

Parameter estimation and inference in dynamic systems described by linear partial differential equations

Gianluca Frasso¹ · Jonathan Jaeger² ·
Philippe Lambert^{1,3}

Received: 13 September 2014 / Accepted: 16 September 2015 / Published online: 12 October 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Differential equations (DEs) are commonly used to describe dynamic systems evolving in one (ordinary differential equations or ODEs) or in more than one dimensions (partial differential equations or PDEs). In real data applications, the parameters involved in the DE models are usually unknown and need to be estimated from the available measurements together with the state function. In this paper, we present frequentist and Bayesian approaches for the joint estimation of the parameters and of the state functions involved in linear PDEs. We also propose two strategies to include state (initial and/or boundary) conditions in the estimation procedure. We evaluate the performances of the proposed strategy through simulated examples and a real data analysis involving (known and necessary) state conditions.

Keywords Linear partial differential equations · Parameter estimation · Penalized tensor B-spline smoothing · State conditions

1 Introduction

Dynamic systems are commonly described by differential equations (DEs). In real data applications, the parameters involved in the DE models are usually unknown and need to be estimated from the available measurements together with the state functions. For

✉ Gianluca Frasso
Gianluca.Frasso@ulg.ac.be

¹ Faculté des Sciences Sociales, Méthodes Quantitatives en Sciences Sociales, Université de Liège, B-4000 Liège, Belgium

² Nestlé Research Center, Lausanne, Switzerland

³ Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Louvain, Belgium

one-dimensional applications, this estimation task has been extensively discussed in the statistical literature. The most popular procedures rely on nonlinear least squares (Biegler et al. 1986). These approaches are computationally intensive and often poorly suited for statistical inference. An attractive alternative is represented by the penalized smoothing framework introduced by Ramsay et al. (2007). This approach can be viewed as a generalization of the L-spline theory (see, for example, Schultz and Varga (1967) among others) where the flexibility acquired by a high-dimensional B-spline basis expansion of the state function is counterbalanced by a penalty term expressed using the (set of) ODE(s) that needs to be solved. The fidelity of the final fit to the hypothesized differential model is then tuned using an ODE-compliance parameter. Jaeger and Lambert (2013) adapt the latter approach to a full Bayesian framework. This Bayesian alternative offers two major advantages with respect to the frequentist one: the selection of the ODE-compliance parameter becomes automatic and uncertainty measures about the parameters can readily be obtained using MCMC.

When dealing with dynamics evolving in more than one dimension, partial differential equations (or PDEs) are usually invoked. Inference in models specified by PDEs has not received a large attention in the literature. Here, we discuss statistical approaches to deal with dynamics described by linear PDEs. Such equations are used to model large classes of phenomena (e.g., diffusion, heat transfer, price of financial instruments, etc.). The general form of a deterministic PDE system can be written as:

$$\mathcal{F}\left(x_1, \dots, x_p, u, \frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_p}, \frac{\partial^2 u}{\partial x_1^2}, \dots, \frac{\partial^2 u}{\partial x_1 \partial x_p}, \dots, \boldsymbol{\theta}\right) = 0, \quad (1)$$

where $u(\mathbf{x})$ is the state function evaluated in $\mathbf{x} = (x_1, \dots, x_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\theta}$ is the vector of PDE parameters. In this paper, we further assume that \mathcal{F} is a linear function in u and its derivatives. Examples of this kind of PDE can be found in Eqs. (14), (15). In particular, we are interested in estimating the vector of unknown $\boldsymbol{\theta}$ using a set of measurements $\boldsymbol{\zeta} = \boldsymbol{\zeta}(\mathbf{x}) = u(\mathbf{x}) + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon}$ can be considered as a vector of independent normally distributed measurement errors.

Xun et al. (2013) have proposed frequentist and Bayesian approaches for the estimation of parameters in model driven by PDEs. The frequentist approach could be viewed as a multidimensional generalization of the framework introduced by Ramsay et al. (2007). The state function solving the differential problem is approximated using tensor product of B-splines. The flexibility of the approximation is then counterbalanced by a penalty related to the PDE. In the frequentist proposal, the PDE penalty is represented by the integral of the PDE operator evaluated at the B-spline approximation to the solution. In their Bayesian approach, they combine a PDE-based penalty with one defined using finite differences [i.e., through a roughness penalty in the spirit of Eilers and Marx (1996)]. In this paper, we present both frequentist and Bayesian frameworks for the joint estimation of the parameters and the state function involved in PDEs. Our methods generalize those proposed in one-dimensional applications by exploiting a PDE-based penalized tensor product B-spline smoothing approach.

B-splines belong to the class of sieve estimators (Grenander 1981; Shen 1997), i.e., they approximate an infinite dimensional parameter space with a finite surrogate. Many

sieves are available for function approximation. Our choice in favor of tensor product B-spline bases can be justified by different arguments. First, they are a fundamental tool in the P-spline framework (Eilers and Marx 1996, 2003, 2010) to which we refer in this paper. Indeed, analogously to P-spline smoothing approaches, we approximate (e.g., smooth) the observed multidimensional dynamics using a rich set of bases functions and shrink the estimated coefficients through a suitable penalty term. The PDE-based shrinkage term we adopt in what follows frames into a B-spline collocation scheme for the (numerical) solution of partial differential equations (see e.g., Golub and Ortega 1992). Within the available bases functions for the approximation of DE solutions, B-splines are particularly convenient. Indeed, their compact supports lead to sparse collocation matrices that can be treated by fast matrix algorithms and allow for efficient storage strategies (see e.g., Botella 2002). Furthermore, their piecewise polynomial nature avoids numerical interpolation distortion effects (such as Runge's phenomena, see, for example, Epperson (1987) for a detailed discussion on this aspect).

The choice of the optimal number of spline knots and their location has been extensively discussed in the literature and both frequentist and Bayesian selection approaches have been proposed. Examples include the stepwise selection method of Friedman and Silverman (1989) or the Bayesian procedures introduced by Denison et al. (1998) and Biller (2000). The knot selection process is usually computationally demanding and some alternatives are available. For example, within the ODE-based penalized smoothing framework (Ramsay et al. 2007), it is common to locate a knot at each data point. However, for multidimensional analyses, this strategy becomes easily impractical given the tensor product nature of the bases functions involved in the estimation process and the amount of available measurements. For this reason, following the suggestion by Eilers and Marx (2010) within the general P-spline framework and by Xun et al. (2013) for PDE-based smoothing, we prefer to define the spline basis over a generous number of knots at fixed locations. This a priori spline knot selection can eventually be guided by prior knowledge about the nature of the phenomenon under study (such an example is discussed in Sect. 5).

Finally, when the dimensionality of the estimation problem increases (e.g., by dealing with PDEs defined over more than two dimensions), tensor product B-splines can be computationally suboptimal if compared with other alternatives such as radial basis functions. Radial bases are defined by kernels evaluated exclusively over the observed domain. This allows for efficient surface smoothing algorithms (see e.g., Holmes and Mallick 1998; Smith et al. 2002). On the other hand, few classes of compact kernel functions are available (see e.g., Wu 1995) and they require the definition of a "compactness" parameter to be selected by some optimality criterion (e.g., AIC, BIC or cross-validation). This selection step can be computationally expensive. For this reason, we recommend to use tensor product B-splines when dealing with differential problems involving up to three dimensions.

This paper is organized as follows. In Sect. 2, we introduce our estimation procedures. Both frequentist and Bayesian approaches can be adapted to include the state (initial, boundary or mixed) conditions in the estimation process (see Sect. 3). In Sect. 4, we evaluate the performance of our approach and compare it with competitors using simulations. In Sect. 5, we illustrate our proposals dealing with a real data analysis based on the Black and Scholes (1973) model. Our application represents one

of those situations in which the state condition of the differential model (namely the no-arbitrage conditions) is known and need to be satisfied to obtain reasonable results. We conclude the paper with a discussion in Sect. 6.

2 Estimation of the unknown differential parameters for unknown state conditions

In this section, we introduce the PDE-based penalized tensor product B-spline smoothing approach both in frequentist and Bayesian frameworks without taking the state conditions (initial and/or boundary) into account.

2.1 Frequentist approach

Assume that one observes $\{(\boldsymbol{\zeta}; \mathbf{x}) = (\zeta_n; x_{1,n}, \dots, x_{p,n}); n = 1, \dots, N\}$ describing the evolution of the state function u whose dynamics is driven by (1). Our task is to jointly estimate the state function u and the vector of parameters $\boldsymbol{\theta}$. Following Ramsay et al. (2007), we suggest to exploit a DE-based-penalized smoothing approach. More precisely, we denote by $\tilde{u}(\mathbf{x})$ the B-spline tensor product approximation of $u(\mathbf{x})$:

$$\begin{aligned} \tilde{u}(\mathbf{x}) &= (\mathbf{B}_{x_p} \otimes \dots \otimes \mathbf{B}_{x_1}) \mathbf{c} \\ &= \mathcal{B}\mathbf{c}, \end{aligned} \tag{2}$$

where \mathbf{B}_{x_p} is the $(N \times M)$ B-spline matrix defined on the p -th direction where a generous numbers of internal knots were placed and \mathbf{c} is the M^p -vector of spline coefficients. The placement of the knots on the x_p -direction can be either regular or not. To simplify the notation, we suppose that the number of B-splines in each direction is the same and equal to M . The spline coefficients are shrunk by adding (to the log-likelihood) a multiple of a penalty defined using the PDE model (1):

$$PEN(\mathbf{c}|\boldsymbol{\theta}) = \int \mathcal{F}(\mathbf{x}, \tilde{u}, \boldsymbol{\theta})^2 d\mathbf{x}. \tag{3}$$

This penalty term will be close to zero, for given PDE parameters $\boldsymbol{\theta}$, if the approximation $\tilde{u}(\mathbf{x})$ to the solution of (1) is consistent with the PDE model. As we are considering linear PDEs, this penalty is just a polynomial of degree 2 in the spline coefficients:

$$PEN(\mathbf{c}|\boldsymbol{\theta}) = \mathbf{c}^\top \mathbf{R}(\boldsymbol{\theta}) \mathbf{c} + 2\mathbf{c}^\top \mathbf{r}(\boldsymbol{\theta}) + l(\boldsymbol{\theta}),$$

where $\mathbf{R}(\boldsymbol{\theta})$ is the penalty matrix, $\mathbf{r}(\boldsymbol{\theta})$ is the penalty vector and $l(\boldsymbol{\theta})$ is a constant not depending on the spline coefficients. A fast procedure to compute these penalty components is given in Appendix 3.

The compromise between data fitting and fidelity to the differential model can be assessed using the penalized least square criterion:

$$J(\mathbf{c}|\boldsymbol{\theta}, \gamma, \boldsymbol{\zeta}) = \frac{N}{2} \log(\tau) - \frac{\tau}{2} \|\boldsymbol{\zeta} - \tilde{u}(\mathbf{x})\|^2 - \frac{\gamma}{2} PEN(\mathbf{c}|\boldsymbol{\theta}), \tag{4}$$

where τ is the inverse of the error variance (i.e., the precision of measurement). Similar to Ramsay et al. (2007), the estimation process for $\boldsymbol{\theta}$ and \mathbf{c} implies the iteration of two profiling steps. First, for a given value of $\boldsymbol{\theta}$, τ and γ , the spline coefficients \mathbf{c} are estimated as the maximizer of J :

$$\hat{\mathbf{c}} = \underset{\mathbf{c}}{\operatorname{argmax}} J(\mathbf{c}|\boldsymbol{\theta}, \tau, \gamma, \boldsymbol{\zeta}) = \left(\tau \mathbf{B}^\top \mathbf{B} + \gamma \mathbf{R}(\boldsymbol{\theta}) \right)^{-1} \left(\tau \mathbf{B}^\top \boldsymbol{\zeta} - \gamma \mathbf{r}(\boldsymbol{\theta}) \right). \tag{5}$$

Then, given the last available spline coefficients $\hat{\mathbf{c}}$, the PDE parameters $\boldsymbol{\theta}$ and the precision of measurements τ are estimated by maximizing:

$$H(\boldsymbol{\theta}, \tau|\hat{\mathbf{c}}, \boldsymbol{\zeta}) = \frac{N}{2} \log(\tau) - \frac{\tau}{2} \|\boldsymbol{\zeta} - \mathbf{B}\hat{\mathbf{c}}\|^2. \tag{6}$$

The γ parameter weighs and controls the relative emphasis on goodness-of-fit and on solving the partial differential equation. This parameter measures the confidence that one has in the PDE to describe the dynamics in the system. Therefore, we suggest to name it ‘‘PDE-compliance parameter’’. As usual in penalized smoothing approach, the parameter γ has to be selected in a higher optimization level (e.g., using cross-validation, AIC, etc.). Following the mixed model interpretation of penalized smoothing splines (Ruppert et al. 2003), the PDE-compliance parameter can be seen as the ratio of the variances of the residuals and of the penalty. For this reason, we recommend to adopt the following EM-type procedure for the estimation (Schall 1991): update till convergence $\gamma = \hat{\sigma}_\epsilon^2 / \hat{\sigma}_{PEN}^2$ with $\hat{\sigma}_{PEN}^2 = \|PEN(\mathbf{c}|\boldsymbol{\theta})\|^2 / ED$ where $ED = \operatorname{tr}(\mathbf{B}(\tau \mathbf{B}^\top \mathbf{B} + \gamma PEN(\mathbf{c}|\boldsymbol{\theta}))^{-1} \tau \mathbf{B}^\top)$ is the effective dimension of the smoother (Hastie and Tibshirani 1990). Note that for high PDE compliance the effective dimension is usually much lower than the number of spline coefficients (an example is presented in Fig. 2).

By estimating σ_ϵ^2 as $\|\boldsymbol{\zeta} - \tilde{u}(\mathbf{x})\|^2 / (N - ED)$, it is possible to profile $H(\boldsymbol{\theta}, \tau|\hat{\mathbf{c}}, \boldsymbol{\zeta})$ w.r.t. the precision of measurements. Then, the function to be minimized to estimate the unknown PDE parameters becomes:

$$H(\boldsymbol{\theta}|\hat{\tau}, \hat{\mathbf{c}}, \boldsymbol{\zeta}) = -\frac{\hat{\tau}}{2} \|\boldsymbol{\zeta} - \mathbf{B}\hat{\mathbf{c}}\|^2. \tag{7}$$

Confidence intervals for the PDE parameters cannot easily be obtained. Xun et al. (2013) studied the asymptotic properties of the described PDE parameters and spline coefficients estimators. When the hypotheses listed in their appendix A.2 hold, the results presented by the authors can be extended to our procedure and their variance formulas can be used to quantify uncertainty for the estimated unknowns. On the other hand, following the suggestions by Rodriguez-Fernandez et al. (2006) and Xue et al. (2010), we found that a reasonable approximation to the variance–covariance of the $\boldsymbol{\theta}$ parameters can be obtained using the observed pseudo-information matrix given by the inverse of the Hessian of H in (7). Furthermore, approximated piecewise

confidence bounds for the spline coefficients can easily be obtained in case of linear PDE penalties. By defining the hat matrix as $A = \mathbf{B}(\tau\mathbf{B}^\top\mathbf{B} + \gamma PEN(\mathbf{c}|\boldsymbol{\theta}))^{-1}\tau\mathbf{B}^\top$ and noting that $\text{var}(\tilde{u}(\mathbf{x})) \approx \hat{\tau}^{-1} (A^\top A)$, the approximate pointwise confidence bands for the estimated state function are computed as $\tilde{u}(\mathbf{x}) \pm 2\sqrt{\text{diag}[\text{var}(\tilde{u}(\mathbf{x}))]}$.

Alternatively, as suggested by [Xue et al. \(2010\)](#), inference on the model parameters can be done using the weighted bootstrap ([Ma and Kosorok 2005](#)). Differently from empirical bootstrap, this approach suggests to solve repeatedly the minimization problem in Eq. (6) by weighting the residual sum of squares by a set of i.i.d. random weights with unitary mean and variance.

2.2 Bayesian approach

We propose to adapt the frequentist proposal into a Bayesian framework. Contrary to [Xun et al. \(2013\)](#), we will use the same PDE penalty than in the frequentist approach. The penalty term in Eq. (4) appears as a quantity subtracted from the log-likelihood. The log of the joint posterior yields the same fitting criterion using the following model specification:

$$\begin{cases} \boldsymbol{\zeta}|\mathbf{c}, \tau \sim \mathcal{N}_N(\mathbf{B}\mathbf{c}; \tau^{-1}\mathbf{I}_N), \\ p(\mathbf{c}|\boldsymbol{\theta}, \gamma) \propto \exp\left(-\frac{\gamma}{2}PEN\right). \end{cases}$$

The corresponding prior distribution for the spline coefficients is a multivariate normal distribution with mean $\mathbf{V}_1^{-1}\mathbf{v}_1$ and variance-covariance \mathbf{V}_1^{-1} where $\mathbf{v}_1 = -\gamma\mathbf{r}(\boldsymbol{\theta})$ and $\mathbf{V}_1 = \gamma\mathbf{R}(\boldsymbol{\theta})$.

Further prior distributions must be specified. For the precision of measurement τ , it is convenient to take a gamma distribution $\mathcal{G}(a_\tau, b_\tau)$ with mean a_τ/b_τ . If no prior information is available on this precision parameter, we recommend either to set $a_\tau = b_\tau$ equal to a small value (e.g., 10^{-6}) or to set a_τ equal to 1.0 and b_τ to a small value (e.g., 10^{-6}).

The PDE-compliance parameter appears as a precision parameter in the specification of the prior distribution for the spline coefficients. It is therefore convenient to take a gamma distribution $\mathcal{G}(a_\gamma, b_\gamma)$ as prior. To express prior confidence in the PDE model, we recommend to set a_γ to one and b_γ to a small value (e.g., 10^{-8}). Indeed, such a prior for γ is rather flat although it puts slightly more weight on value of $\log_{10}(\gamma)$ around $-\log_{10}(b_\gamma)$ ([Jaeger and Lambert 2013, 2014](#)).

For the PDE parameter $\boldsymbol{\theta}$, the prior distribution depends on contextual information and is denoted by $p(\boldsymbol{\theta})$.

The log joint posterior distribution for $(\mathbf{c}, \boldsymbol{\theta}, \gamma, \tau|\boldsymbol{\zeta})$ can be shown to be:

$$\begin{aligned} \log(p(\mathbf{c}, \boldsymbol{\theta}, \gamma, \tau|\boldsymbol{\zeta})) &= \frac{N}{2} \log(\tau) - \frac{\tau}{2} \|\boldsymbol{\zeta} - \mathbf{B}\mathbf{c}\|^2 \\ &+ \frac{1}{2} \log(\det(\mathbf{V}_1)) - \frac{1}{2} \left\{ \mathbf{c}^\top \mathbf{V}_1 \mathbf{c} - 2\mathbf{c}^\top \mathbf{v}_1 + \mathbf{v}_1^\top \mathbf{V}_1^{-1} \mathbf{v}_1 \right\} \\ &+ (a_\tau - 1) \log(\tau) - b_\tau \tau \\ &+ (a_\gamma - 1) \log(\gamma) - b_\gamma \gamma \\ &+ \log(p(\boldsymbol{\theta})). \end{aligned} \tag{8}$$

It can easily be shown that the conditional posterior distribution for the spline coefficients is a multivariate normal with mean $V_2^{-1}v_2$ and variance–covariance V_2^{-1} where $V_2 = \tau B^T B + V_1$ and $v_2 = \tau B^T \zeta + v_1$:

$$c|\theta, \tau, \gamma, \zeta \sim \mathcal{N}_{Mp} \left(V_2^{-1}v_2; V_2^{-1} \right).$$

The conditional posterior distribution for the precision of measurements and for the PDE-compliance parameter is gamma distributed:

$$\tau|c, \zeta \sim \mathcal{G} \left(\frac{N}{2} + a_\tau; \frac{\|\zeta - Bc\|^2}{2} + b_\tau \right),$$

$$\gamma|\theta, c \sim \mathcal{G} \left(\frac{M^p}{2} + a_\gamma; \frac{c^T R(\theta) c + 2c^T r(\theta) + r(\theta)^T R(\theta)^{-1} r(\theta)}{2} + b_\gamma \right).$$

The conditional posterior distribution for the PDE parameters θ is not necessarily of a familiar type. In addition, high posterior correlation could occur between the spline coefficients c and the elements of θ .

From Eq. (8), it appears that the constant of normalization of the prior distribution of the spline coefficients is given by the determinant of the PDE-penalty matrix and hence depends on the unknown θ parameters. Disregarding this normalizing constant when computing the joint posterior is incorrect and leads to biased parameter estimates.

Finally, following Jaeger and Lambert (2013, 2014), we recommend to marginalize the joint posterior distribution with respect to the spline coefficients, as given in Appendix 1. This marginalization enables us to sample only the PDE parameters with a mixing of the MCMC chains not hindered by the posterior correlation with spline coefficients. The latter quantities can be sampled in a second step if found necessary.

3 Estimation of the unknown differential parameters including state conditions

In many applications, state (initial and/or boundary) conditions naturally arise or are implicitly defined by the observed dynamics (a clear real example is the analysis discussed in Sect. 5). Such information can be included in the statistical framework of Sect. 2.

Consider the general partial problem given by:

$$\begin{cases} \mathcal{F}(x, u, \theta) = 0, \\ \frac{\partial^{(i)}}{\partial x_0^{(i)}} u(x_0) = v(x_0), \end{cases} \tag{9}$$

where x_0 is a part of the domain where the state function u and/or its i th order derivatives are forced to be equal to the function v . Using the B-spline approximation, these conditions can be translated into:

$$Hc = v(x_0), \tag{10}$$

where H is a tensor product B-spline matrix evaluated on the desired support points. In the frequentist framework, we propose two methods to introduce the state conditions in the penalized smoothing approach: one is based on least squares and the other on Lagrange multipliers. In the Bayesian setting, we only include the conditions using a least square type strategy.

3.1 Frequentist approach

One way to take into account the conditions related to a generic PDE is to consider them as an extra penalty. In the frequentist approach, the fitting criterion J is modified as follows:

$$J(c|\theta, \gamma, \kappa, \zeta) = \frac{N}{2} \log(\tau) - \frac{\tau}{2} \|\zeta - \tilde{u}(x)\|^2 - \frac{\gamma}{2} PEN(\theta|c) - \frac{\kappa}{2} (Hc - v(x_0))^T (Hc - v(x_0)). \tag{11}$$

Note that with such a least square penalty, the approximated state function is not forced to be exactly equal to the conditions. With this fitting criterion, the optimal spline coefficients are computed as:

$$\hat{c} = \underset{c}{\operatorname{argmax}} J(c|\theta, \tau, \gamma, \kappa, \zeta) = \left(\tau B^T B + \gamma R(\theta) + \kappa H^T H \right)^{-1} \left(\tau B^T \zeta - \gamma r(\theta) + \kappa H^T v(x_0) \right). \tag{12}$$

If κ tends to infinity, we simply force the state function to be equal to the state conditions at the prescribed points. On the other hand, if κ is equal to zero, then we just go back to the case of Sect. 2.

Using Lagrange multipliers in a frequentist framework, we can force the smoothing function to be exactly consistent with the conditions. The Lagrange function for our constrained maximization problem is:

$$\mathcal{L}(c, \omega|\theta, \tau, \gamma, \zeta) = \frac{N}{2} \log(\tau) - \frac{\tau}{2} \|\zeta - \tilde{u}(x)\|^2 - \frac{\gamma}{2} PEN(\theta|c) - \frac{1}{2} \omega^T (Hc - v(x_0)), \tag{13}$$

where ω is the vector of Lagrange multipliers. As in Currie (2013), the maximization with respect to (c, ω) follows from the solution of:

$$\begin{pmatrix} \tau B^T B + \gamma R(\theta) & H^T \\ H & \mathbf{0} \end{pmatrix} \begin{pmatrix} c \\ \omega \end{pmatrix} = \begin{pmatrix} \tau B^T \zeta - \gamma r(\theta) \\ v(x_0) \end{pmatrix}.$$

The precision of measurement τ and the PDE parameters θ are then estimated in the same way as in Eq. (6).

3.2 Bayesian approach

In a Bayesian setting, the frequentist least squares strategy can be translated into the following prior for the spline coefficients:

$$p(\mathbf{c}|\boldsymbol{\theta}, \gamma, \kappa) \propto \exp\left(\frac{\gamma}{2}PEN - \frac{\kappa}{2}(\mathbf{H}\mathbf{c} - v(\mathbf{x}_0))^\top (\mathbf{H}\mathbf{c} - v(\mathbf{x}_0))\right).$$

This prior distribution corresponds to a multivariate normal distribution with mean $\mathbf{V}_1^{-1}\mathbf{v}_1$ and variance-covariance \mathbf{V}_1^{-1} where $\mathbf{v}_1 = -\gamma\mathbf{r}(\boldsymbol{\theta}) + \kappa\mathbf{H}^\top v(\mathbf{x}_0)$ and $\mathbf{V}_1 = \gamma\mathbf{R}(\boldsymbol{\theta}) + \kappa\mathbf{H}^\top\mathbf{H}$.

One can either fix κ to a large value or consider it as random with a gamma prior distribution $\mathcal{G}(a_\kappa, b_\kappa)$ where a_κ is equal to one and b_κ is equal to a small value (e.g., 10^{-6}). This choice for the prior distribution for κ translates prior confidence in the state conditions. The log joint posterior distribution, the log marginalized posterior distribution and all the conditional posterior distributions can be found in Appendix 2. For the Lagrange multipliers strategy, the possible Bayesian translation would require to project the spline coefficients on a parameter sub-space where the constraints are met: we have no practical implementation of such a strategy for the moment.

4 Simulation

An intensive simulation study was set up to study the properties of the proposed estimation strategies. We consider the following advection-decay partial differential equation:

$$\mathcal{F}\left(x_1, x_2, u, \frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \theta_1, \theta_2\right) = u_{x_1} + \theta_1 u_{x_2} + \theta_2 u = 0, \quad \text{with } x_1 \in \mathbb{R}, x_2 \in \mathbb{R}^+, \tag{14}$$

where u_{x_i} denotes the first derivative of the state function with respect to x_i . The general solution of (14) is:

$$u(x_1, x_2) = g(\theta_1 x_2 - x_1) \exp(-2\theta_2 x_1),$$

where $g(\cdot)$ is an unknown function to be determined from the state conditions. For simulation purposes, we take $\theta_1 = 0.5$ and $\theta_2 = 1.5$ and consider the condition $u(x_1, 0) = \frac{1}{1 + x_1^2}$ (note that it is, in a certain sense, analogous to the terminal condition of the equation discussed in Sect. 5). Applying this state condition, the differential problem has the following closed-form solution:

$$u(x_1, x_2) = \frac{\exp(-3x_2)}{1 + 4x_2^2 - 4x_1x_2 + x_1^2}.$$

This expression contaminated with Gaussian noise is used to generate the data. This PDE appears particularly attractive as it has a closed-form solution, enabling an objec-

tive comparison of the merits of the proposed constrained approaches with the nonlinear least squares one (see Sect. 4.1.2). Furthermore, the estimation abilities of our proposals will be compared with the methods presented by Xun et al. (2013) (see Sect. 4.2).

4.1 Simulation using the PDE-P-splines approaches

The data used for the simulations of this section have been obtained by adding a Gaussian noise component to the analytic solution computed on the grid of equidistant points $x_{1,i}$ in $[-3, 3]$, $x_{2,j}$ in $[0, 1]$ with $i = 1, \dots, 50$ and $j = 1, \dots, 50$. Different levels for the precision of measurements $\tau = \{10000, 400, 100\}$ have been used to simulate the data in such a way to obtain a low, medium and large level of noise in the measurements. Figure 1 shows three possible simulated data clouds (the gray and black level lines on the top of each perspective plot indicate cross-sections, i.e., the contour plots, of the simulated data and of the PDE solution, respectively).

We are interested in evaluating the efficiency of the proposed estimation procedures. For this reason, we consider 500 simulated datasets per noise level and compute, for the frequentist approaches, the relative bias (in percent), the relative root mean squared error and the standard deviation of the estimated differential equation parameters together with the precision of measurements and the optimal compliance parameter. On the other hand, with the Bayesian approaches, we compute the relative bias (based on the posterior mean), the relative root mean squared error (based on the posterior mean), the mean posterior standard deviation and the 80 and 95 % coverage probabilities (based on the HPD intervals).

In this simulation study, we stress our proposal with and without taking into account the state conditions. For the frequentist procedures, we include the state conditions using either the least squares or the Lagrange multiplier strategy. For the Bayesian approach, we only investigate the least squares strategy as it enables to marginalize the joint posterior distribution with respect to the spline coefficients. Finally, the performances of our PDE-P-spline approaches are compared with those achieved using a (frequentist) nonlinear least square estimation procedure (that requires an explicit solution of the PDE and the knowledge of the state conditions).

In this simulation study, we consider non-equally spaced knots. We use a set of 28 spline basis in the x_1 direction and 13 spline basis in the x_2 direction. In both direction, more knots have been located in the part of the domain where variation of the signal appears more evident. The selection of knot location is not mandatory as a fine grid of equidistant knots produces similar estimates (but requires higher computational efforts). For the degree of the B-splines, we found that a useful rule of thumb is to set it equal to the degree of the differential equation plus two. This heuristic rule can be justified by examining the properties of B-spline functions and their derivatives. The d th order derivative of a B-spline basis of degree q is obtained by finite differentiation of spline functions of degree $q - d$. On the other hand, a q th degree B-spline is $q - 1$ times continuously differentiable over the domain span (see e.g., Golub and Ortega 1992; Dierckx 1995). By defining bases of degree equal to the order of the PDE plus two, we ensure that the splines used to approximate the highest order derivative appearing in the differential model are at least one time continuously differentiable over the entire domain mesh.

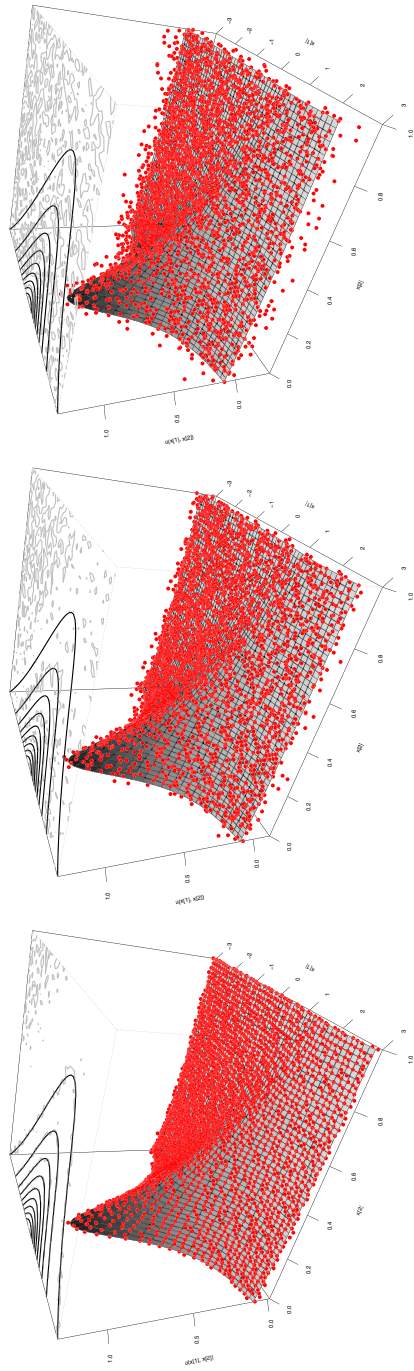


Fig. 1 Three simulated data sets (*dots*) obtained considering an advection-decay dynamics and different levels for the precision of measurements. The surfaces represent the signal used in the simulation. The *contour lines* on the top of the perspective plots give a 1D representation of the simulated data (*solid gray lines*) and of the signal (*solid black lines*)

4.1.1 Simulation results when ignoring the state conditions

The results of the simulation study described above are provided in Table 1. In this case, the state conditions have not been taken into account in the estimation process. The cited table shows some interesting features. The relative bias (in percent) for the

Table 1 Frequentist and Bayesian simulation for advection-decay PDE smoothing example with unknown state conditions

		True parameter values		
		$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 1/\sigma^2$
<i>PDE-P-splines Frequentist</i>				
R-BIAS	$\sigma^2 = 0.01^2$	0.082 %	0.0876 %	2.550 %
	$\sigma^2 = 0.05^2$	0.447 %	0.474 %	2.220 %
	$\sigma^2 = 0.1^2$	1.160 %	1.210 %	0.648 %
R-RMSE	$\sigma^2 = 0.01^2$	0.004	0.005	0.039
	$\sigma^2 = 0.05^2$	0.023	0.025	0.038
	$\sigma^2 = 0.1^2$	0.047	0.681	0.029
R-STD	$\sigma^2 = 0.01^2$	0.004	0.005	0.030
	$\sigma^2 = 0.05^2$	0.022	0.025	0.031
	$\sigma^2 = 0.1^2$	0.045	0.050	0.029
$E(\gamma) = \{2.36E + 08, 5.20E + 07, 4.30E + 07\}$				
<i>PDE-P-splines Bayesian</i>				
R-BIAS	$\sigma^2 = 0.01^2$	0.064 %	0.069 %	0.325 %
	$\sigma^2 = 0.05^2$	1.140 %	1.350 %	0.148 %
	$\sigma^2 = 0.1^2$	4.020 %	4.790 %	0.090 %
R-RMSE	$\sigma^2 = 0.01^2$	0.004	0.005	0.029
	$\sigma^2 = 0.05^2$	0.025	0.028	0.029
	$\sigma^2 = 0.1^2$	0.063	0.072	0.029
MPSD	$\sigma^2 = 0.01^2$	0.002	0.007	0.028
	$\sigma^2 = 0.05^2$	0.011	0.037	1.140
	$\sigma^2 = 0.1^2$	0.023	0.078	2.840
CP80	$\sigma^2 = 0.01^2$	80.2	80.6	80.6
	$\sigma^2 = 0.05^2$	76.6	73.4	80.0
	$\sigma^2 = 0.1^2$	69.4	68.6	79.6
CP95	$\sigma^2 = 0.01^2$	94.4	93.4	92.6
	$\sigma^2 = 0.05^2$	91.0	93.0	92.2
	$\sigma^2 = 0.1^2$	89.2	87.0	93.2
$E(\gamma) = \{1.48E + 07, 2.02E + 07, 3.01E + 07\}$				

At the end of each section in the table, the average optimal compliance parameters (per each simulation setting) are listed

Coverage probabilities in agreement with their nominal values are indicated by bold numbers

estimated parameters is small both for the frequentist and Bayesian approaches. On the other hand, looking at the estimated PDE parameters, the relative bias obtained within the Bayesian framework is positive and larger (but still very small) than for the frequentist estimates, while the estimates for the precision parameter look similar for the two approaches. This could be explained by the choice of the posterior mean for the point estimate. The RMSEs seem to be larger for the PDE parameters in the Bayesian approach and larger for the precision of measurement in the frequentist approach. This can be explained looking at the bias. On the other hand, based on the STDs, the estimation procedure seems quite robust against the noise level. For the Bayesian estimates, the mean posterior standard deviations tend to increase with the level of noise for the PDE parameters and to decrease for the τ parameter. Only some estimated coverage probabilities (at 80 and 95 % level) are in agreement with the nominal values. On the other hand, for a high level of noise, there is evidence of an under coverage effect that could be explained by the fact that the true state conditions are ignored in the estimation process. Finally, the (average) optimal compliance parameters γ shown at the bottom of each sub-table are fairly large, indicating a strong (posterior) confidence in the appropriateness of the differential model used in the penalty to describe the observed dynamics. Figure 2 shows three possible smoothing surfaces coming from this simulation study. As expected, if we do not include the state conditions in the estimation procedure, the fitted surface tends to be sensitive to the level of noise, showing wiggly parts for large values of x_1 and moderate value of x_2 (see first column of Fig. 2).

4.1.2 Simulation results considering the state conditions

The results of the simulation study taking into account the state conditions through a least square penalty (frequentist and Bayesian approaches with fixed $\kappa = 10^6$) or Lagrange multipliers (only frequentist approach) are shown in Table 2. The simulation performances obtained using a nonlinear least square solution of the state function are also shown in Table 3 (as made possible by the availability of a unique and explicit solution when state conditions are known). It appears that the relative bias for the estimated parameters is small both for the frequentist and Bayesian approaches. One can notice that the bias decreases in absolute value when the state conditions are introduced in the estimation procedure in the case of moderate and high noise variability. For frequentist estimates, the RMSEs seem to increase together with the level of measurement noise. On the other hand, based on the STDs, the estimation procedure seems quite robust to the noise level. The estimated standard deviations for the PDE parameters seem to be smaller when the state conditions are ignored. Looking at the frequentist estimates, the results obtained by imposing conditions using least squares or Lagrange multipliers look really similar. For the Bayesian estimates, the mean posterior standard deviations for the PDE parameters increase with the level of noise and decrease for parameter τ . As expected, the mean standard deviation for the PDE parameters tends to decrease when forcing the state conditions. The estimate coverage probabilities are in agreement with the nominal values in almost all the simulation settings. This is probably due to the fact that the conditions are explicitly specified and not estimated. The average optimal compliance parameters γ given in each sub-table indicate a stronger (posterior)

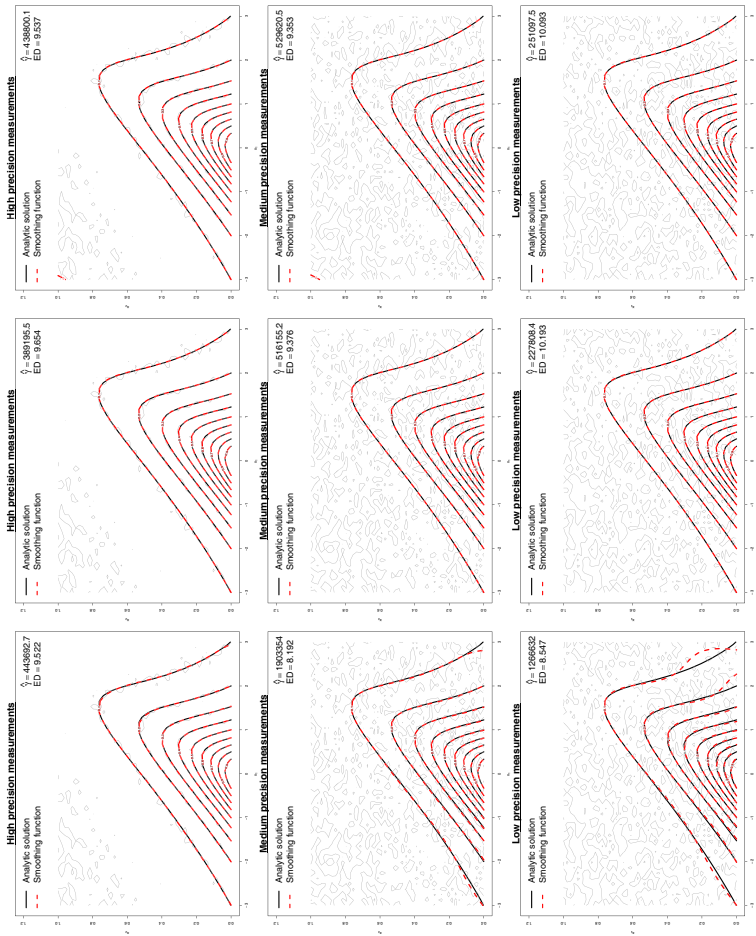


Fig. 2 Typical profiles for the frequentist approaches. The first column shows the smoothing surfaces estimated for unknown conditions. The middle column shows the smoothing surfaces estimated using the state conditions imposed by least squares, while the third column shows the smoothing surface using the Lagrange multipliers approach. The gray *level lines* represent the cross-sections of the simulated cloud of points (*contour lines*). In the upper right legend of each panel, the optimal PDE-compliance parameters and the effective smoother dimensions are reported

Table 2 Frequentist and Bayesian simulation for advection-decay PDE smoothing example considering the state conditions

		True parameter values		
		$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 1/\sigma^2$
PDE-P-splines Frequentist (Least squares)				
R-BIAS	$\sigma^2 = 0.01^2$	0.106 %	0.073 %	1.510 %
	$\sigma^2 = 0.05^2$	0.146 %	0.051 %	2.120 %
	$\sigma^2 = 0.1^2$	0.162 %	0.020 %	0.584 %
R-RMSE	$\sigma^2 = 0.01^2$	0.003	0.004	0.062
	$\sigma^2 = 0.05^2$	0.016	0.018	0.037
	$\sigma^2 = 0.1^2$	0.032	0.668	0.030
R-STD	$\sigma^2 = 0.01^2$	0.003	0.004	0.060
	$\sigma^2 = 0.05^2$	0.016	0.018	0.030
	$\sigma^2 = 0.1^2$	0.033	0.036	0.030
$E(\gamma) = \{2.10E+08, 2.94E+07, 1.21E+07\}$				
PDE-P-splines Frequentist (Lagrange)				
R-BIAS	$\sigma^2 = 0.01^2$	0.112 %	0.061 %	2.240 %
	$\sigma^2 = 0.05^2$	0.144 %	0.047 %	2.130 %
	$\sigma^2 = 0.1^2$	0.159 %	0.014 %	0.578 %
R-RMSE	$\sigma^2 = 0.01^2$	0.003	0.004	0.037
	$\sigma^2 = 0.05^2$	0.016	0.018	0.037
	$\sigma^2 = 0.1^2$	0.033	0.668	0.030
R-STD	$\sigma^2 = 0.01^2$	0.003	0.004	0.030
	$\sigma^2 = 0.05^2$	0.016	0.018	0.030
	$\sigma^2 = 0.1^2$	0.033	0.036	0.029
$E(\gamma) = \{2.11E+08, 3.62E+07, 2.08E+07\}$				
PDE-P-splines Bayesian				
R-BIAS	$\sigma^2 = 0.01^2$	0.016 %	0.008 %	0.278 %
	$\sigma^2 = 0.05^2$	0.242 %	0.224 %	0.233 %
	$\sigma^2 = 0.1^2$	0.814 %	0.821 %	0.228 %
R-RMSE	$\sigma^2 = 0.01^2$	0.003	0.003	0.029
	$\sigma^2 = 0.05^2$	0.016	0.018	0.029
	$\sigma^2 = 0.1^2$	0.034	0.037	0.029
MPSD	$\sigma^2 = 0.01^2$	0.001	0.005	20.84
	$\sigma^2 = 0.05^2$	0.008	0.026	10.14
	$\sigma^2 = 0.1^2$	0.016	0.052	2.840
CP80	$\sigma^2 = 0.01^2$	80.2	80.2	80.6
	$\sigma^2 = 0.05^2$	78.4	78.2	80.6
	$\sigma^2 = 0.1^2$	78.0	79.0	79.8

Table 2 continued

		True parameter values		
		$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 1/\sigma^2$
CP95	$\sigma^2 = 0.01^2$	95.0	94.2	93.2
	$\sigma^2 = 0.05^2$	92.4	93.4	93.0
	$\sigma^2 = 0.1^2$	93.2	93.2	93.0
$E(\gamma) = \{1.05E+08, 1.48E+08, 1.37E+08\}$				

At the end of each section in the table, the average optimal compliance parameters (per each simulation setting) are listed

Coverage probabilities in agreement with their nominal values are indicated by bold numbers

Table 3 Frequentist nonlinear least squares simulation for advection-decay PDE smoothing example

		True parameter values		
		$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 1/\sigma^2$
Frequentist (NLS)				
R-BIAS	$\sigma^2 = 0.01^2$	0.050 %	0.054 %	-0.110 %
	$\sigma^2 = 0.05^2$	0.061 %	0.071 %	0.216 %
	$\sigma^2 = 0.1^2$	-0.070 %	0.140 %	0.865 %
R-RMSE	$\sigma^2 = 0.01^2$	0.006	0.010	0.001
	$\sigma^2 = 0.05^2$	0.027	0.050	0.007
	$\sigma^2 = 0.1^2$	0.053	0.103	0.001
R-STD	$\sigma^2 = 0.01^2$	0.006	0.011	0.001
	$\sigma^2 = 0.05^2$	0.027	0.050	0.007
	$\sigma^2 = 0.1^2$	0.053	0.103	0.001

confidence in the PDE model as descriptor of the dynamics of the observed state. The middle and the third column of Fig. 2 show three smoothing surfaces extracted from the simulation runs. Compared to the first column, the inclusion of the state conditions forces the smoothing surface to be adherent to the closed-form solution over the entire domain even for lower precision of the measurements. Finally, the performance of the proposed frequentist and Bayesian estimation procedures are compatible with those obtained through nonlinear least squares estimation of the parameters involved in the analytic solution (see Table 3) in the exceptional case where it has a closed-form expression.

4.2 Comparison of the PDE-P-splines approaches with the one by Xun et al. (2013)

In their Bayesian framework, Xun et al. (2013) propose to use a penalty combining PDE compliance (in a L^2 sense) and smoothness of the estimated surface (achieved

by an extra finite difference penalty). Our interest is to compare this proposal with our Bayesian approaches (that are, as shown above, usable with known and unknown state conditions). Furthermore, we also consider a modified version of the penalty proposed by Xun et al. (2013) by discarding the extra finite difference penalties to evaluate their impact on the quality of the estimates. The two main differences between our proposals and the one by Xun et al. (2013) lay in the construction of the penalty term and in the inclusion of the normalizing constant of the prior distribution of the spline coefficients when computing the conditional posterior for PDE parameters θ (see Jaeger and Lambert (2013, 2014) for a discussion on this crucial point). Our regularization term is defined as an integrated penalty ensuring the compliance of the B-spline approximation to the proposed PDE. Note that our Bayesian approach is a faithful “Bayesianization” of the frequentist one. As appears in Eq. (8), we compute the normalizing constant of the prior distribution for the spline coefficients as the determinant of the PDE-penalty matrix whereas Xun et al. (2013) approximate it by a power of the smoothing parameters.

The data (100 samples) used for the comparison have been obtained by adding a Gaussian noise component ($\sigma_\epsilon^2 = 0.05^2$) to the analytic solution of the PDE given in Eq. (14) computed on the grid of equidistant points $x_{1,i}$ in $[-4, 4]$, $x_{2,j}$ in $[0, 1]$ with $i = 1, \dots, n$, where $n \times m = 40 \times 20, 60 \times 40, 100 \times 50, 200 \times 100$. The performances of the competing methods are shown in Table 4. It appears that our PDE-P-spline approach (ignoring the state conditions) guarantees better performances than the competing approach in terms of estimation bias. This is particularly evident when looking at the results obtained for smaller sample sizes. For moderate sample sizes, also the RMSEs obtained with the proposed procedure are smaller than those achieved by the competitors. By comparing the estimation performances of the approach by Xun et al. (2013) with those obtained using an analogous model specification without extra finite difference penalties, it appears that the extra smoothing term has the effect to reduce the estimation bias. Finally, similar to the preceding discussion, the inclusion of the state conditions increases the quality of the estimates.

In Figs. 3 and 4, we present the estimates obtained from a single set of data simulated by adding a Student t noise (with 5 degrees of freedom) to the solution of Eq. (14). Figure 3 shows the estimated surfaces obtained using the four approaches. For the large sample sizes, all the estimated smoothing surfaces correctly estimate the underlying state function. When the sample size decreases, our approaches seem to perform better than the competitors in recovering the underlying signal. In the case where the sample is limited, the approach of Xun et al. (2013) tends to oversmooth the data and seems to put (relatively) lesser weight on the PDE-model penalty. On the other hand, for the same simulation settings, the modified version of the model by Xun et al. (2013) without the extra finite difference penalty seems to catch the signal slightly better than its original formulation for intermediate sample sizes. The posterior densities for the PDE parameters in the modified Xun et al. approach appear clearly left shifted (see Fig. 4). The extra finite difference penalty mitigates the introduced bias. The inclusion of the exact normalizing constant in the prior distribution for the spline coefficient centers the marginal posterior distributions around the true PDE parameters. Finally, whatever the sample size, the variability of the posteriors decreases when the state conditions are introduced in the model.

Table 4 Bayesian simulation for advection-decay PDE smoothing example using PDE-P-splines (with and without state conditions) and the Xun et al. (2013) approach (with and without the extra finite difference penalties)

		PDE-P-splines										Xun et al. (2013)							
		N = 40 × 20		N = 60 × 40		N = 100 × 50		N = 200 × 100		N = 40 × 20		N = 60 × 40		N = 100 × 50		N = 200 × 100			
		$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 400$	$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 400$	$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 400$	$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 400$	$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 400$	$\theta_1 = 0.5$	$\theta_2 = 1.5$	$\tau = 400$
R-BIAS	$\theta_1 = 0.5$	2.483 %	0.964 %	0.939 %	0.266 %	-3.822 %	3.350 %	0.974 %	0.598 %										
	$\theta_2 = 1.5$	3.167 %	1.290 %	1.048 %	0.335 %	-3.913 %	1.541 %	1.109 %	0.282 %										
	$\tau = 400$	-1.002 %	0.379 %	0.150 %	-0.134 %	-8.073 %	-3.568 %	-2.367 %	-0.396 %										
R-RMSE	$\theta_1 = 0.5$	0.049	0.033	0.020	0.010	0.075	0.038	0.031	0.018										
	$\theta_2 = 1.5$	0.048	0.026	0.019	0.011	0.052	0.023	0.022	0.017										
	$\tau = 400$	0.058	0.029	0.021	0.012	0.062	0.080	0.031	0.018										
MSTD	$\theta_1 = 0.5$	0.045	0.027	0.010	0.014	0.071	0.030	0.022	0.015										
	$\theta_2 = 1.5$	0.050	0.030	0.034	0.016	0.076	0.032	0.024	0.016										
	$\tau = 400$	0.049	0.029	0.020	0.010	0.054	0.029	0.020	0.010										
CP80	$\theta_1 = 0.5$	77	82	79	70	74	78	78	77										
	$\theta_2 = 1.5$	74	82	84	69	78	80	77	78										
	$\tau = 400$	80	75	81	79	72	82	82	81										
CP95	$\theta_1 = 0.5$	91	91	96	93	90	92	94	93										
	$\theta_2 = 1.5$	90	96	97	94	89	92	92	94										
	$\tau = 400$	92	92	94	93	90	90	90	91										

Table 4 continued

		PDE-P-splines with state conditions					Xun et al. (2013) no extra-penalty						
		$N = 40 \times 20$	$N = 60 \times 40$	$N = 100 \times 50$	$N = 200 \times 100$	$N = 40 \times 20$	$N = 60 \times 40$	$N = 100 \times 50$	$N = 200 \times 100$	$N = 40 \times 20$	$N = 60 \times 40$	$N = 100 \times 50$	$N = 200 \times 100$
R-BIAS	$\theta_1 = 0.5$	1.034 %	0.422 %	0.156 %	0.062 %	-9.403 %	-5.349 %	-3.342 %	0.062 %	-9.403 %	-5.349 %	-3.342 %	-1.593 %
	$\theta_2 = 1.5$	1.122 %	0.376 %	0.205 %	0.093 %	-8.481 %	-5.408 %	-3.394 %	0.093 %	-8.481 %	-5.408 %	-3.394 %	-1.595 %
	$\tau = 400$	0.631 %	-0.238 %	-0.011 %	-0.016 %	-0.333 %	-1.085 %	0.201 %	-0.016 %	-0.333 %	-1.085 %	0.201 %	0.191 %
R-RMSE	$\theta_1 = 0.5$	0.033	0.019	0.011	0.007	0.142	0.081	0.069	0.007	0.142	0.081	0.069	0.033
	$\theta_2 = 1.5$	0.037	0.020	0.012	0.008	0.118	0.091	0.070	0.008	0.118	0.091	0.070	0.033
MSTD	$\tau = 400$	0.049	0.030	0.018	0.011	0.071	0.030	0.019	0.011	0.071	0.030	0.019	0.010
	$\theta_1 = 0.5$	0.031	0.019	0.013	0.008	0.081	0.043	0.032	0.008	0.081	0.043	0.032	0.018
	$\theta_2 = 1.5$	0.035	0.020	0.014	0.009	0.056	0.033	0.019	0.009	0.056	0.033	0.019	0.020
CP80	$\tau = 400$	0.049	0.029	0.020	0.010	0.054	0.039	0.023	0.010	0.054	0.039	0.023	0.015
	$\theta_1 = 0.5$	79	82	79	85	67	58	68	85	67	58	68	77
	$\theta_2 = 1.5$	80	78	78	84	70	70	67	84	70	70	67	78
CP95	$\tau = 400$	80	75	80	79	66	82	81	79	66	82	81	81
	$\theta_1 = 0.5$	96	95	95	97	83	89	88	97	83	89	88	88
	$\theta_2 = 1.5$	94	97	94	94	82	88	88	94	82	88	88	90
	$\tau = 400$	95	92	92	94	94	82	92	94	94	82	92	93

Coverage probabilities in agreement with their nominal values are indicated by bold numbers

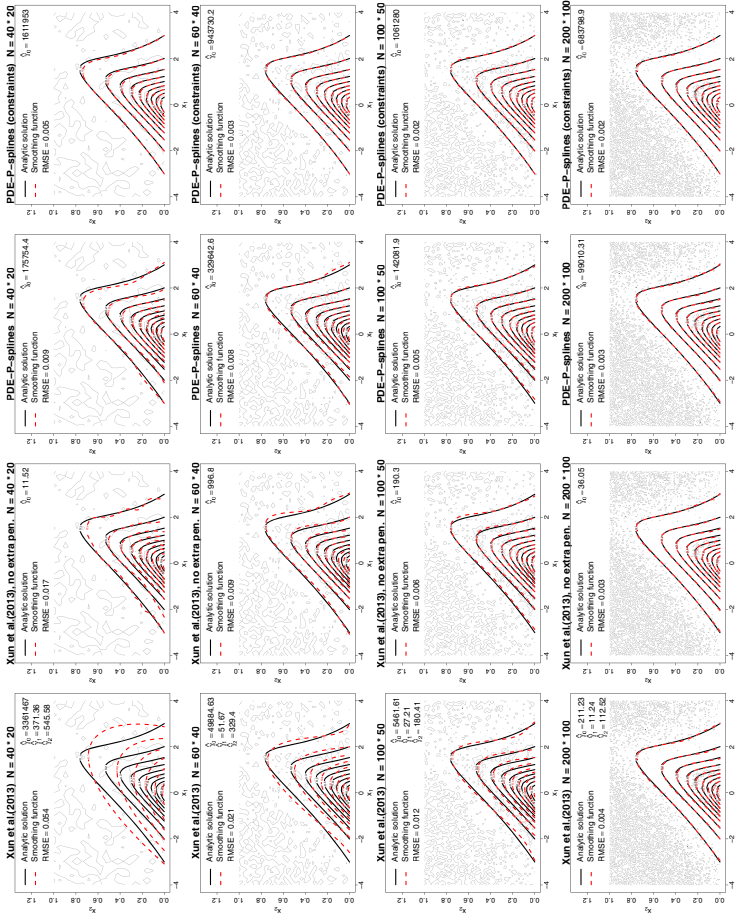


Fig. 3 Posterior pointwise smoothing surfaces estimated either using the [Xun et al. \(2013\)](#) approach with and without the extra finite difference penalty (fourth and third rows) and the PDE-P-splines approaches, without and with the boundary conditions imposed by least squares (first and second rows). In each column, the surfaces estimated for different sample sizes are depicted ($n = 40 \times 20$ in the left column, $n = 60 \times 40$ and $n = 100 \times 50$ in the middle columns and $n = 200 \times 100$ in the right column). The gray level lines indicate cross-sections of the simulated cloud of points (contour lines). The data have been simulated by adding a noise drawn from a Student t distribution with 5 degrees of freedom to the analytic solution of the PDE in (14). The optimal PDE-compliance parameters and the RMSE between the simulated and approximated signals are reported in the legend of each panel

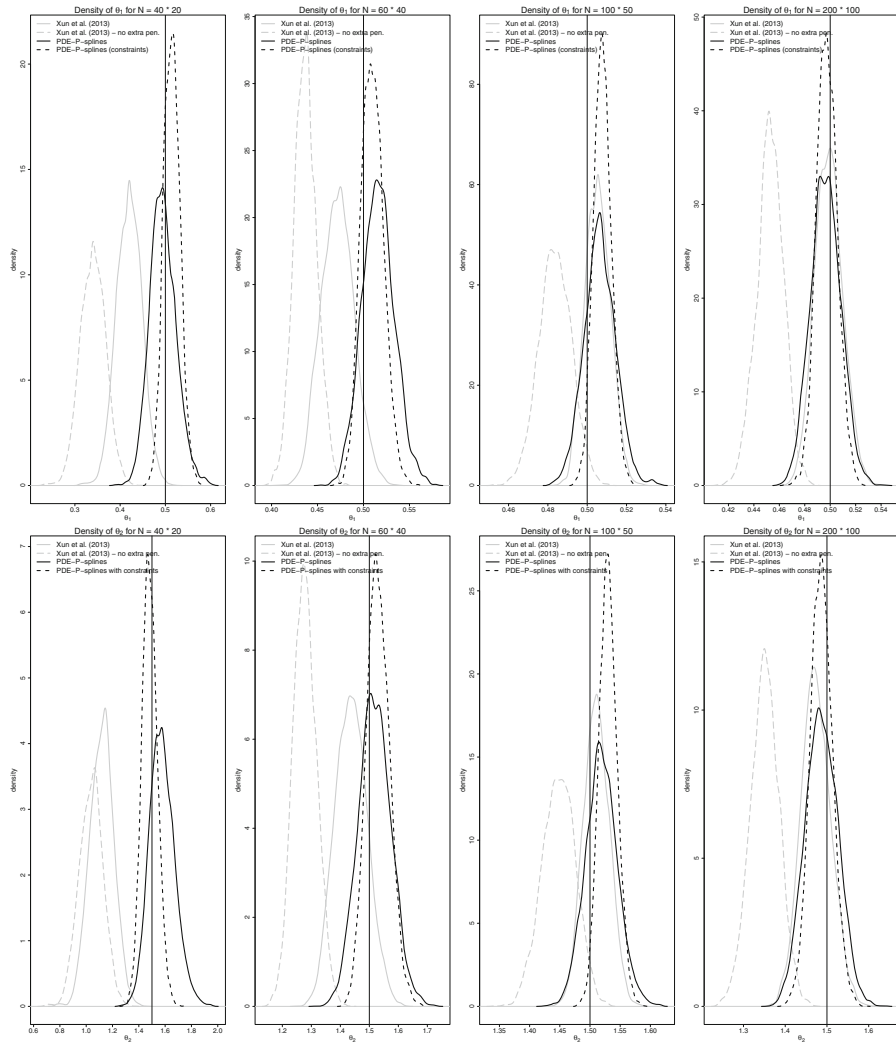


Fig. 4 Posterior densities for $\theta_1 = 0.5$ and $\theta_2 = 1.5$, respectively. Densities in gray are those obtained using the approaches proposed by Xun et al. (2013) (gray curves: solid with the extra finite difference penalty and dashed without it). Densities in black are those obtained using the PDE-P-splines approaches (solid curve for unknown state conditions and dashed for known conditions). In each column, the posterior densities for different sample sizes are depicted ($n = 40 \times 20$ in the left column, $n = 60 \times 40$ and $n = 100 \times 50$ in the middle columns and $n = 200 \times 100$ in the right column). The data have been simulated by adding a 5 degrees of freedom Student t distributed noise to the analytic solution of the PDE in (14)

5 Application

In this section, we evaluate the performance of the proposed approaches in a real data application. We analyze a set of prices of “option contracts” traded for different “strike prices” and “maturities”. Cox et al. (1979), an option is a contract that gives the right,

but not the obligation, to buy or sell a risky asset at a predetermined fixed price within (American style) or at (European style) a given date (maturity of the contract). This kind of financial instrument allows to bet on the future evolution of the price of the underlying asset that can be another contract (a stock for example), a market index or a commodity. Options and other related instruments are often called “derivatives” given that their price is derived from the market value (S) of the underlying risky asset.

Two different kinds of option contracts are distinguished according to the right that they give to the holder. “Call” options give the right to buy the underlying asset, while “put” give the right to sell it. These characteristics determine the payoff functions at maturity (T):

$$C(S, T) = \begin{cases} 0 & \text{if } S_T \leq E \\ S_T - E & \text{if } S_T > E, \end{cases}$$

$$P(S, T) = \begin{cases} 0 & \text{if } S_T \geq E \\ E - S_T & \text{if } S_T < E, \end{cases}$$

where C and P indicate call and put prices, respectively, E is the strike price and S_T represents the market price of the underlying asset at maturity. The current price of the contract is then computed as the present value of the payoff at maturity. For this reason, it is necessary to describe the dynamics of the underlying asset to price the derivative contract “written” on it.

The [Black and Scholes \(1973\)](#) framework is the most famous pricing model for financial derivatives. This model assumes that the price of the underlying risky asset follows a geometric Brownian motion in the instantaneous time with a constant standard deviation (volatility). This assumption leads to the well-known second-order partial differential equation describing the value (C) of a call option:

$$\mathcal{F}\left(S, t, C, \frac{\partial C}{\partial S}, \frac{\partial C}{\partial t}, \frac{\partial^2 C}{\partial S^2}, \sigma\right) = \frac{\partial C}{\partial t} + rS \frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} - rC = 0, \quad (15)$$

where the parameter σ identifies the “implied volatility” and r is the observed risk-free interest rate (e.g., the Libor rate). For European options (exercise allowed only at a given date), a closed-form solution can be found by introducing the following no-arbitrage terminal (related to the expiration of the contract) and boundary conditions:

$$\begin{aligned} C(0, t) &= 0, \\ C(S, t) &\approx S \quad \text{for } S \rightarrow \infty, \\ C(S, T) &= \max(S - E, 0) = (S - E)^+. \end{aligned} \quad (16)$$

This framework provides a concise and easy-to-interpret description of the price dynamics. Here, we aim to approximate the solution of the Black and Scholes PDE and estimate its implied volatility by including (15) and no-arbitrage conditions (16) in a PDE-P-spline approach. Note that we share with the theoretical model the strong assumption of constant (implied) volatility over strikes and maturities. On the other hand, besides the limitations of the differential model, we believe that the proposed

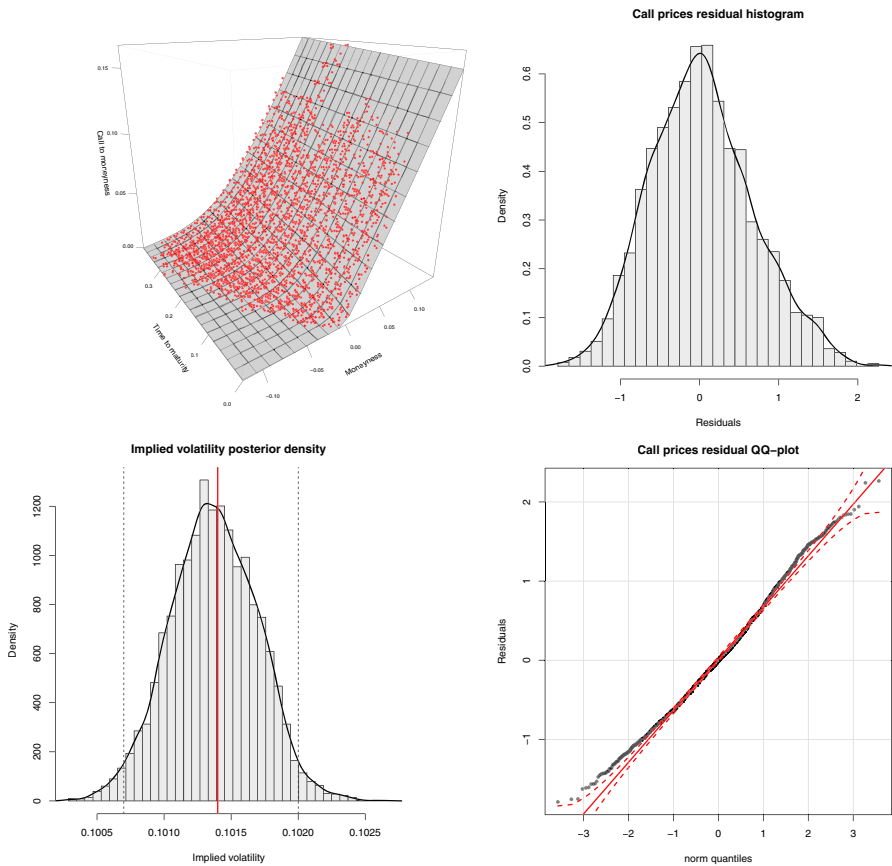


Fig. 5 Left upper panel: observations (*red dots*) and smoothing surface obtained applying a frequentist constrained PDE-P-spline estimation procedure to the SP-500 European call option prices analyzed by Carmona (2004). The prices in the plot are scaled by the observed strikes and the moneyness is defined as the log of the ratio between the observed spot and strike prices. The estimated implied volatility was $\hat{\sigma} = 1.003E - 01$ with an optimal compliance parameter equal to $\hat{\gamma} = 47385275$. Left lower panel posterior density of the implied volatility parameter (Bayesian estimates). Right panels histogram and qq-plot of the residuals (call price scale)

estimation procedure is flexible enough to ensure appropriate estimates. As data example, we consider the option prices already analyzed by Carmona (2004). The dataset counts 2800 prices of European call options written on the SP-500 index and traded in 1993 with expiration between 1993 and 1994. The six variables included in the dataset indicate the value of the index, the strike prices of the option, the time to expiration (in fraction of year), the spot interest rate, the observed implied volatilities and the prices of the options. This application is, in our opinion, particularly interesting since it represents one of those cases in which the state conditions are known a priori (given by theoretical arguments) and need to be taken into account to obtain valid estimates.

The raw data and the estimated smoothing surface (obtained through a soft constrained frequentist approach) are shown in Fig. 5. According to the option-pricing

theory, the payoff surface was expected to be rather flat over large portions of the domain mesh. For this reason, we decided to build our smoother using tensor product of B-splines of degree four defined on 25 equidistant internal knots for both the maturity and moneyness directions. The moneyness parametrization of the PDE (see [Piché and Kannianen \(2007\)](#)) allows for a simple formulation of the penalty and of the necessary state conditions.

The estimated implied volatilities are equal to 1.003E-01 and 1.014E-01 for the frequentist and Bayesian procedures, respectively, and are in agreement with the median of the ones observed in the dataset. The frequentist 95 % approximate confidence interval (obtained using the observed pseudo-information matrix) is [9.967E-02; 1.015E-01], while the corresponding credibility interval has been found to be equal to [1.007E-01; 1.020E-01]. The lower left panel of Fig. 5 shows the posterior distribution of the estimated parameter.

The estimates obtained using our approach show some desirable properties. First of all, we are able to extrapolate the prices associated with strikes and maturities not traded, and these estimates are consistent with the no-arbitrage constraints (due to the compliance to the Black and Scholes model). Furthermore, it is particularly appreciable in real data analysis to have confidence/credible bounds for prices associated to unobserved strikes and/or maturities. On the other hand, the PDE-P-spline approach enables here to obtain point and interval estimates of the implied volatility consistent with the Black and Scholes equation while ensuring a good performance in terms of data fitting.

6 Discussion

In this paper, we propose a PDE-based penalized tensor product B-spline smoothing approach to solve and estimate unknown parameters in partial differential equations. Our aim was to introduce frequentist and Bayesian procedures to analyze data which dynamics is defined over more than one dimension. Both approaches exploit a tensor product B-spline approximation to the state function, while the consistency of the final estimates to the PDE is ensured by PDE-based soft constraints. The compromise between data fitting and consistency with the PDE is tuned by a compliance parameter. As in [Xun et al. \(2013\)](#), we estimate the spline coefficients, the PDE parameters and the precision of measurement using the available observations. If state conditions are not included in the model, our frequentist approach and the one by [Xun et al. \(2013\)](#) are comparable except that we provide an automatic EM-like procedure to select the PDE-compliance parameter ([Schall 1991](#)). In our Bayesian approach, differently from [Xun et al. \(2013\)](#), we use the same PDE-based penalty as with the frequentist approach. In particular, our Bayesian formulation takes into account the functional dependence of the normalizing constant of the spline coefficient prior on PDE parameters. Note that prior confidence in the PDE model is introduced by a specific choice for the prior distribution of the compliance parameter.

Even if their knowledge is not strictly required for the application of the proposed methods, we are able to improve the quality of the estimates by introducing state conditions as extra constraints using either a L^2 penalty or Lagrange multipliers. On

the other hand, the introduction of these conditions sometimes arises naturally from theoretical arguments and cannot be omitted during the estimation process. This is the case for the real data analysis in Sect. 5.

We demonstrate through simulations that our PDE-P-spline approaches show desirable properties and that the inclusion of the state conditions has an influence on the estimation performances by reducing the variability of the PDE parameters and by improving the coverages of credible regions. For known state conditions, the appropriateness of our proposal has been tested by comparing its performances with those achieved via nonlinear least squares inversion of the analytic state function.

We have also compared our Bayesian proposals with the one suggested by [Xun et al. \(2013\)](#). By simulating 100 synthetic data sets with Gaussian noise, we found that, the introduction of an additional finite difference penalty, within their Bayesian approach, helps to reduce the bias of the PDE parameter estimates and mitigates the effect of the miss-specified constant of normalization in the prior of the spline coefficients. The effectiveness of this correction seems to decrease with the size of the data sample. This appears also from the results presented in Figs. 3 and 4 showing the approximated state functions and the PDE parameters posterior densities obtained for a simulated set of data with Student t distributed noise. For small sample sizes, the extra penalty term tends to produce smoothing surface not compliant with the state function. These undesirable features are not shared by our Bayesian proposal. Indeed, even for small sample sizes and/or large level of noise, the use of the proper prior for the spline coefficients ensures good fitting and estimation performances.

As real data application, we have analyzed the SP-500 call option prices discussed in [Carmona \(2004\)](#). The proposed example represents one of those cases where the state conditions (as defined in Sect. 3) are known (due to theoretical arguments) and cannot be ignored to estimate a smoothing surface consistent with the necessary no-arbitrage constraints. We used the well-known Black and Scholes model to define a suitable PDE-based penalty. Despite its simplicity, the presented analysis based on the Black and Scholes differential equation ensured satisfactory results in terms of signal extraction and PDE parameter estimation.

Extensions of our approaches are possible. First of all, the analysis of the call option prices in Sect. 5 could be improved. In the Black and Scholes model, the volatility parameter is supposed to be constant over time. This assumption is obviously unrealistic. One way to allow for time-varying volatility could be to adopt a more realistic differential pricing model (e.g., the stochastic volatility model of [Heston \(1993\)](#)). The same result could be achieved, in our opinion, by allowing the implied volatility parameter in the Black and Scholes equation to vary (e.g., using P-splines) with the time to maturity.

In this paper, we have considered only linear partial differential equations. Moving to nonlinear partial differential equation would lead to additional challenges. First of all, the PDE-penalty term would no longer be a second-order homogeneous polynomial in the spline coefficients. In the frequentist approach, this would require the use of the implicit function theorem to estimate the PDE parameters ([Ramsay et al. 2007](#)). On the other hand, in a Bayesian setting, the constant of normalization for the prior distribution for the spline coefficients would not have an explicit form anymore.

Acknowledgments The authors acknowledge financial support from IAP research network P7/06 of the Belgian Government (Belgian Science Policy), and from the contract ‘Projet d’Actions de Recherche Concertées’ (ARC) 11/16-039 of the ‘Communauté française de Belgique’, granted by the ‘Académie universitaire Louvain’.

Appendix 1: Logarithm of the marginalized posterior distribution

The joint posterior distribution given in Eq. (8) is marginalized with respect to the spline coefficients. The log of the marginalized posterior distribution can be shown to be:

$$\begin{aligned} \log(p(\boldsymbol{\theta}, \gamma, \tau|\boldsymbol{\zeta})) &= \frac{N}{2} \log(\tau) - \frac{\tau}{2} \boldsymbol{\zeta}^\top \boldsymbol{\zeta} \\ &+ \frac{1}{2} \log(\det(V_1)) - \frac{1}{2} \mathbf{v}_1^\top V_1^{-1} \mathbf{v}_1 \\ &- \frac{1}{2} \log(\det(V_2)) + \frac{1}{2} \mathbf{v}_2^\top V_2^{-1} \mathbf{v}_2 \\ &+ (a_\tau - 1) \log(\tau) - b_\tau \tau \\ &+ (a_\gamma - 1) \log(\gamma) - b_\gamma \gamma \\ &+ \log(p(\boldsymbol{\theta})). \end{aligned}$$

Appendix 2: Posterior distributions including the DE conditions by least squares

The log joint posterior distribution can be shown to be:

$$\begin{aligned} \log(p(\mathbf{c}, \boldsymbol{\theta}, \gamma, \kappa, \tau|\boldsymbol{\zeta})) &= \frac{N}{2} \log(\tau) - \frac{\tau}{2} \|\boldsymbol{\zeta} - \mathbf{B}\mathbf{c}\|^2 \\ &+ \frac{1}{2} \log(\det(V_1)) - \frac{1}{2} \left\{ \mathbf{c}^\top V_1 \mathbf{c} - 2\mathbf{c}^\top \mathbf{v}_1 + \mathbf{v}_1^\top V_1^{-1} \mathbf{v}_1 \right\} \\ &+ (a_\tau - 1) \log(\tau) - b_\tau \tau \\ &+ (a_\gamma - 1) \log(\gamma) - b_\gamma \gamma \\ &+ (a_\kappa - 1) \log(\kappa) - b_\kappa \kappa \\ &+ \log(p(\boldsymbol{\theta})). \end{aligned}$$

From this joint posterior distribution, only two conditional posterior distribution can be identified. The conditional posterior distribution for the spline coefficients is a multivariate normal distribution:

$$\mathbf{c}|\boldsymbol{\theta}, \tau, \gamma, \boldsymbol{\zeta} \sim \mathcal{N}_{Mp} \left(V_2^{-1} \mathbf{v}_2; V_2^{-1} \right).$$

where $V_2 = \tau \mathbf{B}^\top \mathbf{B} + V_1$ and $\mathbf{v}_2 = \tau \mathbf{B}^\top \boldsymbol{\zeta} + \mathbf{v}_1$. For the precision of measurement, the conditional posterior distribution is a gamma distribution:

$$\tau | \mathbf{c}, \boldsymbol{\zeta} \sim \mathcal{G} \left(\frac{N}{2} + a_\tau; \frac{\|\boldsymbol{\zeta} - \mathbf{B}\mathbf{c}\|^2}{2} + b_\tau \right).$$

The joint posterior distribution can be marginalized with respect to the spline coefficients. The log of the marginalized posterior distribution can be shown to be:

$$\begin{aligned} \log(p(\boldsymbol{\theta}, \gamma, \tau | \boldsymbol{\zeta})) &= \frac{N}{2} \log(\tau) - \frac{\tau}{2} \boldsymbol{\zeta}^\top \boldsymbol{\zeta} \\ &\quad + \frac{1}{2} \log(\det(V_1)) - \frac{1}{2} \mathbf{v}_1^\top V_1^{-1} \mathbf{v}_1 \\ &\quad - \frac{1}{2} \log(\det(V_2)) + \frac{1}{2} \mathbf{v}_2^\top V_2^{-1} \mathbf{v}_2 \\ &\quad + (a_\tau - 1) \log(\tau) - b_\tau \tau \\ &\quad + (a_\gamma - 1) \log(\gamma) - b_\gamma \gamma \\ &\quad + (a_\kappa - 1) \log(\kappa) - b_\kappa \kappa \\ &\quad + \log(p(\boldsymbol{\theta})). \end{aligned}$$

Appendix 3: Elementary penalty elements

The penalty term, dealing with linear PDEs, can be seen as a homogeneous polynomial of second degree in the spline coefficients. The computations needed to construct the penalty term are:

$$PEN(\mathbf{c} | \boldsymbol{\theta}) = \mathbf{c}^\top \mathbf{R}(\boldsymbol{\theta}) \mathbf{c} + 2\mathbf{c}^\top \mathbf{r}(\boldsymbol{\theta}) + l(\boldsymbol{\theta}).$$

The \mathbf{R} matrix is given by the sum of a series of elementary penalty matrices:

$$\begin{aligned} \mathbf{P}_{i_p, \dots, i_1}^{j_p, \dots, j_1} &= \int (\mathbf{B}_p^{(i_p)}(x_p) \otimes \dots \otimes \mathbf{B}_1^{(i_1)}(x_1)) (\mathbf{B}_p^{(j_p)}(x_p) \otimes \dots \otimes \mathbf{B}_1^{(j_1)}(x_1))^\top dx_p \dots dx_1, \\ &= \int \mathbf{B}_p^{(i_p)}(x_p) (\mathbf{B}_p^{(j_p)}(x_p))^\top dx_p \otimes \dots \otimes \int \mathbf{B}_1^{(i_1)}(x_1) (\mathbf{B}_1^{(j_1)}(x_1))^\top dx_1, \\ &= \mathcal{S}_p^{(i_p, j_p)} \otimes \dots \otimes \mathcal{S}_1^{(i_1, j_1)}, \end{aligned}$$

where the matrix $\mathcal{S}_h^{(i_h, j_h)}$ can be approximated using a trapezoidal rule. The \mathbf{r} vector can be constructed as the sum of elementary penalty vectors:

$$\begin{aligned} \mathbf{p}_{i_p, \dots, i_1} &= \int \mathbf{B}_p^{(i_p)}(x_p) \otimes \dots \otimes \mathbf{B}_1^{(i_1)}(x_1) dx_p \dots dx_1, \\ &= \int \mathbf{B}_p^{(i_p)}(x_p) dx_p \otimes \dots \otimes \int \mathbf{B}_1^{(i_1)}(x_1) dx_1, \\ &= \mathbf{s}_p^{(i_p)} \otimes \dots \otimes \mathbf{s}_1^{(i_1)}, \end{aligned}$$

as before, a trapezoidal rule can be applied to approximate the vector $\mathbf{s}_h^{(i_h)}$.

References

- Biegler, L., Damiano, J., Blau, G.: Nonlinear parameter estimation: a case study comparison. *AIChE J.* **32**(1), 29–45 (1986)
- Billier, C.: Adaptive bayesian regression splines in semiparametric generalized linear models. *J. Comp. Graph. Stat.* **9**(1), 122–140 (2000)
- Black, F., Scholes, M.: The pricing of options and corporate liabilities. *J. Polit. Econ.* **81**(3), 637–654 (1973)
- Botella, O.: On a collocation B-spline method for the solution of the Navier–Stokes equations. *Comp Fluids* **31**(4–7), 397–420 (2002)
- Carmona R.: *Statistical Analysis of Financial Data in S-Plus*. Springer Text in Statistics (2004)
- Cox, J., Ross, S., Rubinstein, M.: Option pricing: A simplified approach. *J. Finan. Econ.* **7**(3), 229–263 (1979)
- Currie, I.D.: Smoothing constrained generalized linear models with an application to the lee-carter model. *Stat. Model.* **13**(1), 69–93 (2013)
- Denison, D.G.T., Mallick, B.K., Smith, A.F.M.: Automatic bayesian curve fitting. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **60**(2), 333–350 (1998)
- Dierckx, P.: *Curve and Surface Fitting with Splines*. Oxford University Press (1995)
- Eilers, P.H.C., Marx, B.D.: Flexible smoothing with B-splines and penalties. *Stati. Sci.* **11**, 89–121 (1996)
- Eilers, P.H.C., Marx, B.D.: Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chem. Intellig. Lab. Syst.* **66**, 159–174 (2003)
- Eilers, P.H.C., Marx, B.D.: Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(6), 637–653 (2010)
- Epperson, J.F.: On the Runge Example. *Am. Math. Monthly* **94**(4), 329–341 (1987)
- Friedman, J.H., Silverman, B.W.: Flexible parsimonious smoothing and additive modeling. *Technometrics* **31**(1), 3–21 (1989)
- Golub, G.H., Ortega, J.M.: *Scientific computing and differential equations*. Academic Press, New York and London (1992)
- Grenander, U.: *Abstract Inference*. Probability and Statistics Series. Wiley (1981)
- Hastie, T.J., Tibshirani, R.J.: *Generalized additive models*. Chapman & Hall, London (1990)
- Heston, S.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev Finan Studies* **6**(2), 327–343 (1993)
- Holmes, C.C., Mallick, B.K.: Bayesian radial basis functions of variable dimension. *Neural Comp.* **10**(5), 1217–1233 (1998)
- Jaeger, J., Lambert, P.: Bayesian P-spline estimation in hierarchical models specified by systems of affine differential equations. *Stat. Model.* **13**(1), 3–40 (2013)
- Jaeger, J., Lambert, P.: Bayesian penalized smoothing approaches in models specified using differential equations with unknown error distributions. *J. App Stat.* 1–18 (2014)
- Ma, S., Kosorok, M.R.: Robust semiparametric m-estimation and the weighted bootstrap. *J. Mult. Anal.* **96**(1), 190–217 (2005)
- Piché, R., Kannianen, J.: *Solving financial differential equations using differentiation matrices*. Lecture Notes in Engineering and Computer Science. Newswood Limited, In World Congress on Engineering (2007)
- Ramsay, J.O., Hooker, G., Campbell, D., Cao, J.: Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society, Series B* **69**, 741–796 (2007)
- Rodriguez-Fernandez, M., Egea, J., Banga, J.: Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics* **7**(1), 483 (2006)
- Ruppert, D., Wand, M.P. and Carroll, R.J.: *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics **12**, (2003)
- Schall, R.: Estimation in generalized linear models with random effects. *Biometrika* **78**(4), 719–727 (1991)
- Schultz, M.H., Varga, R.S.: L-splines. *Numerische Mathematik* **10**(4), 345–369 (1967)
- Shen, X.: On methods of sieves and penalization. *Ann. Statist.* **25**(6), 2555–2591, 12 (1997)
- Smith, M., Yau, P., Shively, T., Kohn, R.: Estimating long-term trends in tropospheric ozone levels. *Int. Stat. Rev.* **70**(1), 99–124 (2002)
- Wu, Z.: Compactly supported positive definite radial functions. *Advances in Computational Mathematics*, **4**(1), 283–292, (1995) (ISSN 1019-7168)

- Xue, H., Miaou, H., Wu, H.: Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Ann. Statist.* **38**(4), 2351–2387 (2010)
- Xun, X., Cao, J., Mallick, B., Carroll, R., Maity, A.: Parameter estimation of partial differential equation models. *J. Am. Stat. Assoc.* **108**(503), 1009–1020 (2013)