiable function with a given shape constraint can be arbitrarily approximated by a spline associated with a control polygon with the same shape constraint. Furthermore, by (2.6), the spline also fulfills the shape constraint if $|\mathbf{t}|$ is small enough.

3 Bayesian regression under shape constraints

Consider the usual regression model with n independent observations $(x_i,y_i) \in [a_0,b_0] \times \mathbb{R}$, $i=1,\ldots,n$:

$$y_i|x_i, f, \sigma^2 \sim N(f(x_i), \sigma^2). \tag{3.1}$$

As in the previous section, denote by $\mathcal{S}$ the set of functions with a given constraint on $[a,b] \subseteq [a_0,b_0]$ and assume it is known that f is smooth and belongs to $\mathcal{S}$. From this prior information, we construct a prior distribution on (f, σ^2) as follows. Fix an order k and a knot sequence $\mathbf{t}$ and set:

$$f = \sum_{j=1}^{m} \beta_j B_j.$$

Clearly, (3.1) reduces to $y|\beta, \sigma^2 \sim N_n(B\beta, \sigma^2 I_n)$, where $y = (y_1, \dots, y_n)'$, $x = (x_1, \dots, x_n)'$, $\beta = (\beta_1, \dots, \beta_m)'$, B is the $n \times m$ matrix whose entry (i, j) is $B_j(x_i)$ and I_n is the $n \times n$ identity matrix. From the shape preserving property, it is easy to derive a set S such that $f \in \mathcal{S}$ for all $\beta \in S$. Then, set

$$(\beta, \sigma^2) \sim NIG^S(\mu, V, \xi, \varsigma),$$

where $NIG^S(\mu, V, \xi, \varsigma)$ is the normal inverse gamma distribution conditioned on $(\beta \in S)$. Clearly, the density of $NIG^S(\mu, V, \xi, \varsigma)$ is simply proportional to the density of the unconditioned normal inverse gamma distribution $NIG(\mu, V, \xi, \varsigma)$ multiplied by the indicator function $\mathbf{1}_S(\beta)$. It is straightforward to check that the posterior distribution is a truncated $NIG^S(\mu^*, V^*, \xi^*, \varsigma^*)$ distribution where the parameters are obtained from the unconditioned case:

$$\begin{array}{rcl} \mu^* & = & (V^{-1} + \mathrm{B}'\mathrm{B})^{-1}(V^{-1}\mu + \mathrm{B}'y), \\ V^* & = & (V^{-1} + \mathrm{B}'\mathrm{B})^{-1}, \\ \xi^* & = & \xi + n/2, \\ \varsigma^* & = & \varsigma + \frac{\mu' V^{-1}\mu + y'y - (\mu^*)'(V^*)^{-1}\mu^*}{2}. \end{array} \tag{3.2}$$

Sampling from the posterior distribution can be done with a Gibbs sampler. Note that the full conditional posterior distributions are given by:

$$\left\{ \begin{array}{rcl} \beta|\sigma^2, y & \sim & N_m^S(\mu^*, \sigma^2 V^*), \\ \sigma^2|\beta, y & \sim & IG\left(\xi^* + m/2, (\beta - \mu^*)'(V^*)^{-1}(\beta - \mu^*)/2 + \varsigma^*\right), \end{array} \right. \tag{3.3}$$

where N_m^S denotes the multivariate normal distribution truncated to S. Geweke (1991) or Jun-wu & liang Tian (2011) provide numerical methods for sampling from a truncated multivariate normal.

It is worth noting that our method can generalize to more complicated models such as hierarchical linear models or linear model with non normal errors. This is easily seen since the posterior distribution for the constrained model is simply the posterior distribution for the unconstrained model conditioned on

$(\beta \in S)$. Thus, if there exists an algorithm for simulating from the posterior distribution in the unconstrained case, it should be possible, at least theoretically, to sample from the posterior distribution in the constrained case by including some Metropolis-Hastings steps. This is illustrated by Algorithm 1 below where Metropolis-Hastings steps are included to simulate from the truncated multivariate normal distribution instead of using the methods cited above. For all $\beta = (\beta_1, \ldots, \beta_m)' \in S$, set

$$S_{\beta,l} = \{z \in \mathbb{R} : (\beta_1, \ldots, \beta_{l-1}, z, \beta_{l+1}, \ldots, \beta_m) \in S\} \cap [\beta_l - \eta, \beta_l + \eta],$$

where $\eta > 0$ allows to control the variance of the proposal distribution. Denote by $p^*(\beta, \sigma^2)$ the posterior density at (β, σ^2) and by $P_{2,l}$ the transition kernel described below.

Transition kernel $P_{2,l}$

`Given` (β, σ^2)`, generate` $(\beta', \sigma^{2\prime})$ `as follows:`

T1) Set $\sigma^{2\prime} = \sigma^2$

T2) `Generate` z `from the uniform distribution on` $S_{\beta,l}$ `and set` $\widetilde{\beta} = (\beta_1, \ldots, \beta_{l-1}, z, \beta_{l+1}, \ldots, \beta_m)$`. Then, take`

$$\beta' = \begin{cases} \widetilde{\beta} & \texttt{with probability} \quad \alpha, \\ \beta & \texttt{with probability} \quad 1-\alpha, \end{cases}$$

`where` $\alpha = \min\{1, p^*(\widetilde{\beta}, \sigma^2)/p^*(\beta, \sigma^2)\}$.

Note that

$$\alpha = \min\left\{1, \exp\left[-\frac{1}{2\sigma^2}\left(Q(\widetilde{\beta}) - Q(\beta)\right)\right]\right\},$$

where

$$Q(\beta) = (\beta - \mu^*)'(V^*)^{-1}(\beta - \ \mu^*).$$

Since $\widetilde{\beta}$ is simply a local perturbation of β: $\widetilde{\beta}_{l'} = \beta_{l'}^{\nu}$ for all $l' \neq l$, it is easy to see that:

$$\begin{aligned} Q(\widetilde{\beta}) - Q(\beta) &= (\widetilde{\beta} - \beta)'(V^*)^{-1}(\widetilde{\beta} + \beta - 2\mu^*) \\ &= (\widetilde{\beta}_l - \beta_l) \ \ [(V^*)^{-1}]_{l.}(\widetilde{\beta} + \beta - 2\mu^*), \end{aligned} \tag{3.4}$$

where $[(V^*)^{-1}]_{l.}$ denotes the lth row of $(V^*)^{-1}$; hence a fast numerical computation of α even for large m. Denote by $\mathcal{P}$ the set of all permutations of $\{1, \dots, m\}$.

Algorithm 1
`Given` $(\beta^\nu, (\sigma^2)^\nu)$`,`

1. `Generate` $(\sigma^2)^{\nu+1}$ `from its full conditional distribution` (3.3).
2. a) `Generate` τ `from the uniform distribution on` $\mathcal{P}$`.`
   b) `Generate` $\beta^{\nu+1}$ `from` $P_{2,\tau(1)} \dots P_{2,\tau(m)}$`.`

It is checked in the Appendix that Algorithm 1 gives an aperiodic and irreducible chain with the posterior distribution as its invariant distribution for classical constraints; hence the theoretical convergence of the algorithm (Tierney, 1994, Theorem 1).

Although any posterior simulation of f necessarily fulfills the shape constraint, the posterior expectation of f may not belong to $\mathcal{S}$ (for instance, the mean of unimodal functions is not necessarily an unimodal function). For this reason, we propose to estimate f by the posterior mode instead of the posterior expectation. Thus, we set

$$\hat{f} = \sum_{j=1}^{m} \hat{\beta}_j B_j,$$

where $\hat{\beta}$ is a minimizer of Q under the constraint $\hat{\beta} \in S$. The estimate $\hat{\beta}$ is computed thanks to a Simulated Annealing Algorithm: given β^ν, a coordinate l is chosen at random uniformly on $\{1, \dots, m\}$ and then, $\beta^{\nu+1}$ is simulated according to $P_{2,l}$ where α is replaced by:

$$\alpha = \min\left\{1, \exp\left[-\left(Q(\beta) - Q(\beta^\nu)\right)/T_\nu\right]\right\},$$

(note that simulating σ^2 is no longer necessary). A temperature parameter $T_\nu = C/\log(\nu + e)$, with $C > 0$ not too small, guarantees the convergence of the algorithm to a global minimizer. The convergence can be proved by checking the sufficient conditions given by Bartoli & Del Moral (2001). (The proof of convergence is very similar to the one given by Abraham, 2009, Section 5.) Remark that the computation of $\hat{\beta}$ is a quadratic programming problem and, hence, can also be solved by non stochastic methods (see Boyd & Vandenberghe, 2004, Chapter 11).

In practice, it can be of interest to test $H_0 : f \in \mathcal{S}_0$ versus $H_1 : f \in \mathcal{S}_1$ for two particular constraints $\mathcal{S}_0$ and $\mathcal{S}_1$. Consider a normal inverse gamma prior as above and denote by π_0 and π_1 the prior distributions conditioned on $\mathcal{S}_0$ and $\mathcal{S}_1$ respectively. In our method, the marginal density of y under H_0 or H_1 is usually intractable but the Bayes factor can numerically be approximated from posterior simulations of the parameters. If $(\beta^{\nu}_{(j)}, (\sigma^2_{(j)})^{\nu})$ is simulated from the posterior under H_0 ($j = 0$) and H_1 ($j = 1$) by Algorithm 1, the Bayes factor can be approximated using the harmonic mean of the likelihood values (Newton & Raftery, 1994; Gelfand & Dey, 1994) by:

$$B_{01} \approx \frac{T_0 \sum_{\nu=1}^{T_1} p(y|\beta^{\nu}_{(1)}, (\sigma^2_{(1)})^{\nu})^{-1}}{T_1 \sum_{\nu=1}^{T_0} p(y|\beta^{\nu}_{(0)}, (\sigma^2_{(0)})^{\nu})^{-1}}, \tag{3.5}$$

where $p(y|\beta, \sigma^2)$ denotes the conditional density of y given (β, σ^2) and where T_j, $j = \{0, 1\}$, is the number of iterations after burn-in.

4 Numerical applications

To provide some validation for our methodology, we first apply it to artificial data (Section 4.1) to illustrate that the approach described in Section 2 enables us to take into account combinations of shape constraints and to localize a shape constraint on a given interval. Next, in Sections 4.2 and 4.3, an applications to the reconstruction of an acidification curve and to Global Warming data are provided.

4.1 A simulation study

We generate data according to model (3.1) with a true regression function defined by $f(x) = \exp\{(x-8)^2/9\}$ on $[a_0, b_0] = [-2, 12]$ with $\sigma = 0.2$ and $n = 26$. The value of σ has been chosen sufficiently large to highlight the fact that large σ can give large errors for the unconstrained estimate while the constrained estimate remains good. We assume that it is known that f is increasing on $[0, 8]$, concave on $[7, 10]$ and continuously differentiable everywhere. Then, we take $k = 3$ and an arbitrary knot sequence

$$\mathbf{t} = (-2.02, -2.01, -2.00, -1.00, 0.50, 2.00, 3.50, 5.00, 6.50, 8.00, 9.50, 11.00, 12.00, 12.01, 12.02).$$

Thus, $m = 12$ and

$$\mathbf{t}^* = (-2.005, -1.500, -0.250, 1.250, 2.750, 4.250, 5.750, 7.250, 8.750, 10.250, 11.500, 12.005),$$

where $\mathbf{t}^*$ denotes the sequence $(t_j^*)_{j=1}^m$. The monotonicity constraint involves the B-spline whose support intersects $[0, 8]$, that is B_j for $j = 2, \ldots, 9$ while the concavity constraint involves the B-spline B_j for $j = 7, \ldots, 11$. From Section 2, $f = \sum_{j=1}^{12} \beta_j B_j$ satisfied the shape and smoothness constraints when $\beta \in S$ with

$$\begin{aligned} S &= S^{monotone} \cap S^{concave}, \\ S^{monotone} &= \{\beta \in \mathbb{R}^{12}, \beta_2 \leq \ldots \leq \beta_9\}, \\ S^{concave} &= \cap_{l=8}^{10} \{\beta \in \mathbb{R}^{12}, (\beta_l - \beta_{l-1})(t_{l+1}^* - t_{l-1}^*) \geq (\beta_{l+1} - \beta_{l-1})(t_l^* - t_{l-1}^*)\}. \end{aligned}$$

We use a rather non informative prior by taking $m = 0, V = 100 I_{12}, \xi = 0.01, \varsigma = 0.01$. The unconstrained estimate $\sum_{j=1}^{12} \mu_j^* B_j$ is plotted in Figure 1 (plain line) with its control polygon (dashed line) and the true regression function f (dotted line). The control points are indicated by filled circles. The vertical lines indicate the locations of the shape constraints. Clearly, the unconstrained estimate is neither increasing on $[0, 8]$ nor concave on $[7, 10]$. Simulations from the posterior distribution is obtained by running for 10^5 iterations the Metropolis-Hastings within Gibbs sampler. Then, we compute a 0.05-credible set for $f(x)$ for every x in a fine mesh of $[-2, 12]$. The computed 0.05-credible set for σ^2 is $[0.0330, 0.0481]$ and the posterior mean is 0.0417 (recall that the data were generated with $\sigma^2 = 0.04$). Several initial values for (β, σ^2) are used and give similar results.

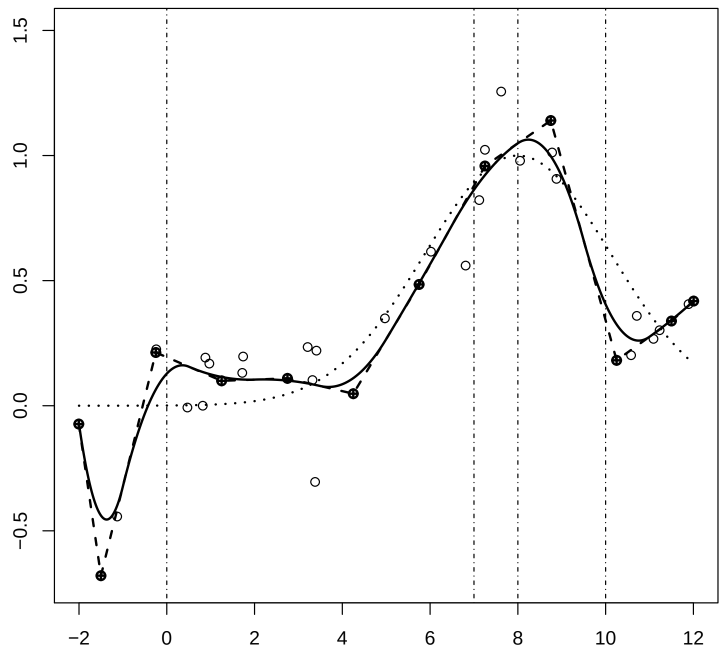

Figure 1: Data (circle), unconstrained estimate (plain line), control polygon (dashed line), control points (filled circle) and true regression function (dotted line)

Convergence diagnostics are checked with the R package `coda` (Plummer et al., 2006). The number of iterations required to estimate the quantiles 0.025 and 0.0975 to within an accuracy of ± 0.01 with probability 0.95, as recommended by Raftery & Lewis (1996), is always less than 2.4×10^4 for each components of (β, σ^2). Similarly, the recommended burn-in period is always less than 100. It can be noted, from the estimated autocorrelation functions (not given here) that, for all parameter, the autocorrelations become negligible for any lag greater than 100 (for some parameters, a lag about 30 is sufficient).

The simulated annealing algorithm is run with several values of β^0 and C and stopped after 10^4 iterations. For different β^0 which gives a spline $f =$

$\sum_{j=1}^{12} \beta_j^0 B_j$ in the neighborhood of the data, all the chains eventually merge and provide almost the same minimizers. We retain the value $C = 2$ as it gives a good balance between exploration and convergence. Then, an accurate estimate $\hat{\beta}$ is computed by running again the algorithm with this parameter and stopping it after 3×10^5 iterations. The constrained estimate is plotted on Figure 2 with a 0.05-credible set.

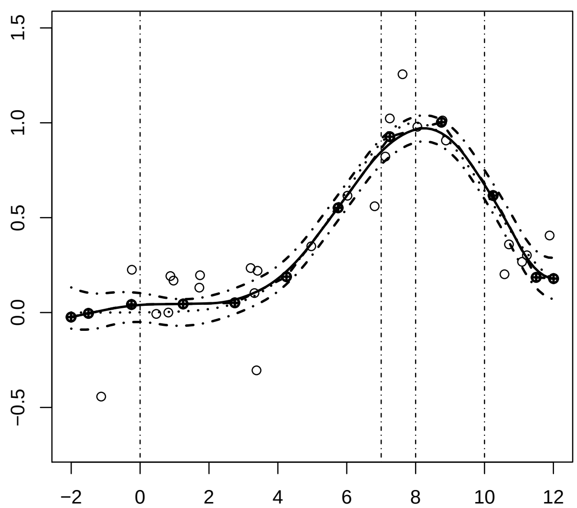

Figure 2: Data (circle), constrained estimate (plain line), true regression function (dotted line) and 95% credible interval for f (- · -).

4.2 Application to the reconstruction of a curve

The control of acidification is an important issue in cheese-making. A recent study was conducted to assess the influence of several environmental conditions (initial dissolved oxygen levels and redox potential) on the decrease of pH (acidification) in milk (Jeanson et al., 2009). The acidification kinetics consists of 1428 measures of pH. Independently of the data, it is known that the pH must be a decreasing function of the time. Some errors have occurred during the recording process because of the sensitivity of the measuring device to external environment. Then, the recorded acidification kinetic had an undesirable unimodal shape for x around 12. As it was not possible to redo the experiment and obtain new data, it is of interest to estimate what would be the true acidification curve if no measurement error had occurred. Extra information from specialists in milk acidification process can be summarized as follows: any point (x_i, y_i) with $x_i \in [8.93, 20.8]$ must be removed and any point (x_i, y_i) with $x_i \notin [7, 21]$ must be retained. Data are illustrated in Figure 3. We compute two constrained estimates of the acidification curve associated with the two data sets $\mathcal{D}_1 = \{(x_i, y_i), x_i \notin [8.93, 20.8]\}$ and $\mathcal{D}_2 = \{(x_i, y_i), x_i \notin [7, 21]\}$. Note that it is also possible to weight each point (x_i, y_i) with weight 0 if $x_i \in [8.93, 20.8]$, 1 if $x_i \notin [7, 21]$, and a weight between 0 and 1 (for instance $\min\{|x_i - 7|/|8.93 - 7|, |x_i - 21|/|20.8 - 21|\}$) for $x_i \in [7, 21] \setminus [8.93, 20.8]$. If we denote by W the diagonal matrix of weights, it is easy to see that the posterior distribution for the weighted data can be obtained by replacing respectively y and B by $W^{1/2}y$ and $W^{1/2}$B in (3.2).

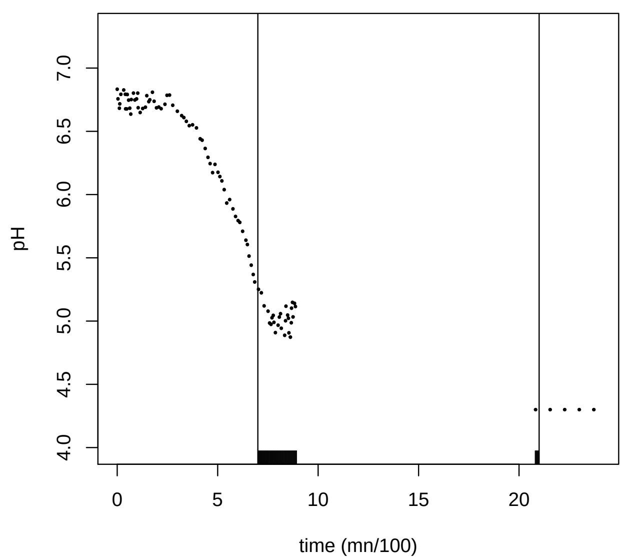

Figure 3: pH v.s. time (mn) rescaled by a factor of 10^{-2}, data in $\mathcal{D}_2 \setminus \mathcal{D}_1$ are indicated by the vertical lines and the horizontal stripes

Then, we take $k = 3$ and an arbitrary knot sequence

$$\mathbf{t} = (-0.50, -0.25, 0.00, 1.00, 3.00, 5.00, 7.00, 9.00, 11.00, 13.00, 15.00, 17.00, 19.00, 21.00, 23.00, 23.80, 24.05, 24.30).$$

The monotonicity constraint involves all the B-spline as the constraint is located on the whole interval $[0, 23.8]$. The shape and smoothness constraints are fulfilled as soon as $\beta \in S$ with

$$S = \{\beta \in \mathbb{R}^{15}, \beta_1 \geq \cdots \geq \beta_{15}\}.$$

We use a non informative prior by taking a prior density for (β, σ^2) proportional to $\sigma^{-2}\mathbf{1}_{\{\beta \in S\}}$. It is easily seen that the posterior density is proportional to the posterior density for the unconstrained prior multiplied $\mathbf{1}_{\{\beta \in S\}}$.

The unconstrained estimates for $\mathcal{D}_1$ and $\mathcal{D}_2$ (not given here) have an undesirable shapes: they are not decreasing in the neighborhood of zero and, most importantly, they are equal to zero for all x in the middle of $[8.93, 20.8]$ as there is no data in the support of some B-splines.

Computations of $\hat{\beta}$ are very similar in both cases ($\mathcal{D}_1$ and $\mathcal{D}_2$): we take $C = 2$ as this value gives a good balance between exploration and convergence and stop the algorithm after 3×10^5 iterations. Several values of β^0 are used and provide the same results. The small difference between the two estimates for $\mathcal{D}_1$ and $\mathcal{D}_2$ (see Figure 4) indicates that the reconstruction process is not very sensitive to the extra information uncertainty. We do not plot the unconstrained estimate for the clarity of the figure but we note that constrained and unconstrained estimates are identical on intervals of the x-axis where the constraints are fulfilled by the unconstrained estimate. Furthermore, we also note that the initial function of the simulated annealing algorithm (not given here for the clarity of Figure 4) is very different from the constrained regression function. These two facts indicate that the simulated annealing algorithm has probably converged to the global minimizer of Q.

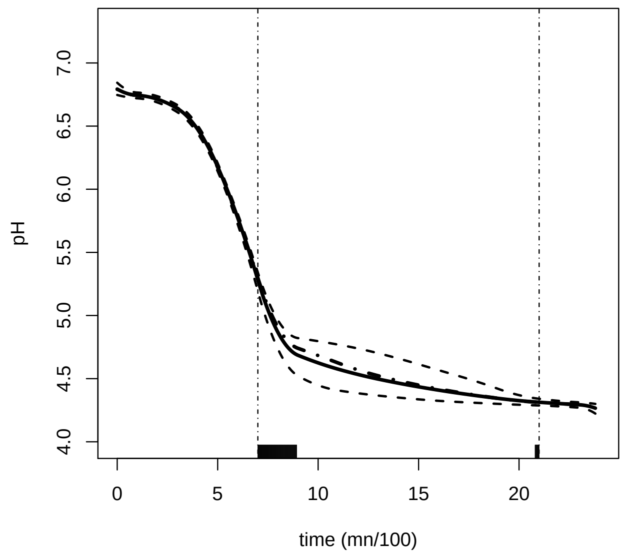

Figure 4: Constrained estimates for $\mathcal{D}_1$ (plain) and $\mathcal{D}_2$ (-.-) and excess credible interval (- -)

For both data sets $\mathcal{D}_1$ and $\mathcal{D}_2$, we run the Metropolis-Hasting algorithm for 4×10^5 iterations and check the convergence of the chain for different initial values in the same way as in Section 4.1. We also compute a posterior credible interval $[f_i^l(x), f_i^u(x)]$ of $f(x)$ for every x in a fine mesh of $[0, 23.8]$ for each data set $\mathcal{D}_i$, $i \in \{1, 2\}$. We plot in Figure (4) an excess credible interval defined by $[\min_i f_i^l(x), \max_i f_i^u(x)]$. The narrowness of the excess credible interval for times outside $[7, 21]$ is a consequence of the large number of observations with $x_i \notin [7, 21]$.

4.3 Application to the Global Warming data

The Global Warming data set (Jones et al., 2011) provides the annual temperature anomalies from 1850 to 2012, expressed in degrees Celsius. Anomalies are departures from the temperatures average between 1961 and 1990. This data set has been used by several authors: Alvarez & Dey (2009) use a Bayesian monotonic change point method to show the existence of an increasing trend in the global annual temperature between 1958 and 2000 and similar conclusions are obtained by Alvarez & Yohai (2012), Wu et al. (2001) and Zhao & Woodroofe (2012) using isotonic estimation methods. In all these papers, observations are supposed to be independent and identically distributed; for simplicity, we make the same assumption in the present paper. The methodology of the present paper enables us to test whether the temperature anomalies sequence is 5-increasing (H_1) or simply 1-increasing (H_0). Note that a 5-increasing shape has never been considered in the papers cited above. We use B-splines of order $k = 3$ with 15 coefficients and the prior distribution of Section 3 (the values of the parameters are given in the Appendix). The computation of the Bayes factor according to (3.5) gives $B_{01} = 5.9048 \times 10^{-11}$; hence a strong evidence in favor of a 5-increasing shape. For the computation of B_{01}, we run the algorithm 1 for 1.5×10^5 iterations and check the convergence in the same way as in Sections 4.2 and 4.1. Estimates of f under H_0 and H_1 are computed by the simulated annealing algorithm with 1.5×10^5 iterations and are plotted in Figure 5. Under H_1, in order to study the global trend of the increasing parts of f and the global trend of the decreasing parts of f, we define $\bar{\beta}_j$, $j \in \{1, \ldots, 5\}$, as the mean of the coefficients β_l associated with the jth monotone part of f. For example, if $\beta_1 \leq \ldots \leq \beta_{n_1} \geq \beta_{n_1+1} \geq \ldots \geq \beta_{n_1+n_2} \leq \beta_{n_1+n_2+1} \leq \ldots$, $\bar{\beta}_1$ is the mean of $\beta_1, \ldots, \beta_{n_1}$, $\bar{\beta}_2$ is the mean of $\beta_{n_1}, \ldots, \beta_{n_1+n_2}$ and so on. Denote by p^* and p_* the posterior probabilities of $\{\bar{\beta}_1 \leq \bar{\beta}_3 \leq \bar{\beta}_5\}$ and $\{\bar{\beta}_2 \leq \bar{\beta}_4\}$ respectively. Under π_1, we obtain $p^* = 1$ and $p_* = 1$ from the posterior simulations of β according to Algorithm 1. Thus, we can conclude that the data have a 5-increasing shape with an increasing global trend.

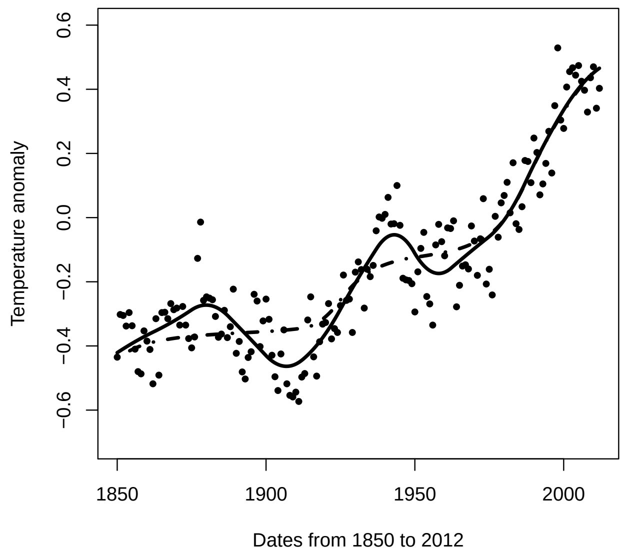

Figure 5: Annual temperatures v.s. time, 5-increasing constrained estimate (plain line) and increasing constrained estimate (dashed line).

5 Discussion

In this paper, a Bayesian method for regression under shape and smoothness constraints with B-splines is provided. It enables us to take into account combinations of shape constraints and to localize each constraint on a given interval. Contrary to Abraham (2012), it is not necessary to use a large dimensioned coefficient vector β to ensure the smoothness condition on the regression function f. As a consequence, autocorrelations of the components of β in the Metropolis-Hastings-within-Gibbs algorithm are reduced and the numerical implementation is easy to perform.

The problem of curves reconstruction as described in Section 4.2 is rather general and suggests some improvements of the method described in Section 4.2. For instance, it could be interesting to consider free knots B-splines and, especially, to detect automatically the interval of erroneous measurements. We plan to pursue these directions in future works.

6 Appendix

6.1 Proof of Proposition 1

For any sequence $\alpha = (\alpha_i)_{i=1}^n$, we denote by $S^-(\alpha)$ the number of signs changes in α, that is the largest integer r with the property that for some $1 \leq j_1 < \cdots < j_{r+1} \leq n$, $\alpha_{j_i}\alpha_{j_{i+1}} < 0$ for $i = 1, \dots, r$. For a function f, we denote by $S^-(f)$ the number of sign changes of f, that is the supremum over all numbers $S^-(f(\tau_1), \dots, f(\tau_r))$ with r arbitrary and $\tau_1 < \cdots < \tau_r$ arbitrary in the domain of f. For simplicity of notation and with no loss of generality, we assume that $[a, b] = [a_0, b_0]$ and we simply denote by C instead of $C_{[a,b]}$ the control polygon associated with the spline $s = \sum_{j=1}^m \beta_j B_{j,k}$. We also highlight the order k of the B-spline by noting $B_{j,k}$ instead of B_j. Following de Boor (2001, page 71), for all $x \notin \{t_1, \dots, t_{m+k}\}$, denote by $D^j s(x)$ the jth derivative of s at x ($D^j s(x)$ does exist since s is a polynomial on $]t_j, t_{j+1}[$) and define $D^j s(t_l)$ by making $D^j s$ continuous on the right. From de Boor (2001, page 117), we have that

$$Ds = (k-1)\sum_{j=2}^{m} \beta_j' B_{j,k-1},$$

$$D^2 s = (k-2)\sum_{j=3}^{m} \beta_j'' B_{j,k-2},$$

with $\beta_j' = (\beta_j - \beta_{j-1})/(t_{j+k-1} - t_j)$ and $\beta_j'' = (\beta_j' - \beta_{j-1}')/(t_{j+k-2} - t_j)$.

a) If C is greater that c then, by definition of C, $\beta_j \geq c$ for all j. Since $\sum_{j=1}^m B_{j,k} = 1$ (de Boor, 2001, page 96), for all $x \in [a, b]$, we have:

$$s(x) = \sum_{j=1}^{m} \beta_j B_{j,k}(x) \geq (\inf_j \beta_j) \sum_{j=1}^{m} B_{j,k}(x) = \inf_j \beta_j \geq c.$$

Hence, s is greater that c. The proof is similar if C is strictly greater, lower or strictly lower than c. Assume that C is increasing, then s is increasing since, for all $x \in [a, b]$:

$$Ds(x) = \sum_{j=2}^{m} \beta_j' B_{j,k-1}(x) \geq (\inf_j \beta_j') \sum_{j=2}^{m} B_{j,k-1}(x) = (\inf_j \beta_j') \geq 0.$$

If C is strictly increasing, decreasing or strictly decreasing, the proof is similar. Finally, for the convex case, it is easy to see that C is convex if and only if $\beta_j'' \geq 0$ for all $j \geq 2$ and we deduce that $D^2 s \geq \inf_j \beta_j'' \geq 0$ by similar arguments. The proof is similar if C is strictly convex, concave or strictly concave.

b) Assume that C is κ-gl than c. Then, it is easily seen that $S^-(C-c) = \kappa - 1$ and, from the definition of C, that $S^-(C-c) = S^-(\beta - c)$ where $\beta = (\beta_1, \dots, \beta_m)$ and $\beta - c = (\beta_1 - c, \dots, \beta_m - c)$. Using that $S^-(\sum_j \alpha_j B_{j,k}) \leq S^-(\alpha)$ for all sequence $\alpha = (\alpha_1, \dots, \alpha_m)$ (de Boor, 2001, Corollary 28 page 139), we have:

$$S^-\left(\sum_{j=1}^m \beta_j B_{j,k} - c\right) = S^-\left(\sum_{j=1}^m (\beta_j - c) B_{j,k}\right) \leq S^-(\beta - c) = \kappa - 1;$$

hence, s is κ'-greater or κ'-lower than c with $1 \leq \kappa' \leq \kappa$. Assume that C is κ-monotone. Then, similarly, $S^-(Ds) \leq S^-(\beta') = \kappa - 1$ and s is κ'-monotone with $1 \leq \kappa' \leq \kappa$. Finally, if C is κ-cc, it can be checked that $S^-(\beta'') = \kappa - 1$ and we conclude by noting that $S^-(D^2 s) \leq S^-(\beta'') = \kappa - 1$.

c) Assume that C is κ-greater than c. Then, by definition of C, there exist $\beta_{i_1}, \dots, \beta_{i_\kappa}$ such that $\beta_{i_1} > c, \beta_{i_2} < c, \beta_{i_3} > c, \dots$ From de Boor (2001, page 134),

$$|\beta_j - s(t_j^*)| \leq \text{const}_k \, |\mathbf{t}|^2 \sup_{x \in [t_{j+1}, t_{j+k-1}]} |D^2 s(x)|, \tag{6.1}$$

where const_k is a constant that may depend on k. Then, for $|\mathbf{t}|$ small enough, $s(t_{i_1}^*) > c, s(t_{i_2}^*) < c, s(t_{i_3}^*) > c, \dots$ Thus, s is κ'-gl than c with $\kappa' \geq \kappa$. Since from b), $\kappa' \leq \kappa$ we conclude that $\kappa' = \kappa$. To prove that s is κ-greater than c, assume that s is κ-lower than c. Then, there exists $x_0 < t_{i_1}^*$ such that $s(x_0) < c$ and s is κ'-greater than c with $\kappa' > \kappa$ which impossible.

Assume that C is κ-increasing. Then, there exists $i_1 < i_2 < \cdots < i_\kappa$ such that $\beta_{i_1} < \beta_{i_1+1}, \beta_{i_2} > \beta_{i_2+1}, \beta_{i_3} < \beta_{i_3+1}, \dots$ Then, for $|\mathbf{t}|$ small enough, by (6.1), we have that $s(t_{i_1}^*) < s(t_{i_1+1}^*), s(t_{i_2}^*) > s(t_{i_2+1}^*), s(t_{i_3}^*) < s(t_{i_1+3}^*), \dots$ Then, s is κ'-monotone with $\kappa' \geq \kappa$ and by b), $\kappa' = \kappa$. If s were κ-decreasing, there exists $x_0 < t_{i_1}^*$ such that $s(x_0) > s(t_{i_1}^*)$ and s is κ'-monotone with $\kappa' > \kappa$ which is impossible. Thus, s is κ'-increasing.

Assume that C is κ-convex. Then, there exist $i_1 < i'_1 < i''_1 < i_2 < i'_2 < i''_2 < \cdots < i_\kappa < i'_\kappa < i''_\kappa <$ such that

$$\frac{\beta_{i'_1} - \beta_{i_1}}{t^*_{i'_1} - t^*_{i_1}} < \frac{\beta_{i''_1} - \beta_{i'_1}}{t^*_{i''_1} - t^*_{i'_1}}, \frac{\beta_{i'_2} - \beta_{i_2}}{t^*_{i'_2} - t^*_{i_2}} > \frac{\beta_{i''_2} - \beta_{i'_2}}{t^*_{i''_2} - t^*_{i'_2}}, \frac{\beta_{i'_3} - \beta_{i_3}}{t^*_{i'_3} - t^*_{i_3}} < \frac{\beta_{i''_3} - \beta_{i'_3}}{t^*_{i''_3} - t^*_{i'_3}}, \ldots$$

Then, we conclude that s is κ-convex by using (6.1) as above. □

6.2 Convergence of Algorithm 1

In this section, an element of the form $(\beta_1, \ldots, \beta_m, \sigma^2) \in S \times [0, \infty)$ will be denoted x or y. For such an element x, it will sometimes be convenient to write $x = (x^m, x_{m+1})$ where $x^m = (x_1, \ldots, x_m) \in S$ and $x_{m+1} \in [0, \infty[$. Recall that, for two transition kernels K_1 and K_2 and μ a measure, the composition $K_1 K_2$ is the transition kernel defined by

$$K_1 K_2(x, A) = \int K_1(x, dy) K_2(y, A), \tag{6.2}$$

and μK_1 is the measure defined by

$$\mu K_1(x, A) = \int \mu(dy) K_1(y, A),$$

for all measurable set A.

Denote by P_1 and $Q_{2,l}$ the transition kernels associated with step 1 and step 2b of Algorithm 1 respectively. Thus, for $1 \leq l \leq m$, we have :

$$\begin{aligned} P_1(x, dy) &= Q_{\sigma^2|\beta}(x^m, dy_{m+1}) \prod_{j=1}^{m} \delta_{x_j}(dy_j), \\ Q_{2,l}(x, dy) &= U_{l,x^m}(dy_l) \prod_{j \neq l} \delta_{x_j}(dy_j), \end{aligned}$$

where $Q_{\sigma^2|\beta}(x^m, .)$ is the posterior conditional distribution of σ^2 given that $\beta = x^m$, δ_{x_j} is the Dirac measure at x_j and U_{l,x^m} is the uniform distribution on $S_{x,l}$ (defined in Section 3). Denote by $\alpha(x, y)$ the acceptance probability of Algorithm 1 for moving from x to y, by Π the posterior distribution and by $P_{2,l}$ the kernel transition associated with steps 2b and 2c of Algorithm 1:

$$P_{2,l}(x, dy) = \alpha(x, y)\ Q_{2,l}(x, dy) + (1 - r_l(x))\ \delta_x(dy),$$

with $r_l(x) = \int \alpha(x,y)\ Q_{2,l}(x,dy)$. Then, by construction of α (and using calculations as in Proposition 9 of Abraham, 2009) it can be seen that:

$$\alpha(x,y)\ Q_{2,l}(x,dy)\ \Pi(dx) = \alpha(y,x)\ Q_{2,l}(y,dx)\ \Pi(dy).$$

Then, following Tierney (1994, page 1705), it is classical to check that Π is stationary for $P_{2,l}$, that is $\Pi P_{2,l} = \Pi$. Denote by P_2 the transition kernels of step 2 of Algorithm 1:

$$P_2 = \frac{1}{m!} \sum_{\tau \in \mathcal{P}} P_{2,\tau(1)} \dots P_{2,\tau(m)},$$

where $\mathcal{P}$ is the set of permutations of $\{1, \dots, m\}$. Then, clearly, Π is stationary for P_2. As Π is also stationary for P_1 (see Tierney, 1994, page 1703), it is stationary for the transition kernel of the chain $P = P_1 P_2$.

In the following, for two measures μ_1 and μ_2 , we shall denote by $\mu_1 \sim \mu_2$ (respectively $\mu_1 \geq \mu_2$) when $\mu_1(A) = 0$ if and only if $\mu_2(A) = 0$ (respectively $\mu_1(A) \geq \mu_2(A)$). For two transition kernels K_1 and K_2, we simply write $K_1 \sim K_2$ (respectively $K_1 \geq K_2$) if $K_1(x,.) \sim K_2(x,.)$ (respectively $K_1(x,.) \geq K_2(x,.)$) for all x.

Proposition 2 *a) Let $x \in S \times [0, \infty[$ and let A be a measurable subset of $S \times [0, \infty[$. If there exist an integer $k \geq 1$ and $(i_1, \dots, i_k) \in \{1, \dots, m\}^k$ such that $P_1 Q_{2,i_1} \dots Q_{2,i_k}(x, A) > 0$, then $P^k(x, A) > 0$.*
b) For all $x \in S \times [0, \infty)$ and all measurable subset $A \subset S \times [0, \infty)$, we have that $P(x, A) \geq \tilde{P}(x, A)$ where

$$\tilde{P}(x, A) = \int \mathbf{1}_A(x')\ Q_{\sigma^2|\beta}(x, dx'_{m+1}) \prod_{l=1}^{m} (1 - r_l(x')),$$

with $x' = (x^m, x'_{m+1})$.

Proof a) Note that, $P_{2,l} \geq \alpha(x,y) Q_{2,l}(x,dy)$ and $P_{2,l} \geq (1 - r_l(x))\delta_x$ for all l. Then, from Lemma 1, if we set $R_l(x) = \prod_{j=1}^{l-1}(1 - r_j(x)) \times \prod_{j=l+1}^{m}(1 - r_j(x))$, we have:

$$P_{2,1} \dots P_{2,l-1} P_{2,l+1} \dots P_{2,m} P_{2,l}(x, dy) \geq R_l(x) \alpha(x,y) Q_{2,l}(x,dy),$$

and

$$P_2(x, dy) \geq \frac{1}{m!} R_l(x) \alpha(x,y) Q_{2,l}(x,dy).$$

If $\widehat{Q}_{2,l}(x,dy)$ denotes the right hand side of the last inequality, we have $P \geq P_1\widehat{Q}_{2,l}$ for all l and

$$P^k \geq (P_1\widehat{Q}_{2,i_1})\dots(P_1\widehat{Q}_{2,i_k}).$$

As $\alpha(x,y) > 0$ and $0 < r_l(x) < 1$ for all $x, y \in S \times [0,\infty[$, $Q_{2,l} \sim \widehat{Q}_{2,l}$. Then, by Lemma 1,

$$(P_1\widehat{Q}_{2,i_1})\dots(P_1\widehat{Q}_{2,i_k}) \sim (P_1Q_{2,i_1})\dots(P_1Q_{2,i_k}).$$

Note that $Q_{2,l}(x,dy)$ only changes the l-th coordinate of x by a uniform distribution whose support does not depend on x_{m+1} and that $P_1(x,dy)$ only changes the $(m+1)$-th coordinate of x by a distribution with support $[0,\infty)$. Thus, from the definitions of P_1 and $Q_{2,l}$, it is easy to see that $P_1Q_{2,l} \sim Q_{2,l}P_1$, and, consequently that

$$\begin{aligned}(P_1Q_{2,i_1})\dots(P_1Q_{2,i_k}) &\sim P_1^kQ_{2,i_1}\dots Q_{2,i_k}, \\ &\sim P_1Q_{2,i_1}\dots Q_{2,i_k};\end{aligned}$$

hence the result.

b) Let $D_l(x,dy) = (1 - r_l(x))\ \delta_x(dy)$. By Lemma 1, it is easy to see that $P_{2,l} \geq D_l$, that $P_2 \geq D_1\dots D_m$ and that $P \geq P_1D_1\dots D_m$. We conclude by noting that $P_1D_1\dots D_m = \tilde{P}$. □

Lemma 1 *Let K_1, K_2, Q_1 and Q_2 be transition kernels. If $K_1 \sim K_2$ and $Q_1 \sim Q_2$ then $K_1K_2 \sim Q_1Q_2$. If $K_1 \geq K_2$ and $Q_1 \geq Q_2$ then $K_1K_2 \geq Q_1Q_2$.*

Lemma 1 is a straightforward consequence of (6.2); its proof is left to the reader.

Roughly speaking, Proposition 2 says that, if it it possible to move from $x \in S \times [0,\infty)$ to a subset $A \subset S \times [0,\infty)$ by only changing a sequence of coordinates alternatively, then $P(x,A) > 0$. For all usual shape constraints such as monotonicity, convexity (concavity) or unimodality, it is intuitively clear, although a formal proof is not easy to write, that it is possible to reach any set A such that $\Pi(A) > 0$ from all $x \in S \times [0,\infty)$ by changing some coordinates alternatively. More formally, for all $x \in S \times [0,\infty)$ and all A such that $\Pi(A) > 0$, there exist k and $(i_1,\dots,i_k)$ such that $P_1Q_{2,i_1}\dots Q_{2,i_k}(x,A) > 0$; hence the Π-irreducibility of P by Proposition 2.

Let us now prove that the chain is aperiodic. To obtain a contradiction, suppose that there exist $d \geq 2$ disjoint subsets $A_1, \ldots, A_d \subset S \times [0, \infty)$ with $P(x, A_{i+1}) = 1$ for all $x \in A_i$ for $1 \leq i \leq d-1$ and $P(x, A_1) = 1$ for all $x \in A_d$, such that $\Pi(A_i) > 0$ for all i. As Π is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m \times [0, \infty)$, we have that

$$0 < \int \mathbf{1}_{A_1}(x)dx = \int \left[\int \mathbf{1}_{A_1}(x)dx_{m+1} \right] dx^m,$$

with $x = (x^m, x_{m+1}) \in \mathbb{R}^m \times [0, \infty)$ and d_x, d_{x^m} and dx_{m+1} denote the Lebesgue measure on $\mathbb{R}^m \times [0, \infty)$, $\mathbb{R}^m$ and $[0, \infty)$ respectively. Thus, there exists $x \in A_1$ such that

$$\int \mathbf{1}_{A_1}(x)dx_{m+1} > 0.$$

If we denotes by $q_{\sigma^2|\beta}(x, .)$ the (inverse gamma) density of $Q_{\sigma^2|\beta}(x, .)$ with respect to the Lebesgue measure on $[0, \infty)$, as $q(x, x'_{m+1})(1 - r_l(x)) > 0$ for all $x'_{m+1} > 0$, we deduce that

$$\int \mathbf{1}_{A_1}(x') \; q(x, x'_{m+1}) \prod_{l=1}^{m} (1 - r_l(x')) \; dx'_{m+1} > 0,$$

with $x' = (x^m, x'_{m+1})$. By Proposition 2, we deduce that $P(x, A_1) > 0$, hence a contradiction. Finally, from Tierney (1994, Theorem 1), we have that, for Π-almost all x, $\|P^\nu(x, .) - \Pi\| \to 0$ as $\nu \to \infty$, with $\|.\|$ denoting the total variation distance.

6.3 Parameters values of Section 4.3

Under H_0, we take $\mu = (-0.65, -0.57, -0.49, -0.41, -0.33, -0.25, -0.17, -0.1, -0.02, 0.05, 0.13, 0.21, 0.29, 0.37, 0.45)$, $V = I_{15}$ (identity matrix), $\xi = 0.01$ and $\varsigma = 0.01$. We take the same values under H_1 except for μ which is replaced by $\mu = (0.44, -0.40, -0.32, -0.27, -0.36, -0.46, -0.38, -0.17, -0.027, -0.17, -0.09, 0.01, 0.22, 0.47, 0.48)$. For both the hypothesis, the knot sequence is $\mathbf{t} = (1849.50, 1849.75, 1850.00, 1862.46, 1874.92, 1887.38, 1899.84, 1912.30, 1924.76, 1937.23, 1949.69, 1962.15, 1974.61, 1987.07, 1999.53, 2012.00, 2012.25, 2012.50)$.

Acknowledgments

We thank a reviewer for a careful reading of the paper and for many helpful suggestions.

References

Abraham C. (2009). A computation method in robust bayesian decision theory. *International Journal of Approximate Reasoning*, **50**, 289–302.

Abraham C. (2012). Bayesian regression under combinations of constraints. *Journal of Statistical Planning and Inference*, **142**, 2672–2687.

Alvarez E.E. & Dey D.K. (2009). Bayesian isotonic changepoint analysis. *Annals of the Institute of Statistical Mathematics*, **61**, 355–370.

Alvarez E.E. & Yohai V.J. (2012). M-estimators for isotonic regression. *Journal of Statistical Planning and Inference*, **142**, 2351–2368.

Bartoli N. & Del Moral P. (2001). *Simulation et algoritmes stochastiques.* Cepadues, Toulouse.

Boyd S. & Vandenberghe L. (2004). *Convex Optimization.* Cambridge University Press.

de Boor C. (2001). *A practical guide to splines.* Springer-Verlag, New-York.

Delecroix M. & Thomas-Agnan C. (2000). Spline and kernel regression under shape restrictions. Dans *Smoothing and Regression: Approaches, Computation, and Application*, réd. M.G. Schimek, pp. 109–133. John Wiley & Sons, Inc.

Gelfand A.E. & Dey D.K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of Royal Statistical Society B*, **56**, 501–514.

Geweke J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraint. Dans *Computing Science and Statistics: Proceedings of the 23-rd Symposium on the Interface*, réd. E.M. Keramidas, pp. 571–578. Interface Foundation of North America, INc., Fairfax.

Gunn L.N. & Dunson D.B. (2005). A transformation approach for incorporating monotone or unimodal constraints. *Biostatistics*, **6**, 434–449.

Holmes C.C. & Heard N.A. (2003). Generalized monotonic regression using random change points. *Statistics in Medicine*, **22**, 623–638.

Jeanson S., Hilgert N., Coquillard M., Seukpanya C., Faiveley M., Neuveu P., Abraham C., Georgescu V., Fourcassi P. & Beuvier E. (2009). Milk acidification by lactococcus lactis is improved by decreasing the level of dissolved oxygen rather than decreasing redox potential in the potential in the milk prior to inoculation. *International Journal of Food Microbiology*, **131**, 75–81.

Jones G., Leung J. & Robertson H. (2009). A mixed model for investigating a population of asymptotic growth curves using restricted b-splines. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 66–78.

Jones P., Parker D., Osborn T. & Briffa K. (2011). Global and hemispheric temperature anomalies, land and marine instrumental records.

Jun-wu & liang Tian G. (2011). Efficient algorithms for generating truncated multivariate normal distributions. *Acta Mathematicae Applicatae Sinica, English Series*, **27**, 601–612.

Lavine M. & Mockus A. (1995). A nonparametric bayes method for isotonic regression. *Journal of Statistical Planning and Inference*, **46**, 235–248.

Mammen E., Marron J., Turlach B. & Wand M. (2001). A general projection framework for constrained smoothing. *Statistical Science*, **16**, 232–248.

Mammen E. & Thomas-Agnan C. (1999). Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics*, **26**, 239–252.

Meyer M.C. (2008). Inference using shape-restricted regression splines. *The Annals of Applied Statistics*, **2**, 1013–1033.

Meyer M.C., Hackstadt A.J. & Hoeting J.A. (2011). Bayesian estimation and inference for generalised partial linear models using shape-restricted splines. *Journal of Nonparametric Statistics*, **23**, 867–884.

Neelon B. & Dunson D.B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics*, **60**, 398–406.

Newton M. & Raftery A. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society (B)*, **56**, 3–48.

Plummer M., Best N., Cowles K. & Vines K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News*, **6**(1), 7–11.

Raftery A.E. & Lewis S. (1996). Implementing mcmc. Dans *Markov Chain Monte Carlo in Practice*, réd. W.R. Gilks, S. Richardson & D.J. Spiegelhalter, pp. 115–130. Chapman & Hall, London.

Ramsay J.O. (1988). Monotone regression splines in action. *Statistical Science*, **3**, 425–461.

Ramsay J.O. (1998). Estimating smooth monotone functions. *Journal of Royal Statistical Society B*, **60**, 365–375.

Shively T.S., Sager T.W. & Walker S.G. (2009). A bayesian approach to non-parametric monotone function estimation. *Journal of The royal Statistical Society B*, **71**, 159–175.

Tierney L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22**, 1701–1762.

Turlach B. (2005). Shape constrained smoothing using smoothing splines. *Computational Statistics*, **20**, 81–103.

Wu W.B., Woodroofe M. & Mentz G. (2001). Isotonic regression: another look et the changepoint problem. *Biometrika*, **88**, 793–804.

Zhao O. & Woodroofe M. (2012). Estimating a monotone trend. *Statistica Sinica*, **22**, 359–378.