Parametrizations, fixed and random effects

Azzouz Dermoune, Cristian Preda

Parametrizations, fixed and random effects

Azzouz Dermoune

Laboratoire Paul Painlevé
USTL-UMR-CNRS 8524.
UFR de Mathématiques, Bât. M2.
59655 Villeneuve d'Ascq Cédex, France.
e-mail azzouz.dermoune@univ-lille1.fr

Cristian Preda

Laboratoire Paul Painlevé and
Institute of Mathematical Statistics and Applied Mathematics
Romanian Academy, USTL-UMR-CNRS 8524
Inria Lille (Modal)
59655 Villeneuve d'Ascq Cédex, France.
e-mail cristian.preda@polytech-lille.fr

Abstract

We consider the problem of estimating the random element $\boldsymbol{s}$ of a finite dimensional vector space S from the continuous data corrupted by noise with unknown variance σ_w^2. The mean $\mathrm{E}(\boldsymbol{s})$ (the fixed effect) of $\boldsymbol{s}$ belongs to a known vector subspace F of S, and the likelihood of the centred component $\boldsymbol{s} - \mathrm{E}(\boldsymbol{s})$ (the random effect) belongs to an unknown supplementary space E of F relative to S and has the PDF proportional to $\exp\{-q(\boldsymbol{s})/2\sigma_s^2\}$, where σ_s^2 is some unknown positive parameter. We introduce the notion of bases separating the fixed and random effects and define comparison criteria between two separating bases using the partition functions and the maximum likelihood method. We illustrate our results for climate change detection using the set S of cubic splines. We show the influence of the choice of separating basis on the estimation of the linear tendency of the temperature and the signal-to-noise ratio σ_w^2/σ_s^2.

Keywords: General linear model, fixed effect, random effect, cubic spline, smoothing parameter, likelihood, climate change detection.

1. Motivation

We consider the problem of climate change detection. The years taken into account and the annual mean temperature are denoted by $t_1 < \ldots < t_{n+1}$ and $y_1, \ldots, y_{n+1}$ respectively. In our work we consider the additive model

$$y(i) = s(t_i) + w_i, \quad i = 1, \ldots, n+1, \tag{1}$$

where $(w_1, \ldots, w_{n+1})$ is a Gaussian white noise with the variance σ_w^2, and $s(t_i)$ is the true temperature at the year t_i. We model the behaviour of the true temperature $\boldsymbol{s}$ by a random element of the set $S_3(t_1, \ldots, t_{n+1})$ of cubic splines having the knots $t_1 < \ldots < t_{n+1}$.

The fixed effect $\mathrm{E}(\boldsymbol{s})$ is a straight line. It belongs to the null space $Q^{-1}(0)$ of the quadratic form

$$Q(\boldsymbol{s}) = \int_{t_1}^{t_{n+1}} |s''(t)|^2 dt. \tag{2}$$

Here s', s'' are respectively the first and the second derivative of the map $\boldsymbol{s}$.

The random effect $\boldsymbol{s} - \mathrm{E}(\boldsymbol{s})$ belongs to an unknown supplementary space E of $Q^{-1}(0)$ relative to $S_3(t_1, \ldots, t_{n+1})$ and has the probability distribution

$$\exp\{-Q(\boldsymbol{s})/2\sigma_s^2\}\mathbf{1}_E(\boldsymbol{s})d\boldsymbol{s}/Z_{\sigma_s}. \tag{3}$$

Here σ_s^2 is a positive parameter which measures the dispersion of the random effect around the space $Q^{-1}(0)$, and may depend on the space E, and Z_{σ_s} is the partition function. We assume that the noise (w_i) is independent of $\boldsymbol{s}$, but its variance σ_w^2 may depend on the space E.

A popular estimator of the temperature $\boldsymbol{s}$ is given by the following penalized estimation technique [15]

$$\hat{\boldsymbol{s}} = \arg\min\{\sum_{i=1}^{n+1} |y(i) - s(t_i)|^2 + \lambda Q(\boldsymbol{s}) : \boldsymbol{s} \in H^2\}, \tag{4}$$

where H^2 is the infinite dimensional space of all functions with square integrable second derivative, and $\lambda > 0$ denotes the smoothing parameter. Generalized cross-validation techniques are among the automatic methods used to estimate the smoothing parameter see, e.g., [2], [9], [15]. The estimator $\hat{\boldsymbol{s}}$ (4), for λ fixed, belongs to the set $S_3(t_1, \ldots, t_{n+1})$ of cubic splines and does not depend on the parametrization of $S_3(t_1, \ldots, t_{n+1})$. See [4], [5] for a similar study.

The concept of fixed and random effects has been applied to the analysis of longitudinal data. See, e.g., [8], [12], [14], [16]. In [11] the authors have modelled the fixed effects nonparametrically using truncated series expansions with B-spline basis. They have selected the fixed effects using lasso methodology, while the random effects are estimated using the Newton Raphson algorithm.

In our work the sum of the fixed and the random effects is a random cubic spline. The fixed effects are straight lines. The random effects are the supplementary spaces of the space of straight lines relative to $S_3(t_1, \ldots, t_{n+1})$. We introduce an original notion of basis separating the space of straight lines (fixed effects) from its supplementary spaces (random effects) relative to $S_3(t_1, \ldots, t_{n+1})$. We interpret the smoothing parameter as the signal-to-noise ratio σ_w^2/σ_s^2. We show that the estimator of the smoothing parameter is a function of such a basis. We also show that there exists an infinite number of separating bases and we propose comparison criteria between two separating bases using the partition functions and the maximum-likelihood method.

The plan of our work is the following. In Section 2, we introduce the notion of separating bases in a general setting , and show that they determine the shape of the fixed effect and the parametrization of the random effect, but are not enough to determine the shape of the random effect and the parametrization of the fixed effect. In Section 3 we return to the climate change detection. We show that the random effect is parametrized by $s''(t_1)$, $\ldots$, $s''(t_{n+1})$, and the fixed effect is parametrized by two independent linear forms of the vectors

$$(s(t_1), \ldots, s(t_{n+1})), (s'(t_1), \ldots, s'(t_{n+1})).$$

We construct four bases separating the fixed and the random effects, and calculate in each basis the maximum likelihood estimators of the fixed and the random effects and the dispersion parameters σ_s^2 and σ_w^2. Finally we consider the separating basis as a parameter and we estimate it by maximizing the likelihood.

2. Separating Basis in a general setting

In this section the set S is any finite dimensional vector space having the dimension p, and $\boldsymbol{s}$ a random element of S. Its probability distribution is defined by the fixed space F having the dimension $k < p$ and the measurable map $q : S \to \mathbf{R}$ such that for all $\nu > 0$ and for all supplementary space E of F relative to S, the partition function

$$Z_\nu := \int_E \exp\{-q(\boldsymbol{s})/2\nu^2\} d\boldsymbol{s} < +\infty.$$

Here $d\boldsymbol{s}$ denotes the Lebesgue measure on E. The random vector $\boldsymbol{s}$ has the mean (fixed effect) $\mathrm{E}(\boldsymbol{s}) \in F$. We say that F is the shape of the fixed effect. The random effect $\boldsymbol{s} - \mathrm{E}(\boldsymbol{s})$ belongs to some supplementary space E of F relative to S. Knowing the space E (the shape of the random effect), $\boldsymbol{s} - \mathrm{E}(\boldsymbol{s})$ has the probability distribution

$$\exp\{-q(\boldsymbol{s})/2\sigma_s^2\} \mathbf{1}_E(\boldsymbol{s}) d\boldsymbol{s}/Z_{\sigma_s}$$

Here σ_s^2 is a positive parameter which measures the dispersion of the random effect.

Now we need to parametrize the set S in order to define properly an element $\boldsymbol{s} \in S$. A parametrization of S is a one-to-one linear map

$$\Theta : \boldsymbol{s} \in S \to \theta \in \mathbb{R}^p.$$

Defining a parametrization Θ is equivalent to the existence of the basis $\mathbf{B} := (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p)$ of S such that for all $\boldsymbol{s} \in S$

$$\boldsymbol{s} = \sum_{i=1}^{p} \theta_i \boldsymbol{b}_i := \mathbf{B}\theta.$$

Definition 2.1. We say that a basis $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p)$ separates $(\theta_1, \ldots, \theta_k)$ from $(\theta_{k+1}, \ldots, \theta_p)$ (or simply $\mathbf{B}$ is a separating basis) if $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$ is a basis of the set of the fixed effect F.

We will show in assertion 2 of Proposition 2.3 below, for all separating bases $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p)$ that the component $\sum_{j=k+1}^{p} \theta_j \boldsymbol{b}_j$ has the PDF

$$\exp\{-q(\sum_{j=k+1}^{p} \theta_j \boldsymbol{b}_j)/2\sigma_s^2\} d\theta_{k+1} \ldots d\theta_p / Z_{\sigma_s}(\boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_p). \tag{5}$$

The notation $Z_{\sigma_s}(\boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_p)$ means that the partition function depends on the family $\boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_p$. However the component $\sum_{j=1}^k \theta_j \boldsymbol{b}_j$ is only a candidate for the fixed effect , i.e., $\sum_{j=1}^k \theta_j \boldsymbol{b}_j$ is a candidate for the mean of $\boldsymbol{s}$.

The following proposition shows that there is an infinite number of separating bases, and the choice of a fixed effect depends on the practitioner's aim.

Proposition 2.2. *There is an infinite number of separating bases.*

Proof. Starting from any basis $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$ of F and using the incomplete basis theorem we can construct an infinite number of bases $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p)$ of S, which achieves the proof. □

Two parametrizations Θ^1 and Θ^2 are related by the passage matrix $\mathbf{P}$:

$$\theta_i^{(2)} = \sum_{j=1}^{p} \mathbf{P}_{ij} \theta_j^{(1)}, \; i = 1, \ldots, p, \tag{6}$$

or equivalently their respective bases $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ are related by

$$\boldsymbol{b}_i^{(2)} = \sum_{j=1}^{p} \mathbf{P}_{ji}^{-1} \boldsymbol{b}_j^{(1)}.$$

Here $\mathbf{P}^{-1}$ denotes the inverse of the passage matrix $\mathbf{P}$. Now we ask the following question. What is the link between two separating bases?

Proposition 2.3. *Let* $\mathbf{B}^{(1)} = (\boldsymbol{b}_1^{(1)}, \ldots, \boldsymbol{b}_p^{(1)})$ *and* $\mathbf{B}^{(2)} = (\boldsymbol{b}_1^{(2)}, \ldots, \boldsymbol{b}_p^{(2)})$ *be two separating bases. Let* $\Theta^{(1)}$ *and* $\Theta^{(2)}$ *be the parametrizations of the element* $\boldsymbol{s}$ *in* $\mathbf{B}^{(1)}$ *and* $\mathbf{B}^{(2)}$ *respectively.*

1) The passage matrix $\mathbf{P}$*, given by* $\Theta^{(2)} = \mathbf{P}\Theta^{(1)}$*, has the following form*

$$\begin{pmatrix} \mathbf{P}(1:k, 1:k) & \mathbf{P}(1:k, k+1:p) \\ 0(k+1:p, 1:k) & \mathbf{P}(k+1:p, k+1:p) \end{pmatrix},$$

where for $1 \leq n_1 < n_2 \leq p$*,* $\mathbf{P}(n_1 : n_2, n_3 : n_4)$ *is the sub-matrix* $(\mathbf{P}_{ij} : n_1 \leq i \leq n_2, n_3 \leq j \leq n_4)$ *of* $\mathbf{P}$ *and* $0(k+1:p, 1:k)$ *denotes the sub-matrix of the null matrix* $0(1:p, 1:p)$*.*

2) We have $(\theta_{k+1}^{(2)}, \ldots, \theta_p^{(2)})^\top = \mathbf{P}(k+1:p, k+1:p)(\theta_{k+1}^{(1)}, \ldots, \theta_p^{(1)})^\top$*.*

3) The components $\theta_1^{(2)}, \ldots, \theta_k^{(2)}$ *may depend on the all parameters* $(\theta_1^{(1)}, \ldots, \theta_p^{(1)})$*.*

4) The elements $(\boldsymbol{b}_1^{(2)}, \ldots, \boldsymbol{b}_k^{(2)})$ *depend only on* $(\boldsymbol{b}_1^{(1)}, \ldots, \boldsymbol{b}_k^{(1)})$*. But the elements* $(\boldsymbol{b}_{k+1}^{(2)}, \ldots, \boldsymbol{b}_p^{(2)})$ *may depend on the whole basis* $(\boldsymbol{b}_1^{(1)}, \ldots, \boldsymbol{b}_p^{(1)})$*.*

Proof. 1) If $\theta_i^{(2)} = 0$, $i = k+1, \ldots, p$, then from

$$\sum_{i=1}^{p} \theta_i^{(2)} \boldsymbol{b}_i^{(2)} = \sum_{i=1}^{p} \theta_i^{(1)} \boldsymbol{b}_i^{(1)},$$

we have

$$\sum_{i=1}^{k} \{\theta_i^{(2)} \boldsymbol{b}_i^{(2)} - \theta_i^{(1)} \boldsymbol{b}_i^{(1)}\} = \sum_{i=k+1}^{p} \theta_i^{(1)} \boldsymbol{b}_i^{(1)}.$$

The left-hand side term belongs to F, and the right-hand side term belongs to the supplementary space $span(\boldsymbol{b}_{k+1}^{(1)}, \ldots, \boldsymbol{b}_p^{(1)})$ of F. It follows that $\theta_i^{(1)} = 0$ for all $i = k+1, \ldots, p$, which achieves the proof of 1). The proof of 2), 3) and 4) is a consequence of 1). □

The vector $\boldsymbol{s}$ has in two separating bases $\mathbf{B}^{(1)}$, $\mathbf{B}^{(2)}$ the following decompositions

$$\begin{aligned} \boldsymbol{s} &= \sum_{i=1}^{k} \theta_i^{(1)} \boldsymbol{b}_i^{(1)} + \sum_{i=k+1}^{p} \theta_i^{(1)} \boldsymbol{b}_i^{(1)} \\ &= \sum_{i=1}^{k} \theta_i^{(2)} \boldsymbol{b}_i^{(2)} + \sum_{i=k+1}^{p} \theta_i^{(2)} \boldsymbol{b}_i^{(2)}. \end{aligned}$$

1) If the vector $\theta_{k+1}^{(2)}, \ldots, \theta_p^{(2)}$ has the PDF

$$\exp\{-q(\sum_{j=k+1}^{p} \theta_j^{(2)} \boldsymbol{b}_j^{(2)})/2\sigma_s^2\} d\theta_{k+1}^{(2)} \ldots d\theta_p^{(2)} / Z_{\sigma_s}(\boldsymbol{b}_{k+1}^{(2)}, \ldots, \boldsymbol{b}_p^{(2)}), \tag{7}$$

then assertion 2 of Proposition 2.3 and the formula of change of variables tell us that the vector $\theta_{k+1}^{(1)}, \ldots, \theta_p^{(1)}$ has the PDF

$$\exp\{-q(\sum_{j=k+1}^{p} \theta_i^{(1)} \boldsymbol{b}_j^{(1)})/2\sigma_s^2\} d\theta_{k+1}^{(1)} \ldots d\theta_p^{(1)} / Z_{\sigma_s}(\boldsymbol{b}_{k+1}^{(1)}, \ldots, \boldsymbol{b}_p^{(1)})$$

where

$$Z_{\sigma_s}(\boldsymbol{b}_{k+1}^{(1)}, \ldots, \boldsymbol{b}_p^{(1)}) = Z_{\sigma_s}(\boldsymbol{b}_{k+1}^{(2)}, \ldots, \boldsymbol{b}_p^{(2)})/|det\{\mathbf{P}(k+1:p, k+1:p)\}|.$$

Hence, a random parameter in the basis $\mathbf{B}^{(2)}$ remains random in the basis $\mathbf{B}^{(1)}$. 2) If $\theta_1^{(2)}, \ldots, \theta_k^{(2)}$ are fixed and $\theta_{k+1}^{(2)}, \ldots, \theta_p^{(2)}$ has the PDF (7), then for $i = 1, \ldots, k$, the parameter

$$\theta_i^{(1)} = \sum_{j=1}^{p} \mathbf{P}_{ij}^{-1} \theta_j^{(2)}$$

may depend on the random vector $\theta_{k+1}^{(2)}, \ldots, \theta_p^{(2)}$. The parameters $\theta_1^{(1)}, \ldots, \theta_k^{(1)}$ are corrupted by $\theta_{k+1}^{(2)}, \ldots, \theta_p^{(2)}$. Hence, a fixed effect in a basis $\mathbf{B}^{(2)}$ is in general no longer fixed in another basis $\mathbf{B}^{(1)}$.

Definition 2.4. We say that two separating bases $\mathbf{B}^{(1)}$, $\mathbf{B}^{(2)}$ are equivalent if the components $\sum_{i=1}^{k} \theta_i^{(1)} \boldsymbol{b}_i^{(1)}$, $\sum_{i=1}^{k} \theta_i^{(2)} \boldsymbol{b}_i^{(2)}$ are the fixed effect of $\boldsymbol{s}$ respectively in the basis $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ and if the passage matrix (see Proposition2.3 assertion 2) satisfies $|det\{\mathbf{P}(k+1:p, k+1:p)\}| = 1$.

Remark 2.5. *Two separating bases $\mathbf{B}^{(1)}$, $\mathbf{B}^{(2)}$ are equivalent if and only if the passage matrix (6) has the following form*

$$\begin{pmatrix} \mathbf{P}(1:k, 1:k) & 0(1:k, k+1:p) \\ 0(k+1:p, 1:k) & \mathbf{P}(k+1:p, k+1:p) \end{pmatrix},$$

with $|det\{\mathbf{P}(k+1:p, k+1:p)\}| = 1$.

Note that the PDF (5) will concentrate around the minimizers of the map q as $Z_{\sigma_s}(\boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_p) \to 0$. Now we can compare two separating bases as follows.

Definition 2.6. We say that the parametrization $\Theta^{(1)} = (\theta_1^{(1)}, \ldots, \theta_p^{(1)})$ is more concentrated than $\Theta^{(2)} = (\theta_1^{(2)}, \ldots, \theta_p^{(2)})$ if for all $\nu > 0$,

$$Z_\nu(\boldsymbol{b}_{k+1}^{(1)}, \ldots, \boldsymbol{b}_p^{(1)}) < Z_\nu(\boldsymbol{b}_{k+1}^{(2)}, \ldots, \boldsymbol{b}_p^{(2)}).$$

3. Four separating bases of cubic splines

We consider the set $S := S_3(t_1, \dots, t_{n+1})$ of C^2 cubic splines having the knots $t_1 < \dots < t_{n+1}$. We recall that an element $\boldsymbol{s} \in S$ is a C^2 map on $[t_1, t_{n+1}]$ and is a polynomial of degree three on each interval $[t_i, t_{i+1}[$ for $i = 1,\dots, n$.

More precisely let

$$\begin{aligned}
&p_1 := s(t_1), \dots, p_{n+1} := s(t_{n+1}),\\
&q_1 := s'(t_1), \dots, q_{n+1} := s'(t_{n+1}),\\
&u_1 := s''(t_1), \dots, u_{n+1} := s''(t_{n+1}),\\
&v_1 = s'''(t_1+), \dots, v_n = s'''(t_n+)
\end{aligned}$$

be respectively the values of $\boldsymbol{s}$ and its derivatives up to order three on the knots. We have for $i = 1, \dots, n$,

$$s(t) = p_i + q_i(t - t_i) + (t - t_i)^2 u_i/2 + (t - t_i)^3 v_i/6, \quad t \in [t_i, t_{i+1}[.$$

The following constraint for $h_i = t_{i+1} - t_i$, $i = 1, \dots, n$ guarantees the hypothesis that $\boldsymbol{s}$ is C^2:

$$p_i + q_i h_i + u_i h_i^2/2 + v_i h_i^3/6 = p_{i+1}, \tag{8}$$
$$q_i + u_i h_i + v_i h_i^2/2 = q_{i+1}, \tag{9}$$
$$v_i = s^{(3)}(t_i) = (u_{i+1} - u_i)/h_i. \tag{10}$$

It is well known [3] that S has the dimension $n + 3$. Hence an element $\boldsymbol{s} \in S$ is completely defined by $n+3$ independent parameters. Note that the quadratic form (2) is equal to

$$\begin{aligned}
\sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} |u_i + t(u_{i+1} - u_i)/h_i|^2 dt &= \sum_{i=1}^{n} (u_i^2 + u_i u_{i+1} + u_{i+1}^2) h_i/3\\
&= \boldsymbol{u}^\top \mathbf{Q}_2 \boldsymbol{u}
\end{aligned}$$

where the column vector $\boldsymbol{u} = (u_1, \dots, u_{n+1})^\top$ and the invertible $(n + 1) \times (n + 1)$ matrix

$$\mathbf{Q}_2 := [q(i, j)] \tag{11}$$

is defined by $q(1,1) = h_1/3$, $q(n+1,n+1) = h_n/3$, for $i = 2,\ldots,n$, $q(i,i) = (h_{i-1}+h_i)/3$, for $i = 1,\ldots,n$, $q(i,i+1) = q(i+1,i) = h_i/6$, and $q(i,j) = 0$ elsewhere.

The PDF (3) means that the vector $\boldsymbol{u}$ is Gaussian, centred with the covariance matrix $\sigma_s^2 \mathbf{Q}_2^{-1}$. It follows for any separating basis $(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{n+3})$ that

$$Z_{\sigma_s}(\boldsymbol{b}_3,\ldots,\boldsymbol{b}_{n+3}) = \sigma_s^{n+1}/det(\mathbf{Q}_2). \tag{12}$$

Any other parametrization of the random effect is just a linear transformation of $(u_1,\ldots,u_{n+1})$, e.g., $(u_1,v_1,\ldots,v_n)$ is also a parametrization of the random effect. See assertion 2 of Proposition 2.3.

From Remark 2.5, we derive that for all $t \in [t_1, t_{n+1}]$ the fixed effect $\mathrm{E}(\boldsymbol{s})$ is given by $\mathrm{E}(s(t)) = \alpha + \beta t$, where

$$\alpha = \sum_{i=1}^{n+1} \{\alpha_i^{(0)} s(t_i) + \alpha_i^{(1)} s'(t_i)\} \tag{13}$$

$$\beta = \sum_{i=1}^{n+1} \{\beta_i^{(0)} s(t_i) + \beta_i^{(1)} s'(t_i)\}. \tag{14}$$

The coefficients $\alpha_i^{(0)}$, $\alpha_i^{(1)}$, $\beta_i^{(0)}$, $\beta_i^{(1)}$ define the parametrization of the fixed effect.

If a priori information tells us that the initial values p_1, p_2 are not random, then the parametrization $\Theta_{002} = (p_1, p_2, u_1, \ldots, u_{n+1})$ and the corresponding basis $\mathbf{B}_{002} = (\boldsymbol{b}_1^{002}, \ldots, \boldsymbol{b}_{n+3}^{002})$ are the right choice for separating the fixed effect (p_1, p_2) from the random effect $(u_1, \ldots, u_{n+1})$. The subscript notation 002 is justified by the fact that

$$p_1 := s(t_1) := s^{(0)}(t_1), p_2 := s(t_2) := s^{(0)}(t_2),$$
$$u_1 := s''(t_1) := s^{(2)}(t_1), \ldots, u_{n+1} := s''(t_{n+1}) := s^{(2)}(t_{n+1}).$$

If a priori information tells us that (p_1, q_1) are not random then the parametrizations

$$\Theta_{012} := (p_1, q_1, u_1, \ldots, u_{n+1}), \quad \Theta_{0123} := (p_1, q_1, u_1, v_1, \ldots, v_n),$$

and the corresponding bases

$$\mathbf{B}_{012} := (\boldsymbol{b}_1^{012}, \ldots, \boldsymbol{b}_{n+3}^{012}), \quad \mathbf{B}_{0123} := (\boldsymbol{b}_1^{0123}, \ldots, \boldsymbol{b}_{n+3}^{0123})$$

are the right choices for separating the fixed effect (p_1, q_1) from the random effect $(u_1, \ldots, u_{n+1})$. The subscript notation 012 is justified by the fact that $p_1 := s^{(0)}(t_1)$, $q_1 := s'(t_1) := s^{(1)}(t_1)$ and $u_1 := s^{(2)}(t_1)$, $\ldots$, $u_{n+1} := s^{(2)}(t_{n+1})$. Similarly, the notation 0123 is justified by the fact that $p_1 := s^{(0)}(t_1)$, $q_1 := s^{(1)}(t_1)$, $u_1 := s^{(2)}(t_1)$ and $v_1 := s'''(t_1+) := s^{(3)}(t_1+), \ldots, v_n = s'''(t_n+) := s^{(3)}(t_n+)$.

It follows for $\boldsymbol{s} \in S$ that

$$\begin{aligned}
\boldsymbol{s} \quad &= p_1 \boldsymbol{b}_1^{002} + p_2 \boldsymbol{b}_2^{002} + \sum_{i=1}^{n+1} u_i \boldsymbol{b}_{2+i}^{002} \\
&= p_1 \boldsymbol{b}_1^{012} + q_1 \boldsymbol{b}_2^{012} + \sum_{i=1}^{n+1} u_i \boldsymbol{b}_{2+i}^{012} \\
&= p_1 \boldsymbol{b}_1^{0123} + q_1 \boldsymbol{b}_2^{0123} + u_1 \boldsymbol{b}_3^{0123} + \sum_{i=1}^{n} v_i \boldsymbol{b}_{3+i}^{0123}.
\end{aligned}$$

The passage matrix from the parametrization Θ_{002} to the parametrization Θ_{012} is given by

$$\begin{aligned}
p_1 \quad &= p_1, \\
q_1 \quad &= (p_2 - p_1)/h_1 - h_1 u_1/3 - h_1 u_2/6, \\
u_1 \quad &= u_1, \\
\vdots \quad &= \vdots \\
u_{n+1} \quad &= u_{n+1}.
\end{aligned}$$

Note that if we decide that (p_1, p_2) is the fixed effect, then (p_1, q_1) is not a fixed effect, because from (8) the parameter q_1 is random with the mean $(p_2 - p_1)/h_1$ and the variance $h_1^2 \mathrm{var}(u_1/3 + u_2/6)$. Inversely, if (p_1, q_1) is the fixed effect, then (p_1, p_2) is not a fixed effect, because again from (8) the parameter p_2 is random with the mean $p_1 + h_1 q_1$ and the variance $h_1^4 \mathrm{var}(u_1/3 + u_2/6)$. We conclude that the parametrizations Θ_{002} and Θ_{012} have the same concentration $\sigma_s^{n+1}/det(\mathbf{Q}_2)$ in the sense of Definition 2.6, but are not equivalent in the sense of Definition 2.4.

3.1. Construction of the bases $\mathbf{B}_{0123}$*,* $\mathbf{B}_{012}$*,* $\mathbf{B}_{002}$

First, we construct the basis $\mathbf{B}_{0123} = (\boldsymbol{b}_1^{0123}, \ldots, \boldsymbol{b}_{n+3}^{0123})$ in which

$$\boldsymbol{s} = s(t_1)\boldsymbol{b}_1^{0123} + s'(t_1)\boldsymbol{b}_2^{0123} + s''(t_1)\boldsymbol{b}_3^{0123} + \sum_{i=1}^{n} s'''(t_i+)\boldsymbol{b}_{3+i}^{0123}, \tag{15}$$

for all cubic spline $\boldsymbol{s}$. By writing

$$s(t) = p_i + q_i(t - t_i) + (t - t_i)^2 u_i/2 + (t - t_i)^3 v_i/6$$

on each sub-interval $[t_i, t_{i+1})$, and using its smoothness, we obtain the following constraints

$$\begin{pmatrix} p_{i+1} \\ q_{i+1} \\ u_{i+1} \end{pmatrix} = \mathbf{M}_i \begin{pmatrix} p_i \\ q_i \\ u_i \end{pmatrix} + v_i \begin{pmatrix} h_i^3/6 \\ h_i^2/2 \\ h_i \end{pmatrix}, \tag{16}$$

where

$$\mathbf{M}_i = \begin{pmatrix} 1 & h_i & h_i^2/2 \\ 0 & 1 & h_i \\ 0 & 0 & 1 \end{pmatrix}.$$

We denote by $\mathbf{M}^i$ the product $\prod_{l=1}^{i} \mathbf{M}_l$. The equation (16) implies for $i = 1, \ldots, n$, that

$$\begin{pmatrix} p_{i+1} \\ q_{i+1} \\ u_{i+1} \end{pmatrix} = \mathbf{M}^i \begin{pmatrix} p_1 \\ q_1 \\ u_1 \end{pmatrix} + \sum_{l=1}^{i} v_l \mathbf{M}^{i-l} \begin{pmatrix} h_l^3/6 \\ h_l^2/2 \\ h_l \end{pmatrix}. \tag{17}$$

We will use the notation $\mathbf{M}_i^{-1} = 0$, and $\mathbf{M}_i^0 = \mathbf{I}_3$.

Let us define, for each $i = 1, \ldots, n$, the piecewise functions,

$$\chi_i(t) = 1_{[t_i,t_{i+1})}(t), \quad \chi_i^1(t) = (t - t_i)\chi_i(t), \quad \chi_i^2(t) = (t - t_i)^2\chi_i(t)/2,$$
$$\chi_i^3(t) = (t - t_i)^3\chi_i(t)/6.$$

Now, using (17), we can announce the basis $\mathbf{B}_{0123} = [\boldsymbol{b}_j^{0123} : j = 1, \ldots, n+3]$.

Proposition 3.1. *We have for $t \in [t_1, t_2)$, that*

$$\begin{aligned} \begin{pmatrix} b_1^{0123}(t) & b_2^{0123}(t) & b_3^{0123}(t) & b_4^{0123}(t) \end{pmatrix} &= \begin{pmatrix} \chi_1(t) & \chi_1^1(t) & \chi_1^2(t) & \chi_1^3(t) \end{pmatrix}, \\ b_l^{0123} &= 0, \quad l = 5, \ldots, n+3. \end{aligned}$$

We have for $i = 1, \ldots, n-1$, $t \in [t_{i+1}, t_{i+2})$, and $l = 1, \ldots, i$, that

$$\begin{aligned} \begin{pmatrix} b_1^{0123}(t) & b_2^{0123}(t) & b_3^{0123}(t) \end{pmatrix} &= \begin{pmatrix} \chi_{i+1}(t) & \chi_{i+1}^1(t) & \chi_{i+1}^2(t) \end{pmatrix} \mathbf{M}^i, \\ b_{3+l}^{0123}(t) &= \begin{pmatrix} \chi_{i+1}(t) & \chi_{i+1}^1(t) & \chi_{i+1}^2(t) \end{pmatrix} \mathbf{M}^{i-l} \begin{pmatrix} h_l^3/6 \\ h_l^2/2 \\ h_l \end{pmatrix}, \\ b_{i+4}^{0123}(t) &= \chi_{i+1}^3, \\ b_j^{0123}(t) &= 0, \quad j = i+5, \ldots, n+3. \end{aligned}$$

The basis $\mathbf{B}_{0123}$ is a separating basis because $(\boldsymbol{b}_1^{0123}, \boldsymbol{b}_2^{0123})$ is a basis of $Q^{-1}(0)$ (2). It is plotted in Figure 1. To compare with (13), (14), we observe that $b_1^{0123}(t) = 1$, $b_2^{0123}(t) = t$ and then

$$s(t_1)b_1^{0123}(t) + s'(t_1)b_2^{0123}(t) := \alpha + \beta t,$$

with

$$\alpha = s(t_1), \quad \beta = s'(t_1).$$

3.2. Construction of the bases $\mathbf{B}_{002}$ *and* $\mathbf{B}_{012}$

Now, we are ready to construct the basis $\mathbf{B}_{012} = [\boldsymbol{b}_1^{012}, \dots, \boldsymbol{b}_{3+n}^{012}]$. From the constraints $(u_{i+1} - u_i)/h_i = v_i, \quad i = 1, \dots, n+1$, we have

$$\begin{aligned} \boldsymbol{s} &= p_1\boldsymbol{b}_1^{0123} + q_1\boldsymbol{b}_2^{0123} + u_1\boldsymbol{b}_3^{0123} + \sum_{i=1}^{n} v_i\boldsymbol{b}_{3+i}^{0123} \\ &= p_1\boldsymbol{b}_1^{0123} + q_1\boldsymbol{b}_2^{0123} + u_1(\boldsymbol{b}_3^{0123} - \boldsymbol{b}_4^{0123}/h_1) + \sum_{i=2}^{n} u_i(\boldsymbol{b}_{2+i}^{0123}/h_{i-1} - \boldsymbol{b}_{3+i}^{0123}/h_i) + \\ &\quad u_{n+1}\boldsymbol{b}_{3+n}^{0123}/h_n. \end{aligned}$$

It follows that $\boldsymbol{b}_1^{012} = \boldsymbol{b}_1^{0123}$, $\boldsymbol{b}_2^{012} = \boldsymbol{b}_2^{0123}$, $\boldsymbol{b}_3^{012} = \boldsymbol{b}_3^{0123} - \boldsymbol{b}_4^{0123}/h_1$, for $i = 4, \dots, n+2$, $\boldsymbol{b}_i^{012} = \boldsymbol{b}_i^{0123}/h_{i-3} - \boldsymbol{b}_{i+1}^{0123}/h_{i+1-3}$, $\boldsymbol{b}_{n+3}^{012} = \boldsymbol{b}_{3+n}^{0123}/h_n$. The passage matrix $(\boldsymbol{b}_3^{0123}, \dots, \boldsymbol{b}_{n+3}^{0123})^\top = \mathbf{P}(\boldsymbol{b}_3^{012}, \dots, \boldsymbol{b}_{n+3}^{012})^\top$ is the inverse of the matrix having the rows $\boldsymbol{l}_1 = (1, -1/h_1, 0, \dots, 0)$, $\boldsymbol{l}_2 = (0, 1/h_1, -1/h_2, 0, \dots, 0)$, $\dots$, $\boldsymbol{l}_n = (0, \dots, 0, 1/h_{n-1}, 1/h_n)$, $\boldsymbol{l}_{n+1} = (0, \dots, 0, 1/h_n)$. It follows that the partition function

$$Z_{\sigma_s}(\boldsymbol{b}_3^{0123}, \dots, \boldsymbol{b}_{n+3}^{0123}) = (\prod_{i=1}^{n} h_i^2)\sigma_s^{n+1}/det(\mathbf{Q}_2). \tag{18}$$

We derive the basis $\mathbf{B}_{002}$ as follows. The relation

$$q_1 = (p_2 - p_1)/h_1 - h_1u_1/3 - h_1u_2/6,$$

and

$$\boldsymbol{s} = p_1\boldsymbol{b}_1^{002} + p_2\boldsymbol{b}_2^{002} + \sum_{i=1}^{n+1} u_i\boldsymbol{b}_{2+i}^{002}$$

imply that

$$\begin{aligned}\boldsymbol{b}_1^{002} &= \boldsymbol{b}_1^{012} - \boldsymbol{b}_2^{012}/h_1, \boldsymbol{b}_2^{002} = \boldsymbol{b}_2/h_1, \boldsymbol{b}_3^{002} = \boldsymbol{b}_3^{012} - h_1\boldsymbol{b}_2^{012}/3,\\ \boldsymbol{b}_4^{002} &= \boldsymbol{b}_4^{012} - h_1\boldsymbol{b}_2^{012}/6, \boldsymbol{b}_i^{002} = \boldsymbol{b}_i^{012}, \quad i = 5, \ldots, n+3.\end{aligned}$$

Observe that $\boldsymbol{b}_3^{002}$ and $\boldsymbol{b}_4^{002}$ are corrupted by $\boldsymbol{b}_2^{012}$. The bases $\mathbf{B}_{012}$, $\mathbf{B}_{0123}$, $\mathbf{B}_{002}$ are separating bases and are plotted in Figure 1. Moreover, according to (18) if $h_i = 1$ for all i, then the bases $\mathbf{B}_{002}$, $\mathbf{B}_{012}$, $\mathbf{B}_{0123}$ have the same concentration.

To compare with (13), (14), we observe that $b_1^{002}(t) = 1 - t/h_1$, $b_2^{002}(t) = t/h_1$ and then $s(t_1)b_1^{002}(t) + s(t_2)b_2^{002}(t) := \alpha + \beta t$, with $\alpha = s(t_1)$, $\quad \beta = \{s(t_2) - s(t_1)\}/h_1$.

3.3. Basis $\mathbf{B}_{s202}$

Here we are interested in the parametrization

$$\Theta_{202} = (s''(t_1), s(t_1), \ldots, s(t_{n+1}), s''(t_{n+1})).$$

The corresponding basis $\mathbf{B}_{202} = (\varphi_1, \ldots, \varphi_{n+3})$ was constructed in [7] and is plotted in Figure 1. The subscript notation 202 is justified by the fact that $s''(t_1) := s^{(2)}(t_1)$, $s(t_1) := s^{(0)}(t_1)$, $\ldots$, $s(t_{n+1}) := s^{(0)}(t_{n+1})$, $s''(t_{n+1}) := s^{(2)}(t_{n+1})$. A cubic spline $\boldsymbol{s}$ is written in the basis $\mathbf{B}_{202}$ as follows

$$\boldsymbol{s} = s''(t_1)\varphi_1 + \sum_{i=1}^{n+1} s(t_i)\varphi_{1+i} + s''(t_{n+1})\varphi_{n+3}.$$

The basis $\mathbf{B}_{202}$ does not contain any straight line (see Figure 1). It follows that $\mathbf{B}_{202}$ is not a separating basis. However we are going to derive from $\mathbf{B}_{202}$ an interesting separating basis denoted $\mathbf{B}_{s202}$.

We can show [7] that the quadratic form (2) equals

$$(s''(t_1), s(t_1), \ldots, s(t_{n+1}), s''(t_{n+1}))\mathbf{Q}_{202}(s''(t_1), s(t_1), \ldots, s(t_{n+1}), s''(t_{n+1}))^\top,$$

where

$$\mathbf{Q}_{202} := [\int_{t_1}^{t_{n+1}} \varphi_i''(t)\varphi_j''(t)dt : i, j = 1, \ldots, n+3].$$

The singular valued decomposition (SVD) of $\mathbf{Q}_{202} = \mathbf{O}\mathbf{D}(0, 0, c_1, \ldots, c_{n+1})\mathbf{O}^\top$ tells us that the diagonal matrix $\mathbf{D}(0, 0, c_1, \ldots, c_{n+1})$ is defined by the null

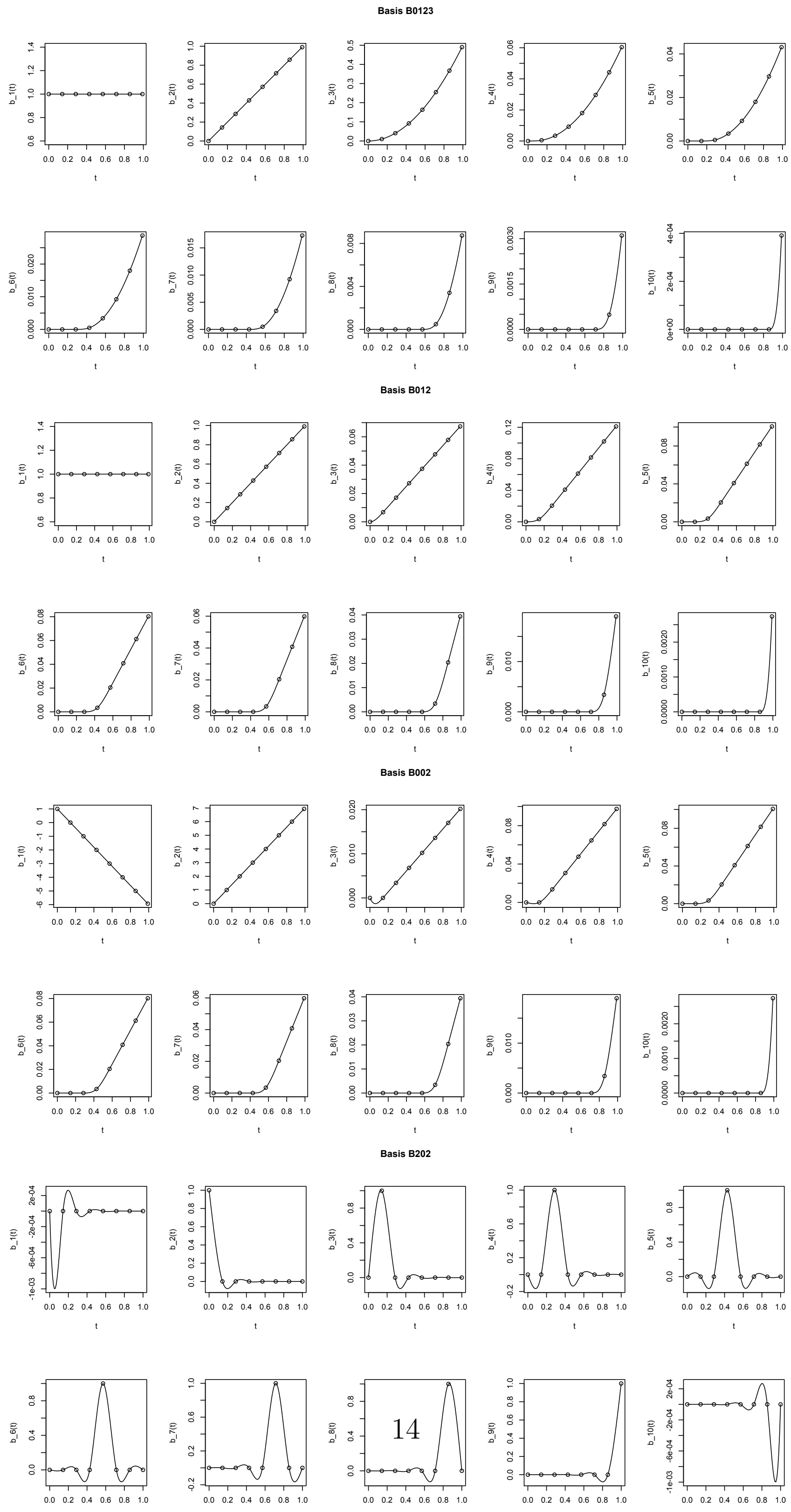

Figure 1: Representation of bases $\mathbf{B}_{012}$, $\mathbf{B}_{0123}$, $\mathbf{B}_{002}$ and $\mathbf{B}_{202}$ for n=7 equidistant intervals in $[0,1]$.

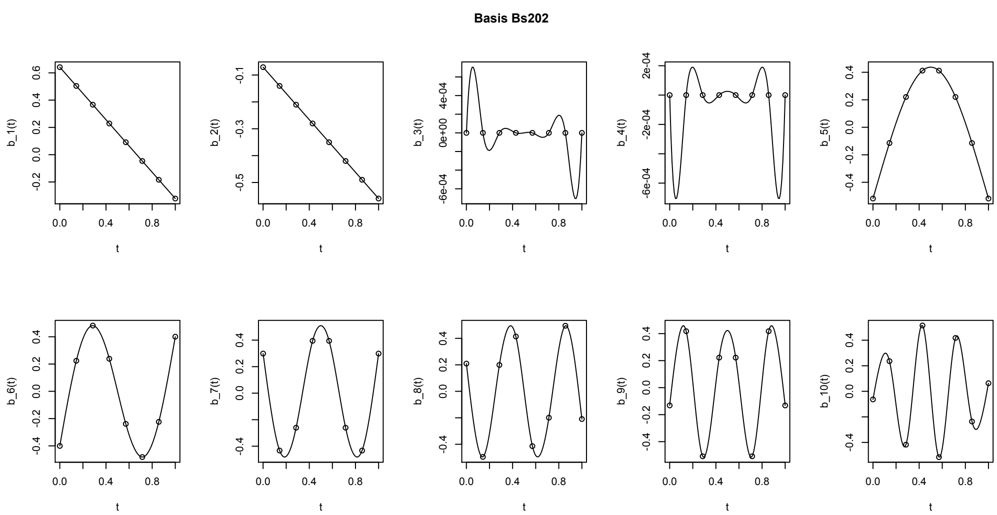

Figure 2: Representation of the basis $\mathbf{B}_{s202}$ for n=7 equidistant intervals in $[0,1]$.

eigenvalue of order 2 and the positive eigenvalues $c_1 \leq \ldots \leq c_{n+1}$ of $\mathbf{Q}_{202}$. Moreover the ith-column $\boldsymbol{o}_i$ of the orthogonal matrix $\mathbf{O}$ is such that $\mathbf{Q}_{202}\boldsymbol{o}_i = 0$ for $i = 1, 2$, and $\mathbf{Q}_{202}\boldsymbol{o}_{2+i} = c_i\boldsymbol{o}_{2+i}$ for $i = 1, \ldots, n+1$.

It follows for all column vector $\theta \in \mathbb{R}^{n+3}$ that

$$\begin{aligned} \theta^\top \mathbf{O}^\top \mathbf{Q}_{202}\mathbf{O}\theta &= \theta^\top \mathbf{D}(0, 0, c_1, \ldots, c_{n+1})\theta \\ &= \sum_{i=1}^{n+1} c_i \theta_{2+i}^2. \end{aligned}$$

The new parametrization

$$\theta = \mathbf{O}^\top(s''(t_1), s(t_1), \ldots, s(t_{n+1}), s''(t_{n+1}))^\top,$$

defines the new basis $\mathbf{B}_{s202} = \mathbf{O}\mathbf{B}_{202} := (\psi_1, \ldots, \psi_{n+3})$ plotted in Figure 2.

By construction ψ_1, ψ_2 is a basis of the null space $Q^{-1}(0)$ of the quadratic form (2), and then $\mathbf{B}_{s202}$ is a separating basis. More precisely, the new parametrization $(\theta_1, \ldots, \theta_{n+3})$ is defined by

$$\theta_1 = \sum_{j=1}^{n+1} o_{1j+1} s(t_j),\ \theta_2 = \sum_{j=1}^{n+1} o_{2j+1} s(t_j),$$

and for $i = 3, \ldots, n+3$, $\theta_i = o_{i1}s''(t_1) + \sum_{j=1}^{n+1} o_{ij+1}s(t_j) + o_{in+3}s''(t_{n+1})$. To compare with (13), (14), we observe that the fixed effect $\theta_1\psi_1 + \theta_2\psi_2 := \alpha + \beta t$ is such that α and β mix all the components $s(t_1)$, $\ldots$, $s(t_{n+1})$.

Making the assumption that $(s''(t_1), \ldots, s''(t_{n+1}))$ is centred, Gaussian with the covariance matrix $\sigma_s^2 \mathbf{Q}_2^{-1}$ (11) is equivalent to saying that $(\theta_3, \ldots, \theta_{n+3})$ is centred, Gaussian with the covariance matrix $\sigma_s^2 \mathbf{Q}_{s202}^{-1}$, where the diagonal matrix $\mathbf{Q}_{s202} = \mathbf{D}(c_1, \ldots, c_{n+1})$. It follows that the partition function of $(\theta_3, \ldots, \theta_{n+3})$ is equal to

$$Z_{\sigma_s}(\psi_3, \ldots, \psi_{n+3}) = \sigma_s^{n+1} \prod_{i=1}^{n+1} 1/c_i. \tag{19}$$

4. Maximum likelihood estimators of the fixed effect and the dispersion parameters

Let $\boldsymbol{s} = \sum_{i=1}^{2} \theta_i \boldsymbol{b}_i + \sum_{i=3}^{n+3} \theta_i \boldsymbol{b}_i$ be a random element of $S_3(t_1, \ldots, t_{n+1})$ in the basis $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n+3})$ separating the fixed effect $\boldsymbol{\beta} := (\theta_1, \theta_2)^\top$ from the random effect $\boldsymbol{r} := (\theta_3, \ldots, \theta_{n+3})^\top$. In the basis $\mathbf{B}$ the PDF of the random effect is centred, Gaussian with the covariance $\sigma_s^2 \mathbf{Q}_B^{-1}$, where

$$\mathbf{Q}_B = [\int_{t_1}^{t_{n+1}} b_i''(t) b_j''(t) dt : \quad i, j = 3, \ldots, n+3].$$

The matrix representation of (1) is the following additive mixed model

$$\boldsymbol{y} := (y_1, \ldots, y_{n+1})^\top = \mathbf{F}\boldsymbol{\beta} + \mathbf{R}\boldsymbol{r} + \boldsymbol{w}.$$

Here the column vectors of the matrices

$$\mathbf{F} = (\boldsymbol{b}_1(\boldsymbol{t}), \boldsymbol{b}_2(\boldsymbol{t})), \quad \mathbf{R} = (\boldsymbol{b}_3(\boldsymbol{t}), \ldots, \boldsymbol{b}_{n+3}(\boldsymbol{t})) \tag{20}$$

are $\boldsymbol{b}_i(\boldsymbol{t}) := (b_i(t_1), \ldots, b_i(t_{n+1}))^\top$.

If the noise $\boldsymbol{w}$ is white and Gaussian with the variance σ_w^2, then the data $\boldsymbol{y}$ is Gaussian with the mean $\mathbf{F}\boldsymbol{\beta}$ and the covariance matrix

$$\boldsymbol{\Sigma}(\sigma_s^2, \sigma_w^2) = \sigma_s^2 \mathbf{G}_0 + \sigma_w^2 \mathbf{G}_1.$$

Here

$$\mathbf{G}_0 = \mathbf{R}\mathbf{Q}_B^{-1}\mathbf{R}^\top, \quad \mathbf{G}_1 = \mathbf{I}_{n+1}.$$

The $-2\,\ln$-likelihood of the data $\boldsymbol{y}$ in the basis $\mathbf{B}$ is equal to

$$\begin{aligned}\ell(\boldsymbol{y}, \mathbf{B}) = {} & (n+1)\ln(2\pi) + \ln[\det\{\boldsymbol{\Sigma}(\sigma_s^2, \sigma_w^2)\}] \\ & + \text{trace}\{(\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta})(\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\sigma_s^2, \sigma_w^2)\}.\end{aligned} \tag{21}$$

The likelihood equations

$$\partial_{\beta_i}\ell = 0, \quad i = 1, 2, \quad \partial_{\sigma_w^2}\ell = 0, \quad \partial_{\sigma_s^2}\ell = 0,$$

are equivalent [1] to

$$\boldsymbol{\beta} = \{\mathbf{F}^\top \boldsymbol{\Sigma}^{-1}(\sigma_s^2, \sigma_w^2)\mathbf{F}\}^{-1}\mathbf{F}^\top \boldsymbol{\Sigma}^{-1}(\sigma_s^2, \sigma_w^2)\boldsymbol{y},$$

$$\begin{pmatrix} \text{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma_s^2, \sigma_w^2)\mathbf{G}_0^2\} & \text{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma_s^2, \sigma_w^2)\mathbf{G}_0\} \\ \text{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma_s^2, \sigma_w^2)\mathbf{G}_0\} & \text{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma_s^2, \sigma_w^2)\} \end{pmatrix} \begin{pmatrix} \sigma_s^2 \\ \sigma_w^2 \end{pmatrix} = \begin{pmatrix} \text{trace}\{(\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta})(\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-2}(\sigma_s^2, \sigma_w^2)G_0\} \\ \text{trace}\{(\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta})(\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-2}(\sigma_s^2, \sigma_w^2)\} \end{pmatrix}.$$

To solve the likelihood equations we use the following iterative scheme [1]. Starting from $\boldsymbol{\beta}(0) = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\boldsymbol{y}$, we obtain initial estimates $\sigma_s^2(0)$, $\sigma_w^2(0)$ by solving the system

$$\begin{pmatrix} \text{trace}(\mathbf{G}_0^2) & \text{trace}(\mathbf{G}_0) \\ \text{trace}(\mathbf{G}_0) & \text{trace}(\mathbf{G}_1) \end{pmatrix} \begin{pmatrix} \sigma_s^2(0) \\ \sigma_w^2(0) \end{pmatrix} = \begin{pmatrix} \text{trace}(\mathbf{C}(0)\mathbf{G}_0) \\ \text{trace}(\mathbf{C}(0)) \end{pmatrix},$$

where $\mathbf{C}(0) = \{\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta}(0)\}\{\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta}(0)\}^\top$. Having $\sigma_s^2(0)$, $\sigma_w^2(0)$, we construct

$$\boldsymbol{\beta}(1) = \{\mathbf{F}^\top\boldsymbol{\Sigma}^{-1}(\sigma_s^2(0), \sigma_w^2(0))\mathbf{F}\}^{-1}\mathbf{F}^\top\boldsymbol{\Sigma}^{-1}(\sigma_s^2(0), \sigma_w^2(0))\boldsymbol{y}.$$

Having $\boldsymbol{\beta}(1)$, we construct $\sigma_s^2(1)$, $\sigma_w^2(1)$ by solving the system

$$\begin{pmatrix} \text{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma_s^2(0), \sigma_w^2(0))\mathbf{G}_0^2\} & \text{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma_s^2(0), \sigma_w^2(0))\mathbf{G}_0\} \\ \text{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma_s^2(0), \sigma_w^2(0))\mathbf{G}_0\} & \text{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma_s^2(0), \sigma_w^2(0))\} \end{pmatrix} \begin{pmatrix} \sigma_s^2(1) \\ \sigma_w^2(1) \end{pmatrix} = \begin{pmatrix} \text{trace}\{\mathbf{C}(1)\boldsymbol{\Sigma}^{-2}(\sigma_s^2(0), \sigma_w^2(0))G_0\} \\ \text{trace}\{\mathbf{C}(1)\boldsymbol{\Sigma}^{-2}(\sigma_s^2(0), \sigma_w^2(0))\} \end{pmatrix},$$

where $\mathbf{C}(1) = \{\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta}(1)\}\{\boldsymbol{y} - \mathbf{F}\boldsymbol{\beta}(1)\}^\top$. If the number $n+1$ of observations $\boldsymbol{y}$ is large then this iterative scheme converges.

If the number $n+1$ is small, then we need N large i.i.d. copies $\boldsymbol{y}(1), \ldots, \boldsymbol{y}(N)$ of $\boldsymbol{y}$. In this case we only need the following one-step algorithm. Calculate OLSE

$$\boldsymbol{\beta}_N(0) = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\bar{\boldsymbol{y}},$$

and obtain initial estimates of $\sigma_{s,N}^2(0)$, $\sigma_{w,N}^2(0)$ by solving the system

$$\begin{pmatrix} \text{trace}(\mathbf{G}_0^2) & \text{trace}(\mathbf{G}_0) \\ \text{trace}(\mathbf{G}_0) & \text{trace}(\mathbf{G}_1) \end{pmatrix} \begin{pmatrix} \sigma_s^2(0) \\ \sigma_w^2(0) \end{pmatrix} = \begin{pmatrix} \text{trace}(\mathbf{C}_N(0)\mathbf{G}_0) \\ \text{trace}(\mathbf{C}_N(0)) \end{pmatrix},$$

Here $\bar{\boldsymbol{y}} = \sum_{k=1}^{N} \boldsymbol{y}(k)/N$ and $\mathbf{C}_N(0) = \sum_{k=1}^{N}(\boldsymbol{y}(k) - \mathbf{F}\boldsymbol{\beta}_N(0))(\boldsymbol{y}(k) - \mathbf{F}\boldsymbol{\beta}_N(0))^\top/N$. Having $\sigma^2_{s,N}(0)$, $\sigma^2_{w,N}(0)$, we construct

$$\boldsymbol{\beta}_N(1) = \{\mathbf{F}^\top\boldsymbol{\Sigma}^{-1}(\sigma^2_{s,N}(0), \sigma^2_{w,N}(0))\mathbf{F}\}^{-1}\mathbf{F}^\top\boldsymbol{\Sigma}^{-1}(\sigma^2_{s,N}(0), \sigma^2_{w,N}(0))\bar{\boldsymbol{y}}.$$

Having $\boldsymbol{\beta}_N(1)$, we construct $\sigma^2_{s,N}(1)$ and $\sigma^2_{w,N}(1)$ by solving the system

$$\begin{pmatrix} \mathrm{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma^2_{s,N}(0), \sigma^2_{w,N}(0))\mathbf{G}_0^2\} & \mathrm{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma^2_{s,N}(0), \sigma^2_{w,N}(0))\mathbf{G}_0\} \\ \mathrm{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma^2_{s,N}(0), \sigma^2_{w,N}(0))G_0\} & \mathrm{trace}\{\boldsymbol{\Sigma}^{-2}(\sigma^2_{s,N}(0), \sigma^2_{w,N}(0))\} \end{pmatrix} \begin{pmatrix} \sigma^2_s(1) \\ \sigma^2_w(1) \end{pmatrix}$$
$$= \begin{pmatrix} \mathrm{trace}\{\mathbf{C}_N(1)\Sigma^{-2}(\sigma^2_{s,N}(0), \sigma^2_{w,N}(0))G_0\} \\ \mathrm{trace}\{\mathbf{C}_N(1)\boldsymbol{\Sigma}^{-2}(\sigma^2_{s,N}(0), \sigma^2_{w,N}(0))\} \end{pmatrix},$$

where $\mathbf{C}_N(1) = \sum_{k=1}^{N}(\boldsymbol{y}(k) - \mathbf{F}\boldsymbol{\beta}_N(1))(\boldsymbol{y}(k) - \mathbf{F}\boldsymbol{\beta}_N(1))^\top/N$. It is known [1] that $\sqrt{N}(\boldsymbol{\beta}_N(1) - \boldsymbol{\beta})$, $\sqrt{N}(\sigma^2_{s,N}(1) - \sigma^2_s, \sigma^2_{w,N}(1) - \sigma^2_w)^T$ have a limiting normal distribution with means 0 and covariance matrices $\{\mathbf{F}^\top\boldsymbol{\Sigma}^{-1}(\sigma^2_s, \sigma^2_w)\mathbf{F}\}^{-1}$, $[\frac{1}{2}\mathrm{trace}\{\boldsymbol{\Sigma}^{-1}(\sigma^2_s, \sigma^2_w)\mathbf{G}_i\boldsymbol{\Sigma}^{-1}(\sigma^2_s, \sigma^2_w)\mathbf{G}_j\} : i, j = 0, 1]^{-1}$ respectively.

Finally having the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2_w$, $\hat{\sigma}^2_s$ of $\boldsymbol{\beta}$, σ^2_w, σ^2_s, the random effect estimate

$$\begin{aligned} \hat{\boldsymbol{r}} &= \arg\min_{\boldsymbol{r}}\{\hat{\lambda}\boldsymbol{r}^\top\mathbf{G}_0\boldsymbol{r} + \|\boldsymbol{y} - (\mathbf{F}\hat{\boldsymbol{\beta}} + \mathbf{R}\boldsymbol{r})\|^2\} \\ &= (\hat{\lambda}\mathbf{G}_0 + \mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top(\boldsymbol{y} - \mathbf{F}\hat{\boldsymbol{\beta}}), \end{aligned}$$

where $\hat{\lambda} := \frac{\hat{\sigma}^2_s}{\hat{\sigma}^2_w}$, and $\|\cdot\|$ denotes the Euclidean norm.

Remark 4.1. *The fixed effect in the bases $\mathbf{B}_{002}$ and $\mathbf{B}_{012}$ does not coincide with the linear regression. It only reflects the linear tendency of the beginning of the period. In the basis $\mathbf{B}_{s202}$ the fixed effect concerns all the period. In the basis $\mathbf{B}_{s202}$ (and unlike the bases $\mathbf{B}_{002}$ and $\mathbf{B}_{012}$) the columns of the matrix $\mathbf{F}$ and the columns of the matrix $\mathbf{R}$ (20) are orthogonal. We can show theoretically that this fact implies that the maximum-likelihood estimator $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}(i) = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\boldsymbol{y}$ for each iteration i. It follows that the fixed effect $\mathbf{F}\hat{\boldsymbol{\beta}}$ coincides with the linear regression. See [4], [6] for a similar study.*

5. Numerical applications

5.1. Simulation: inference from the bases $\mathbf{B}_{002}$, $\mathbf{B}_{012}$

In order to illustrate the importance of the parametrization for the estimation of the true fixed effect, we consider the parametrization $\mathbf{B}_{002}$ and

$\mathbf{B}_{012}$. We construct the spline

$$\begin{aligned}\boldsymbol{s}_0 &= p_1(true)\boldsymbol{b}_1^{002} + p_2(true)\boldsymbol{b}_2^{002} + \sum_{i=1}^{n+1} u_i(true)\boldsymbol{b}_{2+i}^{002} \\ &= p_1(true)\boldsymbol{b}_1^{012} + q_1(true)\boldsymbol{b}_2^{012} + \sum_{i=1}^{n+1} u_i(true)\boldsymbol{b}_{2+i}^{012}\end{aligned}$$

by fixing $p_1(true)$, $q_1(true)$ (fixed effect) and $u_1(true), \ldots, u_{n+1}(true)$ is a realization of the random effect $u_1, \ldots, u_{n+1}$. We obtain the true value

$$p_2(true) = p_1(true) + h_1 q_1(true) + h_1^2 u_1(true)/3 + h_1^3 u_2(true)/6.$$

We generate a sample $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$ of $\boldsymbol{s}_0$, and a sample $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$ of the noise $\boldsymbol{w}$. We obtain a sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ of the data $\boldsymbol{y}$.

Under the $\mathbf{B}_{002}$ parametrization and assuming that p_1, p_2 are the fixed effect, we execute the scoring algorithm and obtain the estimator p_1^{002}, p_2^{002}, $u_1^{002}, \ldots, u_{n+1}^{002}$ of $p_1(true)$, $p_2(true)$, $u_1(true), \ldots, u_{n+1}(true)$. We derive

$$q_1^{002} = (p_2^{002} - p_1^{002})/h_1 - h_1 u_1^{002}/3 - h_1 u_2^{002}/6$$

as an estimator of $q_1(true)$. Now we estimate $p_1(true)$, $q_1(true)$, $u_1(true), \ldots, u_{n+1}(true)$ in the basis $\mathbf{B}_{012}$ and assume that p_1, q_1 are the fixed effect. From the estimator p_1^{012}, q_1^{012}, $u_1^{012}, \ldots, u_{n+1}^{012}$, we obtain

$$p_2^{012} := p_1^{012} + h_1 q_1^{012} + \frac{h_1^2}{3} u_1^{012} + \frac{h_1^3}{6} u_2^{012}$$

as a new estimator of $p_2(true)$.

Table 1 presents the differences between the estimations of the parameters p_1, p_2, q_1, σ_s^2 and σ_w^2 when the true model is the basis $\mathbf{B}_{012}$ and $\mathbf{B}_{002}$ while varying the signal-to-noise ratio $\lambda = \sigma_w^2/\sigma_s^2$.
The results presented in Table 1 show that, for small values of λ, a bad choice of model produces large estimation errors. In our example, for $\lambda = 1/50$ the estimation of $q_1 = 1$ is $q_1^{012} = 0.9775$ in the true model $\mathbf{B}_{012}$, whereas $q_1^{002} = 0.4419$ in the wrong model $\mathbf{B}_{002}$. According to our simulations, these errors decrease with λ (see the results in Table 1 for $\lambda = 1/5$, and $\lambda = 1$).

Parameter θ	p_1	p_2	q_1	σ_s^2	σ_w^2	λ
True values	0	0.056	1	5	0.1	1/50
$\hat{\theta}^{012}$	0.0021	0.1110	0.9775	5.0213	0.0954	1/52.9100
$\hat{\theta}^{002}$	0.0022	0.056	0.4419	5.1343	0.0955	1/53.7634
True values	0	0.172	1	0.5	0.1	1/5
$\hat{\theta}^{012}$	0.0107	0.1566	0.9742	0.4721	0.1025	1/4.6058
$\hat{\theta}^{002}$	0.0101	0.1867	1.1731	0.4578	0.1024	1/4.4707
True values	0	0.1521	1	0.1	0.1	1
$\hat{\theta}^{012}$	-0.0012	0.1373	1.0109	0.1042	0.0998	1/1.0440
$\hat{\theta}^{002}$	-0.0014	0.1517	1.0823	0.1025	0.0912	1/1.12

Table 1: Estimation using $\mathbf{B}_{012}$ and $\mathbf{B}_{002}$ models.

5.2. Real data application: climat change

In the climate change problem we are interested in the annual mean temperature observed in France form 1900 to 2015. For each year, Meteo France (Division des Études et Climatologie, Nord) provided us with the minimum, mean and maximum temperatures in France. In our application we are interested only in the mean temperature. Data is presented in Figure 3.

We illustrate the importance of the parametrization of the fixed effects by considering the bases $\mathbf{B}_{002}$ and $\mathbf{B}_{s202}$. The estimation of the parameters of the two models and the $-2\ln(Likelihood)$ values are presented in Table 2.

Basis	$\hat{\theta}_1$	$\hat{\theta}_2$	σ_s^2	σ_w^2	$-2\ln(likelihood)$
$\mathbf{B}_{002}$	11.486	11.491	1.577×10^{-5}	0.232	172.552
$\mathbf{B}_{s202}$	-114.968	-58.545	2.755×10^{-5}	0.2283	165.495

Table 2: Estimation using $\mathbf{B}_{002}$ and $\mathbf{B}_{s202}$ models for annual mean temperature.

For each model, the fixed and the random effects are plotted in Figure 4. The difference between the two models is illustrated in Figure 5.

The numerical results (Figure 5) confirm the theoretical ones, namely the Remark 4.1. Note that the performance of the basis $\mathbf{B}_{s202}$ is superior to that

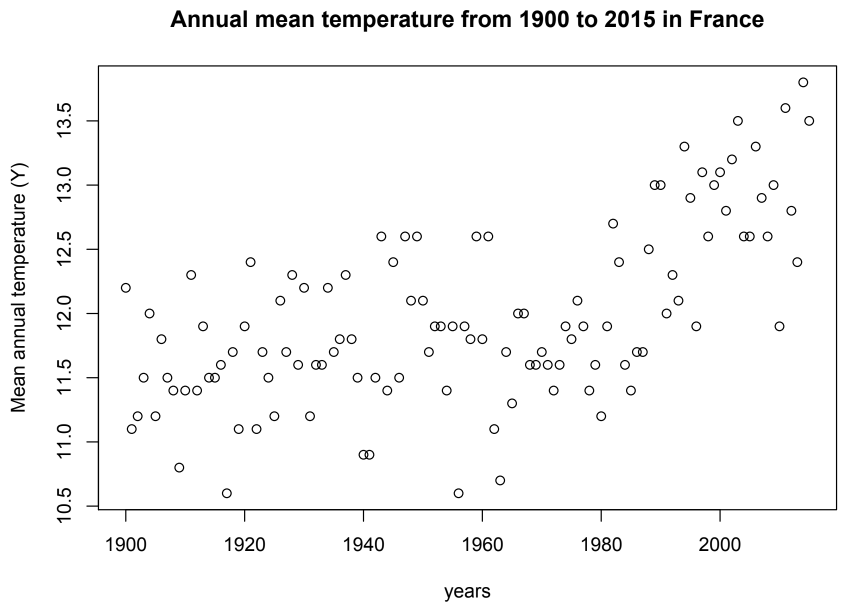

Figure 3: Annual mean temperatures in France from 1900 to 2015.

of $\mathbf{B}_{002}$ in terms of likelihood. As expected, from Figure 5 we see that the fixed effects in the two models are very different, whereas the estimates of the true signal are very close. In the basis $\mathbf{B}_{002}$ the random effect compensates and corrects the fixed effect defined only by the first two observations (years 1900 and 1901). The numerical results also show that the basis $\mathbf{B}_{s202}$ is more concentrated than $\mathbf{B}_{002}$ $(\prod_{i=1}^{116} 1/c_i \approx 1/27 \ll 1/det(\mathbf{Q}_2))$.

Appendix: Inference from fixed and random effects in a general setting

Let $\boldsymbol{s} = \sum_{i=1}^{k} \theta_i \boldsymbol{b}_i + \sum_{i=k+1}^{p} \theta_i \boldsymbol{b}_i$ be a random element of S in the basis $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p)$ separating the fixed effect $\theta_1, \ldots, \theta_k$ from the random effect $\theta_{k+1}, \ldots, \theta_p$. The PDF of $\theta_{k+1}, \ldots, \theta_p$ is equal to

$$f_{\sigma_s^2}(\theta_{k+1}, \ldots, \theta_p) := \frac{1}{Z_s} \exp\{-q(\sum_{i=k+1}^{p} \theta_i \boldsymbol{b}_i)/2\sigma_s^2\}.$$

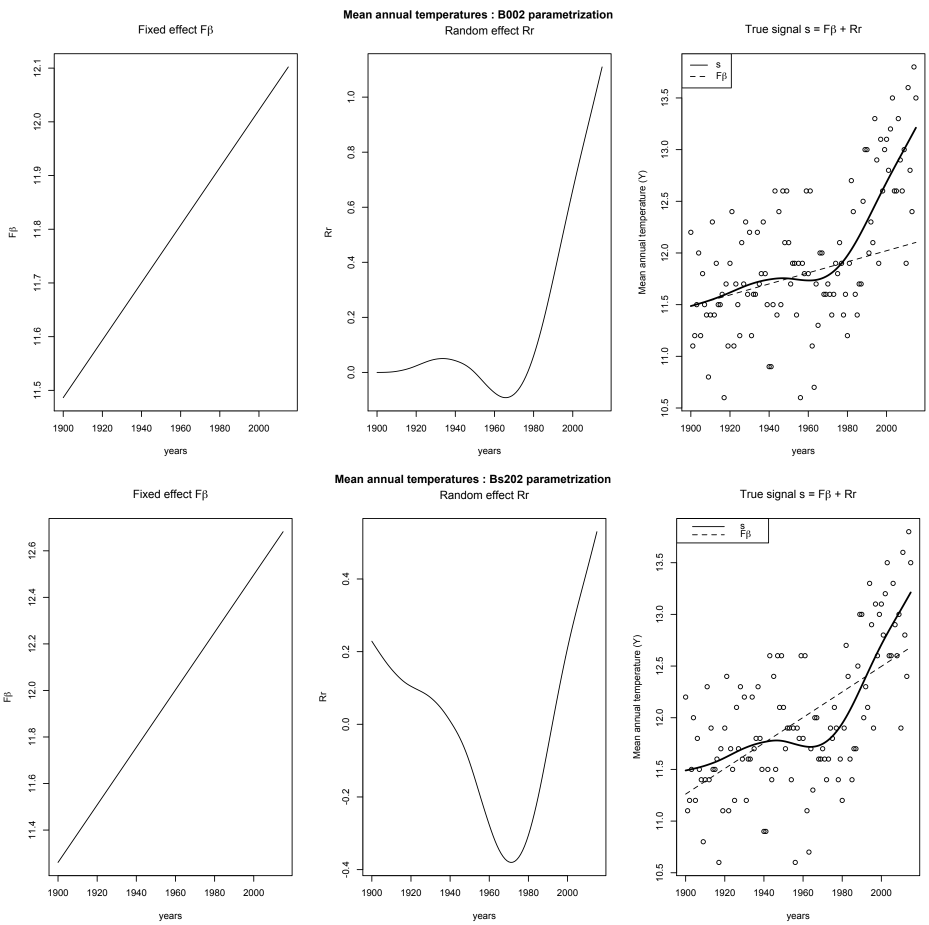

Figure 4: Mean annual temperatures in France from 1900 to 2015.

Let $\mathbf{X} : S \to \mathbb{R}^m$ be a given linear map, and $\boldsymbol{w} \in \mathbb{R}^m$ a random vector with the PDF

$$\frac{1}{Z_w} \exp\{-G(\boldsymbol{w})/2\sigma_w^2\},$$

where G is a measurable map. In the basis $\mathbf{B}$, the additive model $\boldsymbol{y} = \mathbf{X}\boldsymbol{s}+\boldsymbol{w}$ becomes

$$\boldsymbol{y} = \sum_{i=1}^{k} \theta_i \mathbf{X}\boldsymbol{b}_i + \sum_{i=k+1}^{p} \theta_i \mathbf{X}\boldsymbol{b}_i + \boldsymbol{w}. \tag{22}$$

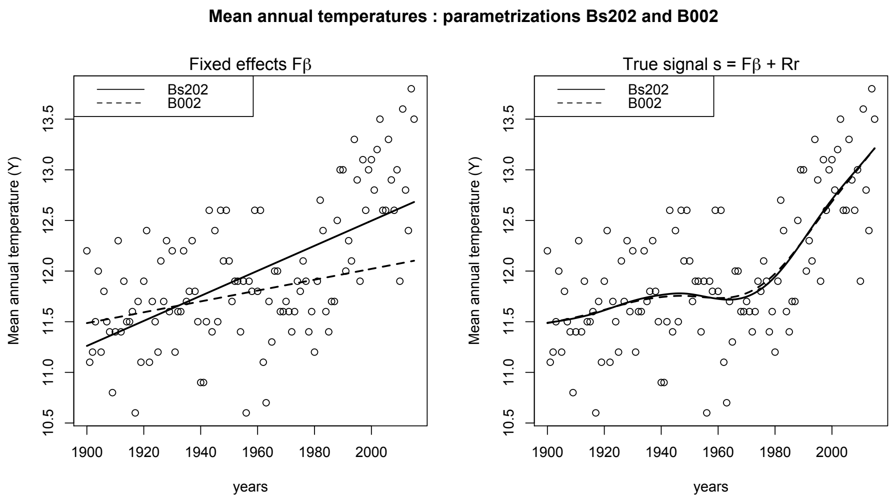

Figure 5: $\mathbf{B}_{002}$ versus $\mathbf{B}_{s202}$ estimators

The mixed model (22) consists of three types of objects: observable random vector $\boldsymbol{y}$ (data), unobservable random vectors $\theta_{k+1}, \ldots, \theta_p$, $\boldsymbol{w}$ and unknown parameters $\tau := (\theta_1, \ldots, \theta_k, \sigma_s^2, \sigma_w^2)$. In general, the probability density function (PDF) of a random vector $\boldsymbol{z}$ is denoted by $f_\tau(\boldsymbol{y})$. Here τ is a fixed parameter and $\boldsymbol{y}$ varies. The likelihood of τ given the data $\boldsymbol{y}$ is denoted by $L(\tau \,|\, \boldsymbol{y})$. The connection between the (PDF) and the likelihood is given by

$$L(\tau \,|\, \boldsymbol{y}) = f_\tau(\boldsymbol{y}).$$

It follows that

$$f_\tau(\boldsymbol{y}) = \int_{\mathbf{R}^{p-k}} f_{(\theta_1,\ldots,\theta_p,\sigma_w^2)}(\boldsymbol{y}) f_{\sigma_s^2}(\theta_{k+1}, \ldots, \theta_p) d\theta_{k+1} \ldots d\theta_p,$$

where

$$f_{(\theta_1,\ldots,\theta_p,\sigma_w^2)}(\boldsymbol{y}) = \frac{1}{Z_w} \exp\{-G(\boldsymbol{y} - \mathbf{X}\boldsymbol{s})/2\sigma_w^2\}$$

is the (PDF) of $\boldsymbol{y}$ known $\boldsymbol{s}$ and σ_w^2.

Parameter estimation: Given the data $\boldsymbol{y}$, we estimate τ by maximizing the likelihood $\tau \to L(\tau \,|\, \boldsymbol{y})$.

The joint (PDF) of the random effect $(\theta_{k+1}, \ldots, \theta_p)$ and $\boldsymbol{y}$ is equal to

$$\begin{aligned} f_\tau(\theta_{k+1}, \ldots, \theta_p, \boldsymbol{y}) &= f_{\sigma_s^2}(\theta_{k+1}, \ldots, \theta_p) f_{(\theta_1,\ldots,\theta_k,\sigma_w^2)}(\boldsymbol{y}) \\ &= L(\tau \,|\, \theta_{k+1}, \ldots, \theta_p, \boldsymbol{y}). \end{aligned}$$

These were called a joint likelihood by [10] in the context of Gaussian case and hierarchical likelihoods by [13].

Random effect estimation. Given the data $\boldsymbol{y}$ and an estimation $\hat{\tau}$ of τ, we estimate the random effect by maximizing the joint likelihood of the random effect $(\theta_{k+1}, \dots, \theta_p)$ and $\boldsymbol{y}$ given $\boldsymbol{y}$ and $\tau = \hat{\tau}$.

$$L(\hat{\tau} \mid \theta_{k+1}, \dots, \theta_p, \boldsymbol{y}).$$

6. Acknowledgements

The research project is supported by Laboratoire Paul Painlevé USTL-UMR-CNRS 8524, the Labex CEMPI ANR-11-LABEX-0007-01 and the project ClinMine - ANR-13-TECS-0009. The authors would like to thank the Editor-in-Chief and the anonymous referees for carefully reading the manuscript and giving many helpful and constructive suggestions resulting in the present version. The authors would also like to thank Meteo France Division des Études et Climatologie Nord for producing the French temperature dataset.

References

[1] T. W. Anderson, Asymptotically efficient estimation of covariance matrices with linear structure, The Annals of Statistics, Vol. 1 No. 1 (1973) 135–141.

[2] P. Craven, G. Wahba, Smoothing noisy data with spline functions, Numer. Math. 31 (4) (1978) 377–390.

[3] C. deBoor, A Practical Guide to Spline, Springer (1978) Springer.

[4] A. Dermoune, B. Djehiche, N. Rahmania, A consistent estimator of the smoothing parameter in the Hodrick-Prescott filter J. Japan Statist. Soc., 38 (2) (2008) 225–241.

[5] A. Dermoune, B. Djehiche, N. Rahmania, Multivariate Extension of the Hodrick-Prescott Filter- Optimality and Characterization, Studies in NonLinear dynamics and Econometrics, 13 (3) (2009) 4–4.

[6] A. Dermoune, N. Rahmania, T. Wei, General Linear mixed model and signal extraction problem with constraint, Journal of Multivariate Analysis, 105 (1) (2012) 311–321.

[7] A. Dermoune, C. Preda, Estimation of noisy cubic spline using a natural basis, Annals of the University of Craiova, Mathematics and Computer Science Series 43 (1) (2016) 33–52.

[8] P. Diggle, P. Heagerty, K Liang, S. Zeger,. Analysis of longitudinal data, Oxford University Press (2002).

[9] P. J. Green, B.W. Silverman, Nonparametric Regression and Generalized linear models, Chapman and Hall, London (1994).

[10] C.R. Henderson, Best linear unbiased estimation and prediction under a selection model, Biometrics 31 (1975) 423–447.

[11] R.C.S. Lai, H.C. Huang, T.C.M. Lee, Fixed and random effects selection in nonparametric additive mixed models, Electronic Journal of Statistics 6 (2012) 810–842.

[12] N. Laird, J. Ware, Random-effects models for longitudinal data, Biometrics 38 (1982) 963–974.

[13] Y. Lee, J. A. Nelder, Likelihood for random-effect models, SORT 29 (2) (2005) 141–164.

[14] K. Liang, S. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.

[15] G. Wahba, Spline models for observational data. Society for Industrial and Applied Mathematics (SIAM) (1990).

[16] S. Zeger, K. Liang, Longitudinal data analysis for discrete and continuous outcomes, Biometrics 42 (1986) 121–130.