



HAL
open science

A novel approach for estimating functions in the multivariate setting based on an adaptive knot selection for B-splines with an application to a chemical system used in geoscience

Mary E. Savino, Céline Lévy-Leduc

► To cite this version:

Mary E. Savino, Céline Lévy-Leduc. A novel approach for estimating functions in the multivariate setting based on an adaptive knot selection for B-splines with an application to a chemical system used in geoscience. 2024. hal-04117880v2

HAL Id: hal-04117880

<https://hal.science/hal-04117880v2>

Preprint submitted on 25 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ADAPTIVE KNOT SELECTION FOR ESTIMATING FUNCTIONS WITH MULTIVARIATE B-SPLINES

ABSTRACT. In this paper, we will outline a novel data-driven method for estimating functions in a multivariate nonparametric regression model based on an adaptive knot selection for B-splines. The underlying idea of our approach for selecting knots is to apply the generalized lasso, since the knots of the B-spline basis can be seen as changes in the derivatives of the function to be estimated. This method was then extended to functions depending on several variables by processing each dimension independently, thus reducing the problem to a univariate setting. The regularization parameters were chosen by means of a criterion based on EBIC. The nonparametric estimator was obtained using a multivariate B-spline regression with the corresponding selected knots. Our procedure was validated through numerical experiments by varying the number of observations and the level of noise to investigate its robustness. The influence of observation sampling was also assessed and our method was applied to a chemical system commonly used in geoscience. For each different framework considered in this paper, our approach performed better than state-of-the-art methods. Our completely data-driven method is implemented in the `glober` R package which is available on the Comprehensive R Archive Network (CRAN).

1. INTRODUCTION

In geochemical models, computing the concentrations of reactive species at equilibrium is well-known to be a challenging task especially when the number of species is large and/or when the reactions involve the dissolution or the precipitation of minerals, see White et al. (1958), Smith (1980) and de Capitani and Brown (1987) for further details. The numerical resolution of these non-linear problems can become so time consuming that coupling them with other physical processes may require to be simplified. For instance in the case of reactive transport, the size of the geometric model has to be drastically limited. To overcome this issue, researchers have been focusing their work on improving the numerical scheme to speed up computations.

However, despite the significant improvements of the numerical solvers and preconditioners over the past few decades, solving three dimensional large scale modelling of complex reactive transport over many time steps is still nearly impossible using standard computers. Consequently, geoscientists are more and more interested in devising approaches which can provide an estimation of the solution of the full simulation model (sometimes also called surrogate model) from a limited set of observations obtained with the full simulation model from specific input values that can thus replace it. Hence, the problem can be reformulated as the estimation of an unknown function f in the following regression model:

$$Y_i = f(x_i) + \varepsilon_i, \quad 1 \leq i \leq n, \quad (1)$$

where the ε_i are i.i.d centered random variables of variance σ^2 and the x_i are observation points which belong to a compact set \mathcal{S} of \mathbb{R}^d , $d \geq 1$. In the reactive transport modelling field (RTM), several surrogate models have been proposed, we refer the reader to Asher et al. (2015) and Jatnieks et al. (2016) for a comparison of the different approaches. Artificial neural networks

Key words and phrases. function estimation, B-splines, generalized lasso, geochemical applications.

have recently gained a huge interest in RTM (Guérillot and Bruyelle (2020)), more especially through Deep Neural Networks (DNNs), since their approximations have a high accuracy compared to other estimators (Laloy and Jacques (2019)). Nevertheless, despite all the effort for improving the efficiency of DNNs via the conjunction of computational advancements for training ever-larger networks and improvements of backpropagation algorithms, DNNs still remain difficult to exploit when the quantity of training data is not sufficient (Karpatne et al. (2018)) especially when a high number of parameters needs to be calibrated. Recently, Savino et al. (2022) proposed an active learning approach to drastically decrease the number of training observations to use by modeling the function to estimate as a sample of Gaussian Processes. This method has given promising results but is not necessarily the most suitable approach for noisy observation sets. In order to circumvent this limitation, nonparametric estimation approaches based on splines are known to be an efficient tool, see Wahba (1990) for further details on this kind of methods.

Nonparametric estimation approaches based on splines consist in approximating the function to estimate by a linear combination of splines which are functions defined by pre-selecting a well chosen set of knots. In this framework, Friedman (1991) proposed an efficient approach called Multivariate Adaptive Regression Splines (MARS) which can be used when the function to estimate has several input variables. However, MARS has not shown better performance than other state-of-the-art methods on a concrete RTM application displayed in Jatnieks et al. (2016). A theoretical and experimental comparison has been undertaken by Eckle and Schmidt-Hieber (2019) and demonstrated that DNNs can outperform MARS but with a specific number of parameters and they do not necessarily give better results for every numerical application. This conclusion was also drawn by Zhang and Goh (2013) and Zhang and Goh (2016) in which the authors demonstrated an equivalent accuracy and performance between a back-propagation neural network architecture and MARS on geotechnical applications but a better interpretability and a higher computational efficiency was demonstrated for the latter.

Other articles proposed approximating the function to estimate by a linear combination of B-splines defined in De Boor (1978) since they display an attractive stability and a computational efficiency. Their ability to approximate complicated functions while being insensitive to noisy observation sets have made them very interesting in the past few decades. Since their definition depends on a pre-defined sequence of knot locations, many strategies have been developed to optimize the selection of these points in order to avoid overfitting and so to ensure the best approximation of the underlying function. O’Sullivan (1986) described an innovative method, introduced as O-splines by Wand and Ormerod (2008), to estimate a function f by selecting simultaneously the number and the locations of knots from an arbitrary set of values. Its main goal was to penalize an integrated square of the second order derivative also called roughness in order to determine the coefficients of the linear combination of B-splines. However, the computation of this method was tedious with higher order derivatives. To circumvent this issue, Eilers and Marx (1996) proposed a discrete version of this method called P-splines which uses a discrete penalty matrix and a ℓ_2 -norm penalized least-square criterion (ridge approach) to determine the coefficients of the B-splines defined from evenly-spaced knots (see Wand and Ormerod (2008) for a detailed comparison between O-splines and P-splines). These P-splines have been used in an impressive list of work of curve fitting (see Eilers et al. (2015) for a review) and have been extended to the multivariate setting, see Eilers and Marx (2003) for an application to smoothing two-dimensional signals. Li and Cao (2022) have adapted these P-splines to apply them to unevenly-spaced knots by defining a general weighted difference penalty matrix adapted to regular and irregular knot spacing. To drastically limit the

number of knots, Goepp et al. (2018) have proposed a weighted adaptive ridge method called A-splines which aims at discarding the less relevant knots and by defining new B-splines from the selected knots. This method appeared to be more interpretable than the P-splines method but their statistical performance is equivalent.

Another approach was introduced for B-spline curve fitting by Yuan et al. (2013). Their idea is to first select the most pertinent B-splines from a multi-resolution basis by applying the Lasso criterion to get the locations of the knots. Then, after a pruning step to reduce once again the number of knots, the final B-splines are built from these small sets of knots. The estimation of the function is obtained by fitting a linear combination of these B-splines to the observations by a least-square approach. This method seems to have promising results but is not available for two-dimensional functions yet.

In this paper, we propose a novel data-driven approach for estimating the function f in the multivariate nonparametric regression model (1) based on an adaptive knot selection for B-splines. Since the knots of a B-spline basis can be seen as changes in the derivatives of f , we propose finding the most relevant ones, based on the work of Denis et al. (2020), by using the generalized lasso described in Tibshirani and Taylor (2011) and further studied in Tibshirani (2014). A B-spline basis is then defined from these selected knots and a least-square approach is undertaken to determine the coefficients of the linear combination of B-splines. Sadhanala et al. (2021) have proposed a multivariate version of trend filtering (a specific generalized lasso form) called Kronecker trend filtering (KTF) to extend it to smoothing functions with multiple input variables. It implies the use of a huge penalty matrix defined as the Kronecker product of univariate trend filtering penalty operators and of a unique regularization parameter common to every input variables. In order to drastically reduce the dimensions of the difference penalty matrix and to allow a better flexibility in the regularization step, the extension of our method to functions with two input variables is presented by simply considering each dimension separately to reduce the problem to the one-dimensional setting. We also propose a way to extend our approach to higher dimensional settings where the observation points do not necessarily come from a cartesian product of the sets in each dimension.

This paper is organized as follows. Section 2 describes the methodology that we propose for our adaptive knot selection method for the one and two-dimensional settings. Section 3 investigates the performance of our approach through numerical experiments. In Section 4, we apply our method to the data that motivated this study. Finally, in Section 5, we extend our approach to more general observation point settings.

2. METHODOLOGY

In this section, we describe our innovative nonparametric method to estimate the function f defined in (1). We will introduce our method first for one-dimensional functions ($d = 1$), then in a second section we will extend it to the two-dimensional case ($d = 2$).

2.1. Description of our method in the one-dimensional case. We propose estimating the function f appearing in (1) by approximating it with a linear combination of B-splines of order M ($M \geq 1$) introduced by De Boor (1978) in Chapter 9.

Let $\mathbf{t} = (t_1, \dots, t_K)$ be a set of K points called knots which are crucial in the definition of the B-spline basis. We define the augmented knot sequence $\boldsymbol{\tau}$ such that:

$$\begin{aligned} \tau_1 &= \dots = \tau_M = x_{min}, \\ \tau_{j+M} &= t_j, \quad j = 1, \dots, K, \end{aligned}$$

$$x_{max} = \tau_{K+M+1} = \dots = \tau_{K+2M},$$

$$\boldsymbol{\tau} = (\tau_1, \dots, \tau_{K+2M}) = \underbrace{(x_{min}, \dots, x_{min})}_{M \text{ times}}, \underbrace{(t_1, \dots, t_K)}_{\mathbf{t}}, \underbrace{(x_{max}, \dots, x_{max})}_{M \text{ times}},$$

where x_{min} and x_{max} are the lower and upper bounds of \mathcal{S} , respectively.

B-splines are defined by (De Boor, 1978, p. 89-90) and (Hastie et al., 2009, p. 160) as follows. Denoting by $B_{i,m}(x)$ the i th B-spline basis function of order m for the knot sequence $\boldsymbol{\tau}$ with $m \leq M$, they are defined by the following recursion:

$$B_{i,1}(x) = \begin{cases} 1 & \text{if } \tau_i \leq x < \tau_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i = 1, \dots, K + 2M - 1, \quad (2)$$

and for $m \leq M$,

$$B_{i,m}(x) = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_{i+m} - x}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}(x), \quad (3)$$

for $i = 1, \dots, (K + 2M - m)$.

In the next section we will describe how to choose the set of knots \mathbf{t} to estimate f .

2.1.1. *Creation of a candidate set of knots.* Let $\mathbf{Y} = (Y_1, \dots, Y_n)$ and $\mathbf{x} = (x_1, \dots, x_n)$ where Y_i and x_i are defined in (1). In the following, we shall assume that $x_1 < \dots < x_n$ and $M = q + 1$, with $q \geq 0$. Hence, when $q = 0$ (resp. $q = 1$, $q = 2$) f is approximated with piecewise constant (resp. linear, quadratic) functions.

Since the knots of a B-spline basis can be seen as changes in the $(q + 1)$ th derivative of f , we propose finding them by using the generalized Lasso described in Tibshirani and Taylor (2011) and further studied in Tibshirani (2014). In the latter, they define the polynomial trend filtering which consists in approximating f by $\widehat{\boldsymbol{\beta}}(\lambda)$ defined as follows:

$$\widehat{\boldsymbol{\beta}}(\lambda) = \underset{\boldsymbol{\beta} \in \mathbb{R}^n}{\operatorname{argmin}} \{ \|\mathbf{Y} - \boldsymbol{\beta}\|_2^2 + \lambda \|D \boldsymbol{\beta}\|_1 \}, \quad (4)$$

where $\|y\|_2^2 = \sum_{i=1}^n y_i^2$ for $y = (y_1, \dots, y_n)$ and $\|u\|_1 = \sum_{i=1}^m |u_i|$ for $u = (u_1, \dots, u_m)$, λ is a positive constant which has to be tuned and $D \in \mathbb{R}^{m \times n}$ is a specified penalty matrix, defined recursively as follows:

$$D = D_{tf,q+1} = D_0 \cdot D_{tf,q} \quad q \geq 0, \quad (5)$$

where “ tf ” is the abbreviation of “trend filtering”, $(q + 1)$ is the order of differentiation, $D_{tf,0} = \operatorname{Id}_{\mathbb{R}^n}$, the identity matrix of \mathbb{R}^n , and D_0 is the penalty matrix for the one-dimensional fused Lasso:

$$D_0 = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}.$$

The penalty matrix D is the discrete difference operator of order $(q + 1)$ and thus, $D\widehat{\boldsymbol{\beta}}$ estimates the $(q + 1)$ st order derivative of f . Hence, observing the locations where $D\widehat{\boldsymbol{\beta}} \neq 0$ provides a way of finding the B-spline knots.

The matrix D is well-adapted when the observation points are evenly spaced. When it is not the case, it should be replaced by the following matrix $\Delta^{(q+1)}$ defined recursively as follows:

$$\Delta^{(q+1)} = \mathbf{W}_{(q+1)} \cdot D_0 \cdot \Delta^{(q)}, \quad q \geq 0,$$

where $\Delta^{(0)} = \text{Id}_{\mathbb{R}^n}$ and $\mathbf{W}_{(q+1)}$ is the diagonal weight matrix defined by:

$$\mathbf{W}_{(q+1)} = \text{diag} \left(\frac{1}{(x_{(q+1)+1} - x_{(q+1)})}, \frac{1}{(x_{(q+1)+2} - x_{(q+1)+1})}, \dots, \frac{1}{(x_n - x_{n-1})} \right).$$

In both cases (evenly or unevenly-spaced observations), the number of rows of D and $\Delta^{(q+1)}$ equals $m = n - q - 1$.

Let us now more precisely explain how to choose the B-spline knots. Let $\Lambda = (\lambda_1, \dots, \lambda_k)$ be a grid of penalization parameters λ_i . We define the resulting differentiated column vector $\mathbf{a}(\lambda)$ by:

$$\mathbf{a}(\lambda) = \Delta^{(q+1)} \cdot \widehat{\boldsymbol{\beta}}(\lambda), \quad (6)$$

where $\widehat{\boldsymbol{\beta}}(\lambda)$ is the solution of problem (4) when $D = \Delta^{(q+1)}$ and λ belongs to Λ .

The ordered vector of selected knots associated to λ is defined as follows:

$$\widehat{\mathbf{t}}_\lambda = (\widehat{t}_j)_{j=1, \dots, K_\lambda} = (x_{p_j})_{j=1, \dots, K_\lambda}, \quad \text{with } p_j \in \mathcal{P}_\lambda, \quad (7)$$

where

$$\mathcal{P}_\lambda = \{\ell + 1, a_\ell(\lambda) \neq 0\} \quad \text{and} \quad K_\lambda = \sum_{\ell=1}^m \mathbb{1}\{a_\ell(\lambda) \neq 0\}, \quad (8)$$

$a_\ell(\lambda)$ denoting the ℓ th component of $\mathbf{a}(\lambda)$ and $\mathbb{1}\{A\} = 1$ if the event A holds and 0 if not.

The corresponding B-spline basis $B_{i,M}$ is defined by replacing the t_j in the augmented knot sequence $\boldsymbol{\tau}$ appearing in (2) and (3) by \widehat{t}_j found in (7). Thus, we obtain the following estimator of f for each λ of Λ :

$$\widehat{f}_\lambda(x) = \sum_{i=1}^{q+K_\lambda+1} \widehat{\gamma}_i B_{i,M}(x), \quad (9)$$

where $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_i)_{1 \leq i \leq q+K_\lambda+1}$ is obtained using the following least-square criterion:

$$\widehat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{q+K_\lambda+1}}{\text{argmin}} \|\mathbf{Y} - \mathbf{B}(\lambda) \boldsymbol{\gamma}\|_2^2, \quad (10)$$

where $\mathbf{B}(\lambda)$ is a $n \times (q + K_\lambda + 1)$ matrix having as i th column $(B_{i,M}(x_k))_{1 \leq k \leq n}$, i belonging to $\{1, \dots, q + K_\lambda + 1\}$.

2.1.2. Choice of the penalization parameter of the regularized method. In order to choose the penalization parameter λ which leads to the best selection of knots, we use a criterion defined by Chen and Chen (2008) and recommended in Goepp et al. (2018), namely the extended Bayesian information criterion also called EBIC:

$$\text{EBIC}(\lambda) = \text{SS}(\lambda) + (q + K_\lambda + 1) \log n + 2 \log \binom{q + K_{\max} + 1}{q + K_\lambda + 1}, \quad (11)$$

where K_{\max} is the maximum number of knots that we can select (here $K_{\max} = n$) and $\text{SS}(\lambda)$ is the sum of squares defined by:

$$\text{SS}(\lambda) = \|\mathbf{Y} - \widehat{\mathbf{Y}}(\lambda)\|_2^2, \quad (12)$$

where

$$\widehat{\mathbf{Y}}(\lambda) = \mathbf{B}(\lambda) \widehat{\boldsymbol{\gamma}},$$

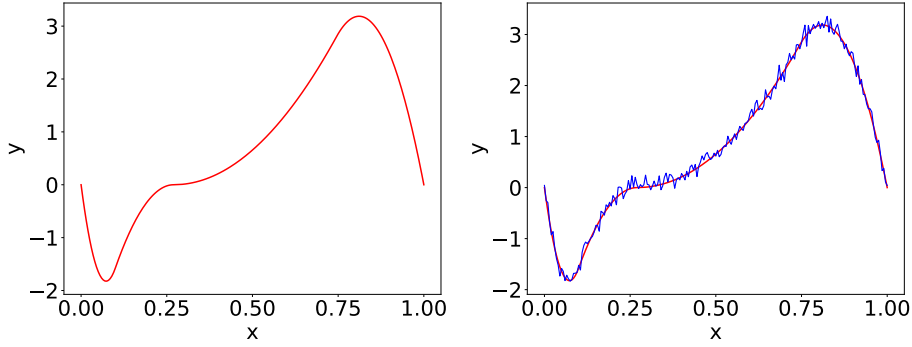


FIGURE 1. Function f_1 to estimate (left) and a noisy set of observations Y_1, \dots, Y_{201} with $\sigma = 0.1$ (right).

with $\hat{\gamma}$ and $\mathbf{B}(\lambda)$ being defined in (10). This criterion allows us to get a trade-off between a good approximation of the underlying function without using too many parameters. The final estimator of f is defined as follows:

$$\hat{f}(x) = \hat{f}_{\lambda_{\text{EBIC}}}(x), \quad (13)$$

where $\hat{f}_{\lambda}(x)$ is defined in (9) and

$$\lambda_{\text{EBIC}} = \underset{\lambda \in \Lambda}{\operatorname{argmin}} \{ \text{EBIC}(\lambda) \}. \quad (14)$$

2.1.3. *Illustration of our method on a simple case.* In order to illustrate our method we apply it to a noisy set of observations $\mathbf{Y} = (Y_1, \dots, Y_n)$ where the Y_i are defined in (1) and $f = f_1$ is a linear combination of quadratic B-splines ($M = 3$) with $\mathbf{t} = (0.1, 0.27, 0.745)$ defined as follows:

$$f_1(x) = -2.5B_{2,3}(x) + 4.3B_{5,3}(x), \quad x \in [0, 1]. \quad (15)$$

In (1), the ε_i are i.i.d Gaussian centered random variables with $\sigma = 0.1$. The set of knots \mathbf{t} belongs to the observation set $\{x_1, \dots, x_n\}$. The corresponding $(f_1(x_i))_{1 \leq i \leq n}$ and $(Y_i)_{1 \leq i \leq n}$ are displayed in Figure 1 for $n = 201$. Since we want to approximate quadratic B-splines, we must choose q such that the method can detect the changes in the third derivative so here $q + 1 = 3$. In order to assess the performance of our knot selection procedure, we compute the Hausdorff distance defined as follows:

$$d(\mathbf{t}, \hat{\mathbf{t}}_{\lambda}) = \max \left(d_1(\mathbf{t}, \hat{\mathbf{t}}_{\lambda}), d_2(\mathbf{t}, \hat{\mathbf{t}}_{\lambda}) \right), \quad (16)$$

where

$$d_1(\mathbf{u}, \mathbf{v}) = \sup_{v \in \mathbf{v}} \inf_{u \in \mathbf{u}} |u - v|,$$

$$d_2(\mathbf{u}, \mathbf{v}) = d_1(\mathbf{v}, \mathbf{u}).$$

Figure 2 displays the boxplots of the first and second part of the Hausdorff distance and of the number of selected knots K_{λ} for $\lambda = \lambda_{\text{EBIC}}$ obtained from 10 different samplings of x_1, \dots, x_n . The first boxplots are obtained for $n = 7$ then new observation points are randomly added to the current observation sets in order to have an increasing number of observations such that $n \leq 100$. We can see from this figure that from $n = 70$ the second part of the Hausdorff distance is close to 0 which means that the estimated knots are near from the real ones. These results are obtained with an almost constant number of selected knots $K_{\lambda_{\text{EBIC}}} = 6$.

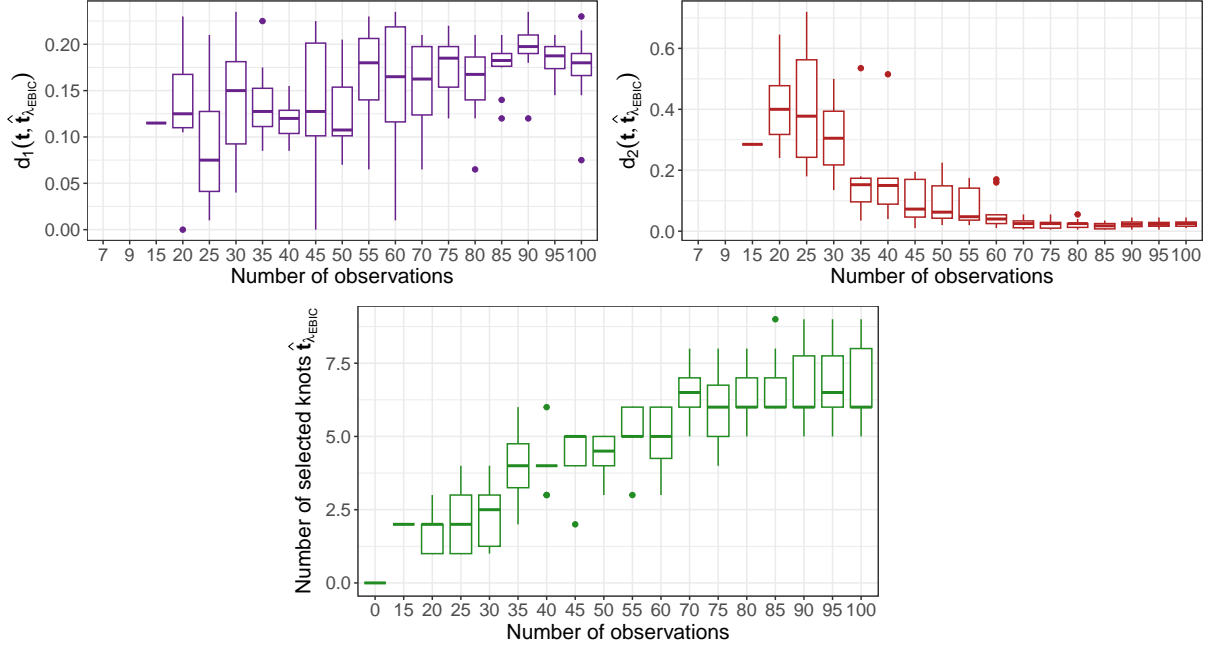


FIGURE 2. Top left: Boxplots for the first part of the Hausdorff distance as a function of n . Top right: boxplots for the second part of the Hausdorff distance as a function of n . Bottom: number of estimated knots as a function of n with $\lambda = \lambda_{\text{EBIC}}$ for estimating f_1 .

For a comparison purpose, we displayed in Figure 3 the results obtained for $\lambda = \lambda_{\text{opt}}$, where λ_{opt} is defined by:

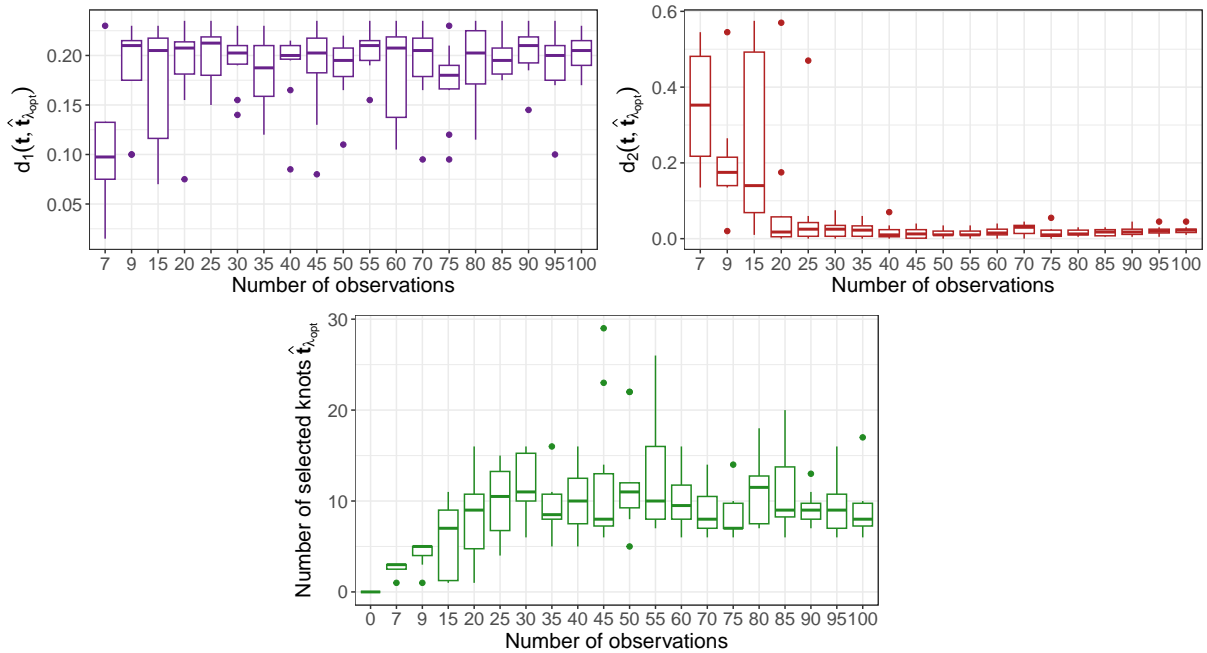
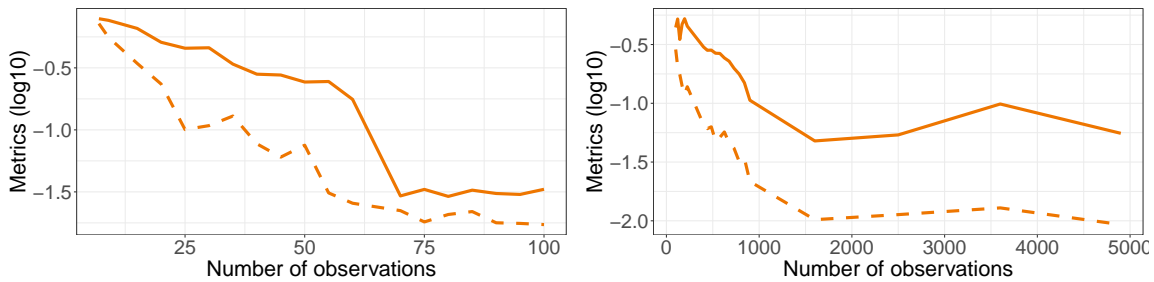
$$\lambda_{\text{opt}} = \underset{\lambda \in \Lambda}{\text{argmin}} (\text{Normalized sup norm}(\lambda)),$$

with

$$\text{Normalized sup norm}(\lambda) = \max_{1 \leq k \leq N} \frac{|f(x_k) - \hat{f}_\lambda(x_k)|}{f_{\max} - f_{\min}}, \quad (17)$$

where \hat{f}_λ is defined in (9). In (17), N ($N > n$) is the cardinality of the set of evenly-spaced points $\{x_1, \dots, x_N\}$ of $[0, 1]$ which contains the observation points x_1, \dots, x_n as well as additional points where f has not been observed. Moreover, f_{\min} and f_{\max} denote the minimum and maximum values of f evaluated on $\{x_1, \dots, x_N\}$, respectively. We can see from this figure that the performance obtained when λ is optimally chosen is on a par with that of λ_{EBIC} which means that our procedure for choosing λ is almost optimal.

The corresponding performance is shown on the right part of Figure 4 for $N = 201$. This figure displays the average of the most stringent metric (Normalized Sup Norm) obtained from 10 different samplings of x_1, \dots, x_n for each n . We observed from this figure that the Normalized Sup Norm reaches $10^{-1.75}$ (resp. $10^{-1.5}$) for λ_{opt} (resp. λ_{EBIC}) which represents a normalized maximum absolute error of 2% (resp. 3%). Once again, these results show that the choice of λ does not alter the performance of our approach.

FIGURE 3. Similar to Figure 2 by choosing $\lambda = \lambda_{\text{opt}}$ for estimating f_1 .FIGURE 4. Statistical performance (Normalized Sup Norm) of the method using λ_{opt} (dashed) and λ_{EBIC} (solid) for the estimation of f_1 (right) and f_2 (left) obtained from 10 replications.

2.2. Extension to the two-dimensional case. In this section, we will extend the previous method for estimating a two-dimensional function f from the observations $(Y_i)_{1 \leq i \leq n}$ defined in Model (1) when $d = 2$.

Here \mathcal{S} is defined as the cartesian product of two compact sets \mathcal{S}_1 and \mathcal{S}_2 of \mathbb{R} . More precisely, we will consider $(x_{11}, \dots, x_{1n_1})$ belonging to \mathcal{S}_1 and similarly $(x_{21}, \dots, x_{2n_2})$ belonging to \mathcal{S}_2 the n_1 and n_2 observation values for the first and second variables of f at which f is evaluated, respectively. Thus, the set of observations belonging to \mathcal{S} will be defined as:

$$\begin{aligned} & ((x_{11}, x_{21}), (x_{11}, x_{22}), \dots, (x_{11}, x_{2n_2}), (x_{12}, x_{21}), (x_{12}, x_{22}), \dots, (x_{12}, x_{2n_2}), \dots, (x_{1n_1}, x_{2n_2})) \\ & = ((x_{1k}, x_{2\ell}))_{1 \leq k \leq n_1, 1 \leq \ell \leq n_2}. \end{aligned}$$

For estimating f , we shall approximate it by a linear combination of multidimensional B-splines defined in (Hastie et al., 2009, p. 162-163) as the tensor product of the B-splines of order M introduced in Section 2.1. More precisely, $f(x) = f(x_1, x_2)$ will be approximated by:

$$\sum_{i=1}^{Q_1} \sum_{j=1}^{Q_2} \gamma_{ij} B_{1,i,M}(x_1) B_{2,j,M}(x_2), \quad (18)$$

where $B_{1,i,M}$ and $B_{2,j,M}$ are the B-spline basis of order M defined in (3) for the first and second dimension, respectively. In (18), $Q_1 = q + K_1 + 1$, $Q_2 = q + K_2 + 1$ with K_1 and K_2 the number of knots defined in the B-spline basis of the first and second variables, respectively and $M = q + 1$.

2.2.1. *Creation of a candidate set of knots.* The idea is to consider the two dimensions independently and thus, by fixing one dimension at a time, the problem can be rewritten as an estimation problem in the one-dimensional framework.

First, we shall consider the knot selection of the B-spline basis of the first dimension by fixing the second dimension to a certain value of x_2 belonging to $\{x_{21}, \dots, x_{2n_2}\}$. Thus, we can apply the polynomial trend filtering method described in Section 2.1 and get the grid of penalization parameters $(\lambda_{(1,i),k})_{1 \leq k \leq s_i}$, with i belonging to $\{1, \dots, n_2\}$ and s_i corresponds to the number of penalization parameters. The index 1 in $(1, i)$ denotes the first dimension and i indexes the i th value of $\{x_{21}, \dots, x_{2n_2}\}$. For each value $\lambda_{(1,i),k}$ with k belonging to $\{1, \dots, s_i\}$, we can get the corresponding selected knots by following the procedure described in Section 2.1.2: after calculating $\mathbf{a}(\lambda_{(1,i),k})$ as in (6), we can determine the set of knots $\hat{\mathbf{t}}_{1,\lambda_{(1,i),k}}$ as in (7). In order to take into account all the information obtained for each value of x_2 , we gather the selected knots into a single vector depending on the value of the penalization parameters. Nevertheless, because not all the vectors $(\lambda_{(1,i),k})_{1 \leq k \leq s_i}$ have exactly the same values $\lambda_{(1,i),k}$ and the same number of penalization parameters s_i when i varies, we shall define the set of *equivalent regularization parameters* $\tilde{\Lambda}_1$ and the minimal number of penalization parameters s_{min_1} :

$$\tilde{\Lambda}_1 = \{\tilde{\lambda}_{1,1}, \dots, \tilde{\lambda}_{1,s_{min_1}}\} \quad \text{and} \quad s_{min_1} = \min_{1 \leq i \leq n_2} s_i, \quad (19)$$

where

$$\tilde{\lambda}_{1,k} = (\lambda_{(1,i),k})_{1 \leq i \leq n_2}, \quad 1 \leq k \leq s_{min_1}. \quad (20)$$

In (20), $\tilde{\lambda}_{1,k}$ can be seen as the vector of parameters which penalize (4) at an equivalent strength for each fixed value of x_2 . We can therefore get the vector of selected knots $\hat{\mathbf{t}}_{1,\tilde{\lambda}_{1,k}}$ for the first dimension by grouping together and ordering all the corresponding selected knots of $\tilde{\lambda}_{1,k}$.

We proceed the same way to get the set of equivalent parameters $\tilde{\Lambda}_2$ for the second dimension by fixing this time the value of x_1 and with s_{min_2} defined similarly as in (19) for i belonging to $\{1, \dots, n_1\}$. Analogously as in (19) and (20), we have:

$$\tilde{\Lambda}_2 = \{\tilde{\lambda}_{2,1}, \dots, \tilde{\lambda}_{2,s_{min_2}}\} \quad \text{and} \quad \tilde{\lambda}_{2,\ell} = (\lambda_{(2,i),\ell})_{1 \leq i \leq n_1}, \quad 1 \leq \ell \leq s_{min_2}.$$

Moreover, as well as for the first dimension, the vector of selected knots for the second dimension for each $\tilde{\lambda}_{2,\ell}$ is defined as $\hat{\mathbf{t}}_{2,\tilde{\lambda}_{2,\ell}}$, ℓ belonging to $\{1, \dots, s_{min_2}\}$.

In the following, let us consider two generic penalization parameters $\tilde{\lambda}_1$ belonging to $\tilde{\Lambda}_1$ and $\tilde{\lambda}_2$ belonging to $\tilde{\Lambda}_2$. Thus, we can define the candidate sets of knots for both dimensions $\hat{\mathbf{t}}_{1,\tilde{\lambda}_1}$ and $\hat{\mathbf{t}}_{2,\tilde{\lambda}_2}$. We must now determine which combination of penalization parameters $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ and hence, which combination of selected knots for the first and second dimension, allows us to get an optimal estimator of f .

2.2.2. *Choice of the penalization parameters of the regularized method.* In order to choose the penalization parameters leading to the best selection of knots, we consider the following EBIC criterion which can be seen as the adaptation to the two-dimensional case of the one defined in (11):

$$\text{EBIC}(\tilde{\lambda}_1, \tilde{\lambda}_2) = \text{SS}(\tilde{\lambda}_1, \tilde{\lambda}_2) + \tilde{Q}_1 \tilde{Q}_2 \log n + 2 \log \binom{(q+n_1+1)(q+n_2+1)}{\tilde{Q}_1 \tilde{Q}_2}. \quad (21)$$

where $\tilde{Q}_1 = q + K_{\tilde{\lambda}_1} + 1$ and $\tilde{Q}_2 = q + K_{\tilde{\lambda}_2} + 1$, $K_{\tilde{\lambda}_1}$ and $K_{\tilde{\lambda}_2}$ being the number of selected knots with the parameters $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ for the first and second dimension, respectively.

In (21), $\text{SS}(\tilde{\lambda}_1, \tilde{\lambda}_2)$ is defined as:

$$\text{SS}(\tilde{\lambda}_1, \tilde{\lambda}_2) = \left\| \mathbf{Y} - \hat{\mathbf{Y}}(\tilde{\lambda}_1, \tilde{\lambda}_2) \right\|_2^2,$$

where

$$\hat{\mathbf{Y}}(\tilde{\lambda}_1, \tilde{\lambda}_2) = \mathbf{B}(\tilde{\lambda}_1, \tilde{\lambda}_2) \hat{\boldsymbol{\gamma}}, \quad (22)$$

and $\mathbf{B}(\tilde{\lambda}_1, \tilde{\lambda}_2)$ is defined as:

$$\mathbf{B}(\tilde{\lambda}_1, \tilde{\lambda}_2) = \mathbf{B}(\tilde{\lambda}_1) \otimes \mathbf{B}(\tilde{\lambda}_2), \quad (23)$$

$E \otimes F$ denoting the Kronecker product of the matrices E and F . In (23), $\mathbf{B}(\tilde{\lambda}_1)$ is a $n_1 \times \tilde{Q}_1$ matrix having as i th column $(B_{1,i,M}(x_{1k}))_{1 \leq k \leq n_1}$, i belonging to $\{1, \dots, \tilde{Q}_1\}$ and $\mathbf{B}(\tilde{\lambda}_2)$ is a $n_2 \times \tilde{Q}_2$ matrix having as j th column $(B_{2,j,M}(x_{2\ell}))_{1 \leq \ell \leq n_2}$, j belonging to $\{1, \dots, \tilde{Q}_2\}$.

In (22), $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_{ij})_{1 \leq i \leq \tilde{Q}_1, 1 \leq j \leq \tilde{Q}_2}$ is obtained using the following least-square criterion:

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{\tilde{Q}_1 \tilde{Q}_2}}{\text{argmin}} \left\| \mathbf{Y} - \mathbf{B}(\tilde{\lambda}_1, \tilde{\lambda}_2) \boldsymbol{\gamma} \right\|_2^2. \quad (24)$$

As in Equation (14), we shall define $\tilde{\lambda}_{1,\text{EBIC}}$ and $\tilde{\lambda}_{2,\text{EBIC}}$ the penalization parameters which verify:

$$\left(\tilde{\lambda}_{1,\text{EBIC}}, \tilde{\lambda}_{2,\text{EBIC}} \right) = \underset{\tilde{\lambda}_1 \in \tilde{\Lambda}_1, \tilde{\lambda}_2 \in \tilde{\Lambda}_2}{\text{argmin}} \left\{ \text{EBIC}(\tilde{\lambda}_1, \tilde{\lambda}_2) \right\}.$$

Hence, the final estimator of f is defined as:

$$\hat{f}(x_1, x_2) = \hat{f}_{\tilde{\lambda}_{1,\text{EBIC}}, \tilde{\lambda}_{2,\text{EBIC}}}(x_1, x_2),$$

with $\hat{f}_{\tilde{\lambda}_1, \tilde{\lambda}_2}$ defined as:

$$\hat{f}_{\tilde{\lambda}_1, \tilde{\lambda}_2}(x) = \hat{f}_{\tilde{\lambda}_1, \tilde{\lambda}_2}(x_1, x_2) = \sum_{i=1}^{\tilde{Q}_1} \sum_{j=1}^{\tilde{Q}_2} \hat{\gamma}_{ij} B_{1,i,M}(x_1) B_{2,j,M}(x_2). \quad (25)$$

In (25), $\hat{\gamma} = (\hat{\gamma}_{ij})_{1 \leq i \leq \tilde{Q}_1, 1 \leq j \leq \tilde{Q}_2}$ is obtained as in (24).

2.2.3. Illustration of our method on a simple case. To illustrate the extension of our method to the two-dimensional case, we propose estimating a function $f = f_2$ which is a linear combination of tensor product of quadratic B-splines ($M = 3$) with $\mathbf{t}_1 = (0.24, 0.545)$ and $\mathbf{t}_2 = (0.395, 0.645)$:

$$f_2(x_1, x_2) = 2.3B_{1,3,3}(x_1)B_{2,3,3}(x_2) - 1.5B_{1,4,3}(x_1)B_{2,5,3}(x_2), \quad (x_1, x_2) \in [0, 1]^2, \quad (26)$$

where $B_{i,j,M}$ is defined in (18) with \mathbf{t}_1 and \mathbf{t}_2 the knots involved in the definition of $B_{1,j,M}$ and $B_{2,j,M}$, respectively. We shall apply our method to a noisy set of observations $\mathbf{Y} = (Y_1, \dots, Y_n)$ where the Y_i are defined in (1) and the ε_i are i.i.d Gaussian centered random variables with $\sigma = 0.01$. The set of knots \mathbf{t}_1 and \mathbf{t}_2 are a part of the observation set $\{x_{11}, \dots, x_{1n_1}\}$ and $\{x_{21}, \dots, x_{2n_2}\}$, respectively. The corresponding $(Y_i)_{1 \leq i \leq n}$ (resp. $(f_2(x_{1,k}, x_{2,\ell}))_{1 \leq k \leq n_1, 1 \leq \ell \leq n_2}$) are displayed in the right (resp. left) part of Figure 5 for $n = n_1 n_2 = 201^2 = 40401$.

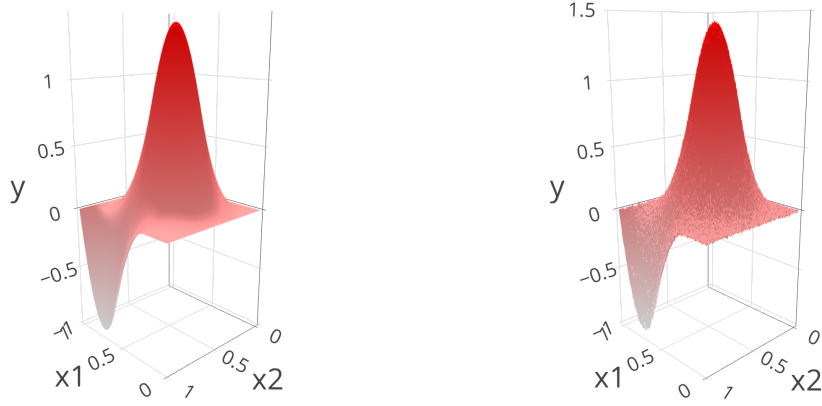


FIGURE 5. Function f_2 to estimate (left) and a noisy set of observations Y_1, \dots, Y_{40401} with $\sigma = 0.01$ (right).

In order to assess the performance of our knot selection procedure, we shall use the Hausdorff distance defined in (16) for each dimension independently. The results for the first and second part of the Hausdorff distance and the number of selected knots for both dimensions are displayed in the boxplots of Figure 18 of the Appendix for $\tilde{\lambda}_1 = \tilde{\lambda}_{1,\text{EBIC}}$ and $\tilde{\lambda}_2 = \tilde{\lambda}_{2,\text{EBIC}}$ and from 10 different samplings of x_{11}, \dots, x_{1n_1} and x_{21}, \dots, x_{2n_2} . New observation points are then randomly added to the current observation sets in order to have an increasing number of observations. We can see from this figure that from $n = 1600$ and so from 40 observation points by dimension, the second part of the Hausdorff distance is close to 0 which means that the estimated knots are near from the real ones. The numbers of selected knots required to get these results are between 5 and 10. Similarly as for the one-dimensional case, we compare these results with those obtained for $\tilde{\lambda}_1 = \tilde{\lambda}_{1,\text{opt}}$ and $\tilde{\lambda}_2 = \tilde{\lambda}_{2,\text{opt}}$, two optimal parameters defined as:

$$\left(\tilde{\lambda}_{1,\text{opt}}, \tilde{\lambda}_{2,\text{opt}} \right) = \underset{\tilde{\lambda}_1 \in \tilde{\Lambda}_1, \tilde{\lambda}_2 \in \tilde{\Lambda}_2}{\text{argmin}} \left\{ \text{Normalized sup norm} \left(\tilde{\lambda}_1, \tilde{\lambda}_2 \right) \right\}$$

with Normalized sup norm being defined in (17) depending here on the values of $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$, with \hat{f}_λ becoming $\hat{f}_{\tilde{\lambda}_1, \tilde{\lambda}_2}$ defined in (25) and x_k belonging to the set:

$$\begin{aligned} \{x_1, \dots, x_N\} = \\ \left\{ (x_{11}, x_{21}), (x_{11}, x_{22}), \dots, (x_{11}, x_{2N_2}), (x_{12}, x_{21}), (x_{12}, x_{22}), \dots, (x_{12}, x_{2N_2}), \dots, (x_{1N_1}, x_{2N_2}) \right\}. \end{aligned} \quad (27)$$

N is the cardinality of $\{x_1, \dots, x_N\}$ and is such that $N = N_1 N_2$. We can see from Figure 19 of the Appendix, where the results are displayed, that they are comparable to those found for $\tilde{\lambda}_{1,\text{EBIC}}$ and $\tilde{\lambda}_{2,\text{EBIC}}$. This means that our choice of the penalization parameters does not alter the performance of our approach. The corresponding performance is shown on the right part of Figure 4 for $N = 40401$ and from 10 different samplings of x_{11}, \dots, x_{1n_1} and x_{21}, \dots, x_{2n_2} . The most stringent metric (Normalized Sup Norm) reaches 10^{-2} (resp. $10^{-1.4}$) for $\lambda_{1,\text{opt}}$ and $\lambda_{2,\text{opt}}$ (resp. $\lambda_{1,\text{EBIC}}$ and $\lambda_{2,\text{EBIC}}$) which represents a normalized maximum absolute error of 1% (resp. 4%). Once again, these results show that the choice of $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ does not alter the performance of our approach. We can see for both penalization parameters, the optimal and the ones from the EBIC criterion, that the performance reaches a plateau from $n = 1600$ for $K_{\tilde{\lambda}_{1,\text{opt}}} = K_{\tilde{\lambda}_{1,\text{EBIC}}} = 6$ and $K_{\tilde{\lambda}_{2,\text{opt}}} = K_{\tilde{\lambda}_{2,\text{EBIC}}} = 9$, which is on a par with what has been found for the number of selected knots for the one-dimensional case in Figures 2 and 3 for $n \geq 40$.

3. NUMERICAL EXPERIMENTS

In this section, we will study the behavior of our method using the EBIC criterion called GLOBER for Generalized Lasso for knot selection in multivariate B-spline Regression and implemented in the `glober` R package when the variance of the noise σ^2 increases and when the observation set changes.

To assess the efficiency of our method, we will compare it to state-of-the-art approaches: Gaussian Processes (GP) described in Rasmussen and Williams (2006) and implemented in the Python package `scikit-learn`, Multivariate Adaptive Spline Regression (MARS) introduced in Friedman (1991) and implemented in the R package `earth` and Deep Neural Networks (DNN) implemented in the R package `keras`.

For the GP, we chose the squared exponential covariance function as defined in Savino et al. (2022). For the MARS approach, we used the default settings proposed in the `earth` package. It has to be noticed that for the two-dimensional case the interaction terms are included in the model in order not to penalize it. The architecture of the DNN was chosen arbitrarily since our goal is not to optimize it in this paper. More precisely, we used a 2-hidden-layered structure composed of 10 neurons per layer. The activation function of the hidden layers was the RELU function since it is one of the most used functions. In order to train this DNN, we used the stochastic gradient descent method Adam as the optimizer and the Mean Squared Error (MSE) as the loss function. According to the analysis of loss function curves during a pre-processing step, we trained our DNN over 300 epochs for functions of $d = 1$ and 50 epochs for functions of $d = 2$ to avoid overfitting.

3.1. Influence of σ on the statistical performance of the method. We first investigate the influence of the level of noise on the performance of GLOBER. To do so, we applied our method to observations corrupted with two different levels of noise and we computed the

average Normalized Sup Norm in (17), for 10 different samplings of the observations. In both cases ($d = 1$ or 2), the set of knots used to define the underlying function to estimate belongs to the observation set.

3.1.1. *One-dimensional case ($d = 1$)*. We first study the estimation of the function f_1 defined in (15) from a noisy set of observations. The corresponding $(Y_i)_{1 \leq i \leq n}$ for $\sigma = 0.05$ (resp. $\sigma = 0.25$) and $n = 201$ are displayed in the left (resp. right) part of Figure 6.

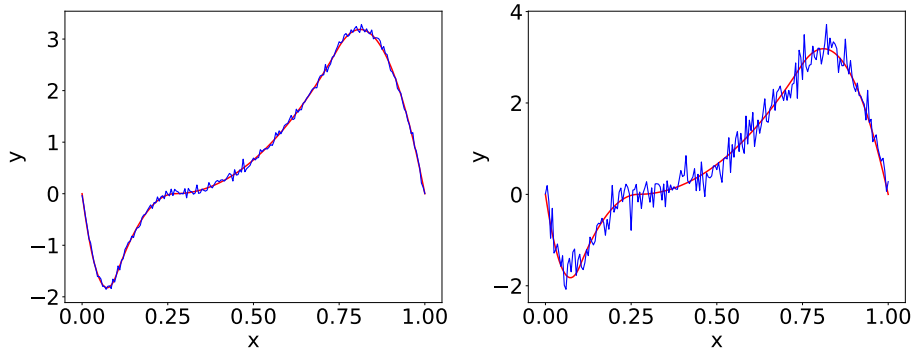


FIGURE 6. Function f_1 to estimate with a noisy set of observations Y_1, \dots, Y_{201} of $\sigma = 0.05$ (left) and $\sigma = 0.25$ (right).

The two parts of the Hausdorff distance between the real knots and the estimated ones as well as the number of selected knots obtained for the noisiest observation set ($\sigma = 0.25$) are displayed in Figure 7. We can see from these results that even the highest value of σ ($\sigma = 0.25$) does not alter the second part of the Hausdorff distance d_2 and that the number of selected knots remains the same as the one previously found for $\sigma = 0.1$.

The corresponding results for the statistical performance defined in (17) for n varying from 7 to 100 are displayed in Figure 8 for $N = 201$. We can see from this figure that the level of noise deteriorates the performance of every method. However, our approach still has high levels of precision since the Normalized Sup Norm varies from $10^{-1.75}$ to $10^{-1.25}$ for $n = 100$ which allows it to outperform the other methods.

3.1.2. *Two-dimensional case ($d = 2$)*. In this part, we focus on the estimation of the function f_2 for $d = 2$ from a noisy set of observations $(Y_i)_{1 \leq i \leq n}$ obtained with $\sigma = 0.005$ (resp. $\sigma = 0.05$) and $n = 40401$. The corresponding Y_i s are displayed in the left (resp. right) part of Figure 9.

The two parts of the Hausdorff distance between the estimated and the real knots as well as the number of selected knots obtained for the noisiest observation set ($\sigma = 0.05$) are shown in Figure 20 of the Appendix. As for the one-dimensional case, we can see that the results are similar to those previously obtained with $\sigma = 0.01$ in Figure 19 of the Appendix: the number of selected knots is the same and the second part of the Hausdorff distance tends to 0.

Figure 21 of the Appendix displays the average of the statistical performance obtained from 10 random samplings of the set of observations for $N = 40401$ where the statistical measure is defined in (17). We can see that even though an alteration of the performance is visible, our approach still outperforms the other methods since the normalized metric keeps decreasing, contrary to the DNNs and the MARS approaches which seem to reach a plateau and the Gaussian Processes which led to very poor accuracy on noisy observations. Once again, our

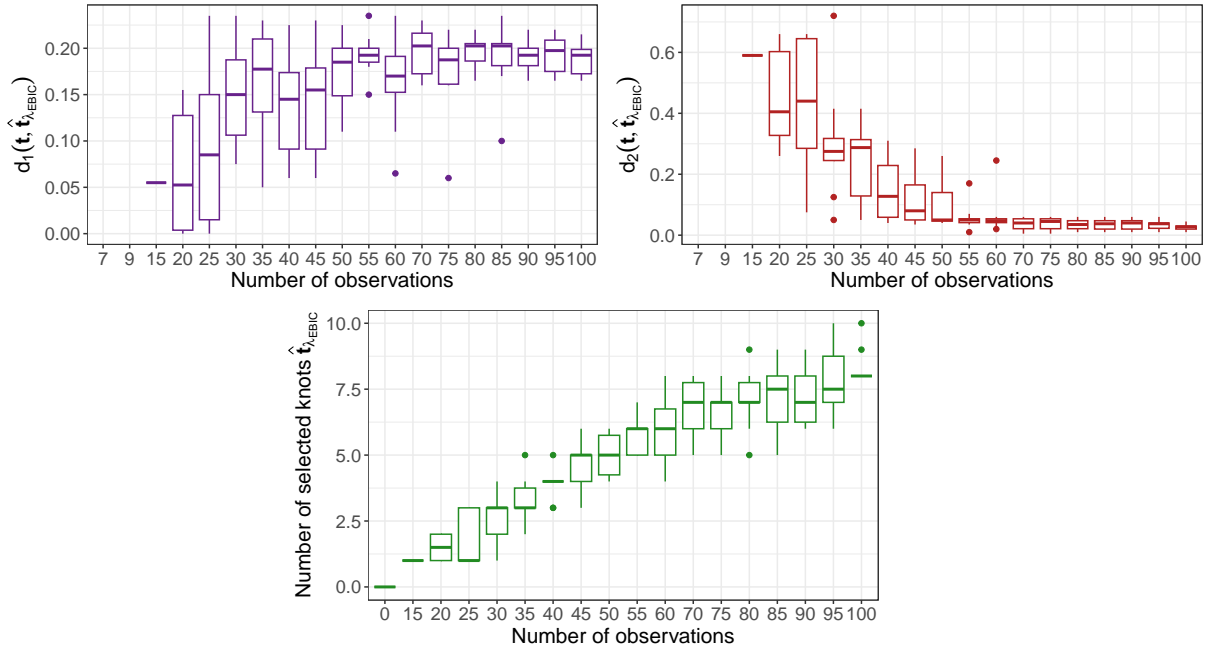


FIGURE 7. Similar to Figure 2 with $\lambda = \lambda_{\text{EBIC}}$ for estimating f_1 when $\sigma = 0.25$.

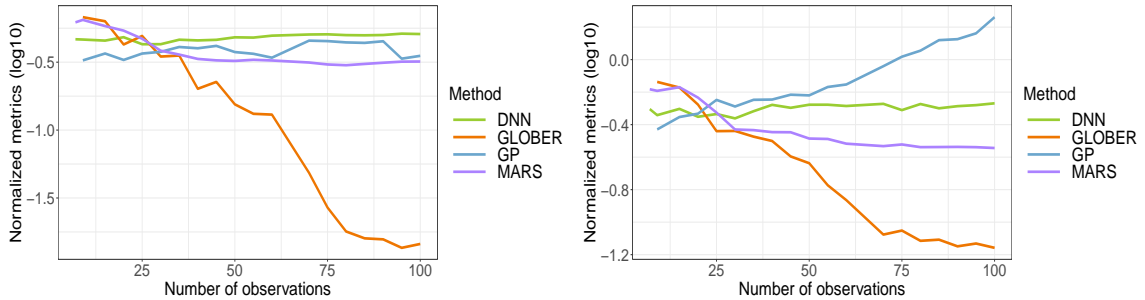


FIGURE 8. Statistical performance (Normalized Sup Norm) of GLOBER for estimating f_1 when $\sigma = 0.05$ (left) and $\sigma = 0.25$ (right) and of the state-of-the-art methods obtained from 10 replications.

method remains robust with highly noisy observation sets both in the one-dimensional and in the two-dimensional case.

3.2. Influence of the sampling of the observation set. We now assess the influence of the sampling of the observation set on the performance of our approach. To do so, we apply our method on randomly chosen observations and we calculate the average Normalized Sup Norm defined in (17) on 10 different samplings of the observations. In such situations, the knots used to define the function to estimate are not necessarily included in the set of observations and in this case, cannot thus be chosen as knots of the B-spline basis. Then, we compare it

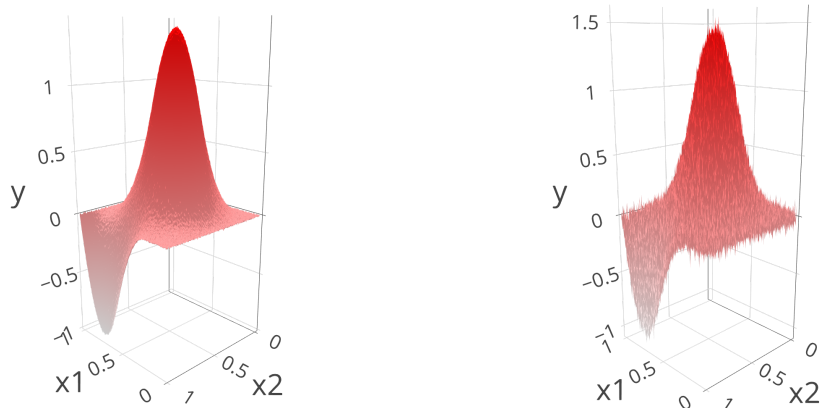


FIGURE 9. Function f_2 to estimate with a noisy set of observations Y_1, \dots, Y_{40401} with $\sigma = 0.005$ (left) and $\sigma = 0.05$ (right).

to the case where the set of observations necessarily contains the set of knots used to define the underlying function f to estimate.

3.2.1. *One-dimensional case ($d = 1$).* We first assess the estimation of the function f_1 defined in (15) from a noisy set of observations obtained with $\sigma = 0.05$. Figure 10 shows the Hausdorff distance between the set of knots \mathbf{t} of the function f_1 and the observation set \mathbf{x} and the Hausdorff distance between the set of knots \mathbf{t} and its estimation with our method in the case where the knots are not necessarily included in the observation set. We can see from this figure that for each element of the observation set, there exists at least one knot at a distance smaller than 0.25. Moreover, for each knot there exists at least one point of the observation set at a distance very close to 0 for large enough n . In addition, we can see from the plot on the bottom right that for large enough n , there exists for each knot an estimated one which is very close.

In the case where \mathbf{t} belongs to \mathbf{x} , the results are displayed in Figure 22 of the Appendix. The boxplot on the top left shows the same results for the distance $d_1(\mathbf{t}, \mathbf{x})$ as for the random sampling and the boxplot on the top right confirms that \mathbf{t} belongs to \mathbf{x} since $d_2(\mathbf{t}, \mathbf{x}) = 0$ at every value of n . Furthermore, we can see similar results on the bottom left and bottom right boxplots since distances $d_1(\mathbf{t}, \hat{\mathbf{t}}_{\lambda_{\text{EBIC}}})$ and $d_2(\mathbf{t}, \hat{\mathbf{t}}_{\lambda_{\text{EBIC}}})$ have exactly the same behavior as for the random sampling case.

Finally, the number of selected knots displayed in Figure 11 shows comparable results between the random sampling of the observations and when \mathbf{t} belongs to \mathbf{x} . Therefore, the random sampling of the observations does not seem to affect the knot selection of our method.

The results of the statistical performance of our method for the estimation of the function f_1 are displayed in Figure 12 for $N = 201$. We can clearly see that the random sampling of the observation set does not deteriorate the performance of our method in comparison to the case where the observation set contains all the knots: the value of the Normalized Sup Norm reaches in both cases $10^{-1.75}$ for $n = 100$ and our method still outperforms the other ones.

3.2.2. *Two-dimensional case ($d = 2$).* Similarly to the previous part, we study the estimation of the function f_2 from a noisy set of observations obtained with $\sigma = 0.01$. Figure 23 of the

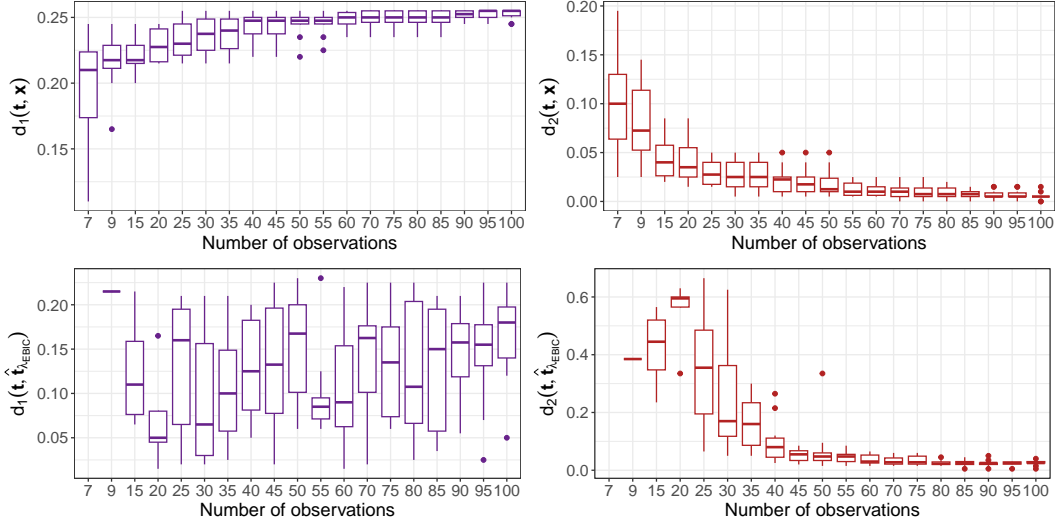


FIGURE 10. Boxplots of the first part of the Hausdorff distance as a function of n : $d_1(\mathbf{t}, \mathbf{x})$ (top left) and $d_1(\mathbf{t}, \hat{\mathbf{t}}_{\lambda_{\text{EBIC}}})$ (bottom left) and for the second part of the Hausdorff distance as a function of n : $d_2(\mathbf{t}, \mathbf{x})$ (top right) and $d_2(\mathbf{t}, \hat{\mathbf{t}}_{\lambda_{\text{EBIC}}})$ (bottom right) for estimating f_1 when the observation set is randomly chosen and $\sigma = 0.05$.

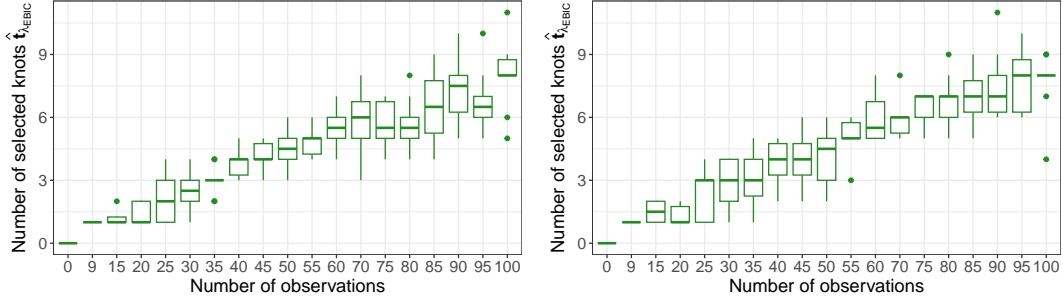


FIGURE 11. Number of estimated knots as a function of n for the estimation of f_1 with GLOBER from a random sampling of observations (left) and when \mathbf{t} belongs to \mathbf{x} (right) with $\sigma = 0.05$.

Appendix shows the Hausdorff distance between the set of knots \mathbf{t}_1 (resp. \mathbf{t}_2) of the function f_2 and the observation set $\mathbf{x}_1 = \{x_{11}, \dots, x_{1n_1}\}$ (resp. $\mathbf{x}_2 = \{x_{21}, \dots, x_{2n_2}\}$) for the first (resp. second) dimension. It also displays the Hausdorff distance between the set of knots \mathbf{t}_1 (resp. \mathbf{t}_2) and its estimation with our method $\hat{\mathbf{t}}_{1, \tilde{\lambda}_1}$ (resp. $\hat{\mathbf{t}}_{2, \tilde{\lambda}_2}$), with $\tilde{\lambda}_1 = \tilde{\lambda}_{1, \text{EBIC}}$ (resp. $\tilde{\lambda}_2 = \tilde{\lambda}_{2, \text{EBIC}}$) for the first (resp. second) dimension in the case where the knots do not necessarily belong to the observation set. We can see from this figure that for each element of the observation set, there exists at least one knot at a distance smaller than 0.45 (resp. 0.38) for the points of the first dimension (resp. second dimension). Furthermore, for each knot of each dimension there exists at least one point of the observation set of the corresponding

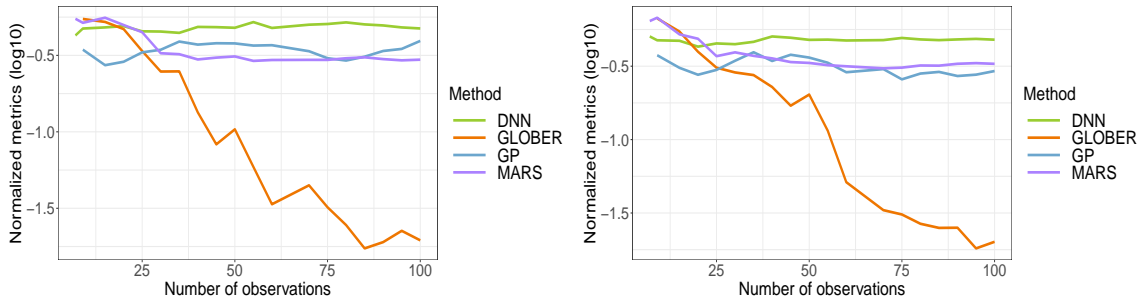


FIGURE 12. Statistical performance (Normalized Sup Norm) of GLOBER for estimating f_1 from a random sampling of the observation set (left) and with \mathbf{t} belonging to the observation set (right) with $\sigma = 0.05$. Comparison to the performance of the state-of-the-art methods obtained from 10 replications.

dimension at a distance very close to 0 when n is large enough (at least 20 points per dimension). For the case where \mathbf{t}_1 and \mathbf{t}_2 belong to \mathbf{x}_1 and \mathbf{x}_2 , respectively, we can see from Figure 24 of the Appendix that the distances $d_1(\mathbf{t}_1, \mathbf{x}_1)$ and $d_1(\mathbf{t}_2, \mathbf{x}_2)$ have the same behavior as for the random sampling. Moreover, the boxplots on the top right shows that $d_2(\mathbf{t}_1, \mathbf{x}_1) = 0$ and $d_2(\mathbf{t}_2, \mathbf{x}_2) = 0$ at every n value which confirms that the sets of knots belong to the observation set. Similar results can be observed for the evolution of the distances $d_1(\mathbf{t}_1, \hat{\mathbf{t}}_{1, \tilde{\lambda}_1})$ (resp. $d_1(\mathbf{t}_2, \hat{\mathbf{t}}_{2, \tilde{\lambda}_2})$) and $d_2(\mathbf{t}_1, \hat{\mathbf{t}}_{1, \tilde{\lambda}_1})$ (resp. $d_2(\mathbf{t}_2, \hat{\mathbf{t}}_{2, \tilde{\lambda}_2})$) for the random sampling and when the knots belong to the observation set. Indeed, in both cases we can see from the plot on the bottom right of Figures 23 and 24 of the Appendix that for large enough n , there exists for each knot of each dimension an estimated one which is very close. Moreover, the number of selected knots displayed in Figure 25 of the Appendix shows comparable results between the random sampling of the observations and when \mathbf{t} belongs to \mathbf{x} . Therefore, the random sampling of the observations does not seem to affect the knot selection of our method for $d = 2$.

As it is the case for $d = 1$, we can see from Figure 26 of the Appendix where the statistical performance of our method for estimating f_2 for $N = 40401$ are displayed that the performance of our approach is not altered by the sampling of the observation set. More precisely, the value of the Normalized Sup Norm reaches in both cases $10^{-1.5}$ for $n = 1600$ and our method still outperforms the other approaches which seem to reach a constant value.

3.3. Numerical performance. The goal of this section is to investigate the computational times of our approach GLOBER implemented in the `glober` R package available on the CRAN as a function of the number of observation points. The timings were obtained on a workstation with 31.2GB of RAM and Intel Core i7 (3.8GHz) CPU. The average computational times and their standard deviation obtained from 30 independent executions are displayed in Figure 13. We can see from this figure that it only takes 600ms to estimate the underlying function with our approach in the one-dimensional. In the two-dimensional case, the computational time is larger but even in this case the computational burden of our approach is low since it only takes 110 seconds for processing 1600 observations.

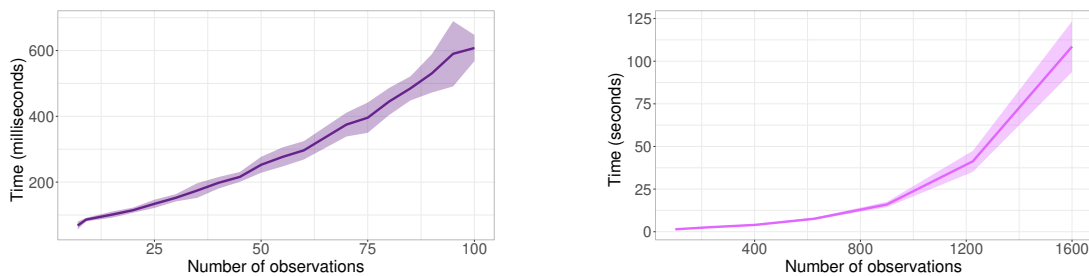


FIGURE 13. Average and standard deviation of the execution time measured in milliseconds for $d = 1$ (left) and in seconds for $d = 2$ (right) on 30 independent executions.

4. APPLICATION TO GEOCHEMICAL SYSTEMS

In this section, we apply our method on simple chemical problems with one or two input concentrations which correspond to the estimation of a function f for $d = 1$ or $d = 2$, respectively. We consider hereafter the estimation of the amount of a "Salt" mineral as a function of the concentrations of its constituents Sp_a^+ and Sp_b^- as in Savino et al. (2022). For this example, the thermodynamic constants of the halite salt (NaCl) were considered because there are only two constitutive elements and they do not depend on the pH of the solution. Following the law of mass action, the dissolution reaction of this mineral writes:



At equilibrium, the activity of these elements $a_{Sp_a^+}$ and $a_{Sp_b^-}$ obey the solubility product

$$K_{\text{Salt}} = a_{Sp_a^+} a_{Sp_b^-} = 10^{1.570}.$$

In the following, the increasing number of observations is obtained by randomly adding new points from a given grid of N points to the pre-existing observation sets. Since the functions to estimate come from chemical processes and are known to be smooth, we propose taking $q \geq 2$. Nevertheless, we observed that using $q = 3$ does not yield better results compared to $q = 2$. This is the reason why we will present in the sequel results obtained for $q = 2$.

4.1. One-dimensional application ($d = 1$). The amount of Salt is first calculated with PHREEQC as in Parkhurst and Appelo (2013) as a function of the normalized concentration of Sp_a^+ so that it belongs to $[0, 1]$. The corresponding function f_3 to estimate is defined as:

$$\text{Salt} = f_3(\text{Normalized } Sp_a^+)$$

and is displayed on the left part of Figure 14. Our method is used to estimate this function by using a set of a varying number of observation points ($7 \leq n \leq 100$) and is then compared to the state-of-the-art methods described in Section 3. Figure 15 displays the illustration of the estimation of f_3 for a varying number of n . We can see by adding new points to the observation set (blue crosses) that GLOBER has better chance to choose knots (blue bullets) among observation points which depict more precisely changes in the underlying curve. However, it seems that 40 observations are enough to have a good approximation of our function f_3 since the estimation (black curve) fits quite perfectly the function to estimate (red curve).

The corresponding statistical performance is displayed on the right side of Figure 14 for $N = 1140$. We can see from this figure that our method outperforms the other approaches

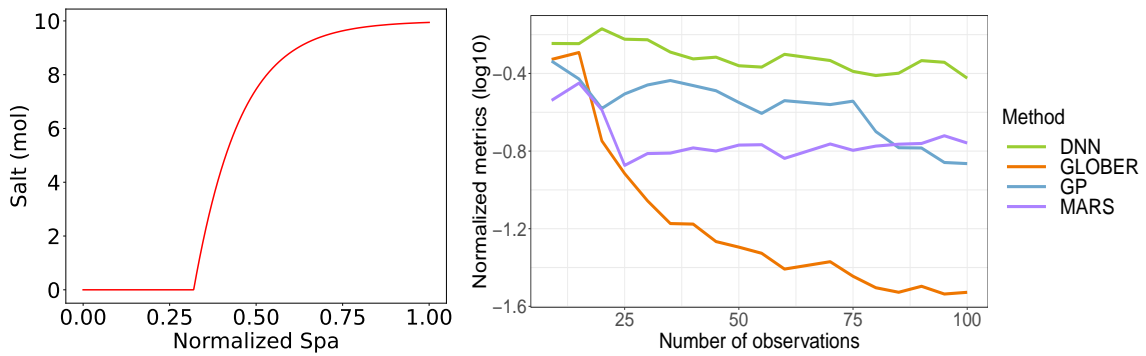


FIGURE 14. Left: Function f_3 to estimate when $d = 1$ with Y_1, \dots, Y_{1140} . Right: Statistical performance (Normalized Sup Norm) of GLOBER and of the state-of-the-art methods for estimating f_3 . The average values are obtained from 10 replications.

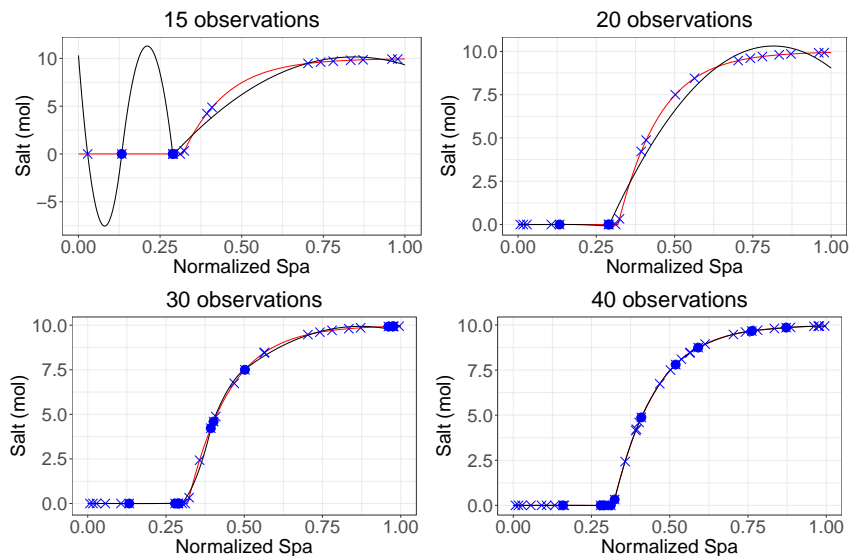
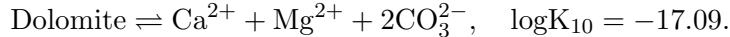


FIGURE 15. Illustration of the estimation of the amount of Salt depending on the normalized concentration of Sp_a^+ for 15 (top left) 20 (top right) 30 (bottom left) and 40 observations (bottom right). The red curve describes the true underlying function f_3 to estimate, the black curve corresponds to the estimation with GLOBER, the blue crosses are the observation points and the blue bullets are the observation points chosen as estimated knots.

and allows us to get a very good accuracy of the amount of Salt since the average Normalized Sup Norm reaches $10^{-1.5}$ for only $n = 100$.

4.2. Two-dimensional application ($d = 2$). Our method is then applied to a more complex chemical problem which derives from the calcite dissolution and precipitation study described in Kolditz et al. (2012). The corresponding thermodynamic data for aqueous species and

minerals are available in the Phreeqc.dat distributed with PHREEQC. The compositional system actually solved consists of 14 species in solution, 2 mineral components, 8 geochemical reactions and 2 mineral dissolution-precipitation reactions. However, in this paper, we will only consider the dolomite precipitation:



The amount of dolomite is computed with PHREEQC as a function of the total elemental concentrations (C, Ca, Cl, Mg), the pH (as $-\log(\text{H}^+)$) and the amount of calcite. In this article, we will only consider the dolomite precipitation for $C=5 \times 10^{-4}$ mol/kgw, $\text{Cl}=2 \times 10^{-3}$ mol/kgw, $\text{pH}=10$, $\text{calcite}=0$ mol in order to reduce the problem to a two-dimensional case. Thus, the function f_4 to be estimated is defined as:

$$\text{Dolomite} = f_4(\text{Normalized Ca}, \text{Normalized Mg}).$$

We seek to estimate f_4 by applying our method to an increasing number of observations ($100 \leq n \leq 1600$) which corresponds to an increasing number of points per dimension ($10 \leq n_1, n_2 \leq 40$ with $n = n_1 n_2$). The resulting illustration is displayed in Figure 27 of the Appendix and shows an improvement of the fitting of GLOBER (green curve) to the underlying curve (red curve) by adding new points to the observation set (orange bullets) in order to get a perfect overlapping for $n \geq 900$. In the right part of Figure 16, the statistical performance of GLOBER compared to the state-of-the-art approaches is displayed for $N = 40000$. Similarly to what has been shown for the precipitation of Salt, our method gives satisfactory results as the Normalized Sup Norm reaches 10^{-1} . Moreover, our method still outperforms the other ones for which the statistical metric seems to rapidly reach a constant value.

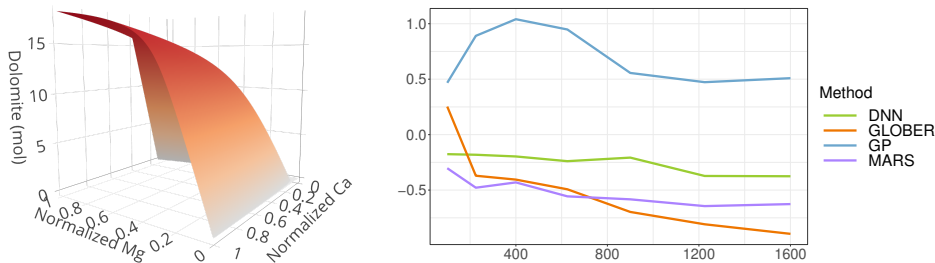


FIGURE 16. Left: Function f_4 to estimate when $d = 2$ with Y_1, \dots, Y_{40000} . Right: Statistical performance (Normalized Sup Norm) of GLOBER and of the state-of-the-art methods for estimating f_4 . The average values are obtained from 10 replications.

5. EXTENSION TO HIGHER DIMENSIONAL AND MORE GENERAL OBSERVATION SETTINGS

Here, let us consider the case where $x_i \in \mathbb{R}^d$ with $d \geq 2$ and f is still a function to estimate from (1). In this section, the grid of N points and the observation sets are no longer constrained to result from a cartesian product of d compact sets, such as in Section 2.2.

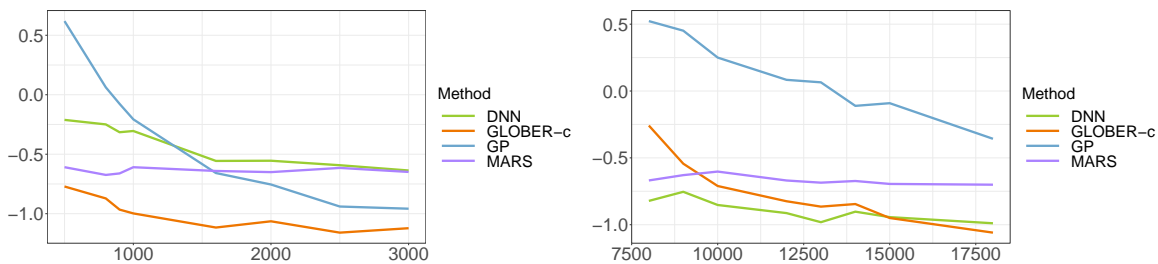


FIGURE 17. Statistical performance (Normalized Sup Norm) of GLOBER-c and of the state-of-the-art methods for estimating f_4 (left) and f_5 (right). The average values are obtained from 10 replications.

5.1. Adaptation of the knot selection method by using clustering. In order to avoid finding the knots associated to a given dimension for each fixed value of the others, we propose in this section to cluster the n observation points by using a k -means approach. More precisely, for each dimension j , the n observation points are gathered into clusters based on their multidimensional characteristics excluding dimension j . Then, the one-dimensional approach described in Section 2.1 is applied to the Y 's associated to the x 's belonging to this cluster to find the knots in dimension j . The number k of clusters is chosen as the integer part of n divided by 25 to guarantee a large enough amount of points per cluster. Thus, since sets of knots are found for each $\tilde{\lambda}_j$ of $\tilde{\Lambda}_j$ we gather the knots $\hat{\mathbf{t}}_{\tilde{\lambda}_j}$ into clusters. The final knot sets are built by keeping the median value of each cluster of knots. The number of knot clusters is chosen as the value where there is a change in the slope of the within-sum of squares. Finally, the EBIC criterion defined in (21) is used to find the final combination of knot sets.

5.2. Case where $d = 2$. This method was first validated on the two-dimensional framework application by comparing the results obtained for the clusterized version called GLOBER-c on the Dolomite precipitation case described in Section 4.2. The obtained results are displayed in the left part of Figure 17. We can see from this figure that compared to the results obtained in Section 4.2 with the non-clusterized version, the maximal absolute error is smaller than 10^{-1} for $n = 1600$ and that our method still outperforms the three others.

5.3. Case where $d = 3$. In this section, we apply our method to an extension of the geochemical system described in Section 4.2. Our goal is to estimate f_5 which describes the relationship between the precipitation of Dolomite and three chemical elements:

$$\text{Dolomite} = f_5(\text{Normalized C}, \text{Normalized Ca}, \text{Normalized Mg}).$$

The obtained results are displayed in the right part of Figure 17. We can see from this figure that our method is the only one that displays a maximal absolute error below 10^{-1} . Moreover, it outperforms the three others when the number of observations is larger than 15000 observations.

REFERENCES

Asher, M. J., Croke, B. F. W., Jakeman, A. J., and Peeters, L. J. M. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, 51(8):5957–5973.

- Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771.
- De Boor, C. (1978). *A practical guide to splines*, volume 27. Springer-Verlag New York.
- de Capitani, C. and Brown, T. H. (1987). The computation of chemical equilibrium in complex systems containing non-ideal solutions. *Geochimica et Cosmochimica Acta*, 51(10):2639–2652.
- Denis, C., Lebarbier, E., Lévy-Leduc, C., Martin, O., and Sansonnet, L. (2020). A novel regularized approach for functional data clustering: An application to milking kinetics in dairy goats. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 69(3):623–640.
- Eckle, K. and Schmidt-Hieber, J. (2019). A comparison of deep networks with ReLU activation function and linear spline-type methods. *Neural Networks*, 110:232–242.
- Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science*, 11(2):89–121.
- Eilers, P. H. and Marx, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and intelligent laboratory systems*, 66(2):159–174.
- Eilers, P. H., Marx, B. D., and Durbán, M. (2015). Twenty years of P-splines. *SORT: statistics and operations research transactions*, 39(2):0149–186.
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1):1–67.
- Goepp, V., Bouaziz, O., and Nuel, G. (2018). Spline regression with automatic knot selection. arXiv preprint arXiv:1808.01770.
- Guérillot, D. and Bruyelle, J. (2020). Geochemical equilibrium determination using an artificial neural network in compositional reservoir flow simulation. *Computational Geosciences*, 24(2):697–707.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data mining, inference, and prediction*. Springer New York Inc., New York, NY, USA.
- Jatnieks, J., De Lucia, M., Dransch, D., and Sips, M. (2016). Data-driven surrogate model approach for improving the performance of reactive transport simulations. *Energy Procedia*, 97:447–453.
- Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. (2018). Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554.
- Kolditz, O., Görke, U.-J., Shao, H., and Wang, W. (2012). *Thermo-hydro-mechanical-chemical processes in porous media: benchmarks and examples*, volume 86. Springer Science & Business Media.
- Laloy, E. and Jacques, D. (2019). Emulation of CPU-demanding reactive transport models: a comparison of Gaussian processes, polynomial chaos expansion, and deep neural networks. *Computational Geosciences*, 23:1193–1215.
- Li, Z. and Cao, J. (2022). General P-splines for non-uniform B-splines. arXiv preprint arXiv:2201.06808.
- O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1(4):502–518.
- Parkhurst, D. L. and Appelo, C. (2013). *Description of input and examples for PHREEQC version 3: a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations*. U.S.G.S. Techniques and Methods.

- Rasmussen, C. E. and Williams, C. K. I. (2006). Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press.
- Sadhanala, V., Wang, Y.-X., Hu, A. J., and Tibshirani, R. J. (2021). Multivariate trend filtering for lattice data. arXiv preprint arXiv:2112.14758.
- Savino, M., Lévy-Leduc, C., Leconte, M., and Cochebin, B. (2022). An active learning approach for improving the performance of equilibrium based chemical simulations. Computational Geosciences, 26(2):365–380.
- Smith, W. R. (1980). The computation of chemical equilibria in complex systems. Industrial & Engineering Chemistry Fundamentals, 19(1):1–10.
- Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. The Annals of Statistics, 42(1):285–323.
- Tibshirani, R. J. and Taylor, J. (2011). The solution path of the generalized lasso. The Annals of Statistics, 39(3):1335 – 1371.
- Wahba, G. (1990). Spline models for observational data. Society for Industrial and Applied Mathematics.
- Wand, M. P. and Ormerod, J. T. (2008). On semiparametric regression with O’Sullivan penalized splines. Australian & New Zealand Journal of Statistics, 50(2):179–198.
- White, W. B., Johnson, S. M., and Dantzig, G. B. (1958). Chemical equilibrium in complex mixtures. The Journal of Chemical Physics, 28(5):751–755.
- Yuan, Y., Chen, N., and Zhou, S. (2013). Adaptive B-spline knot selection using multi-resolution basis set. IIE Transactions, 45(12):1263–1277.
- Zhang, W. and Goh, A. T. (2013). Multivariate adaptive regression splines for analysis of geotechnical engineering systems. Computers and Geotechnics, 48:82–95.
- Zhang, W. and Goh, A. T. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geoscience Frontiers, 7(1):45–52.

6. APPENDIX

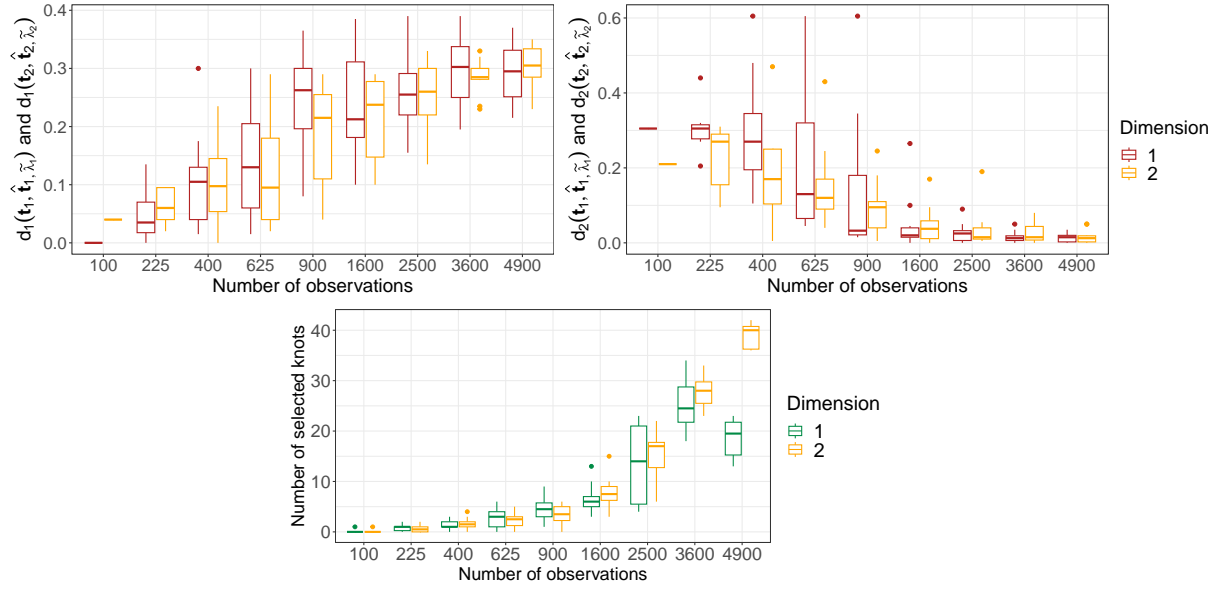


FIGURE 18. Top left: Boxplots for the first part of the Hausdorff distance (d_1) as a function of $n = n_1 n_2$ and top right: boxplots for the second part of the Hausdorff distance (d_2) as a function of $n = n_1 n_2$ between the two sets of knots \mathbf{t}_1 and $\hat{\mathbf{t}}_1, \tilde{\lambda}_1$ and \mathbf{t}_2 and $\hat{\mathbf{t}}_2, \tilde{\lambda}_2$ for the first and second dimension, respectively. Bottom: number of estimated knots as a function of $n = n_1 n_2$ by choosing $\tilde{\lambda}_1 = \tilde{\lambda}_{1, \text{EBIC}}$ and $\tilde{\lambda}_2 = \tilde{\lambda}_{2, \text{EBIC}}$ for the estimation of f_2 .

FRENCH NATIONAL RADIOACTIVE WASTE MANAGEMENT AGENCY (ANDRA) AND UNIVERSITÉ PARIS-SACLAY, AGROPARISTECH, INRAE, UMR MIA PARIS-SACLAY

UNIVERSITÉ PARIS-SACLAY, AGROPARISTECH, INRAE, UMR MIA PARIS-SACLAY
Email address, Corresponding author: mary.savino@outlook.fr

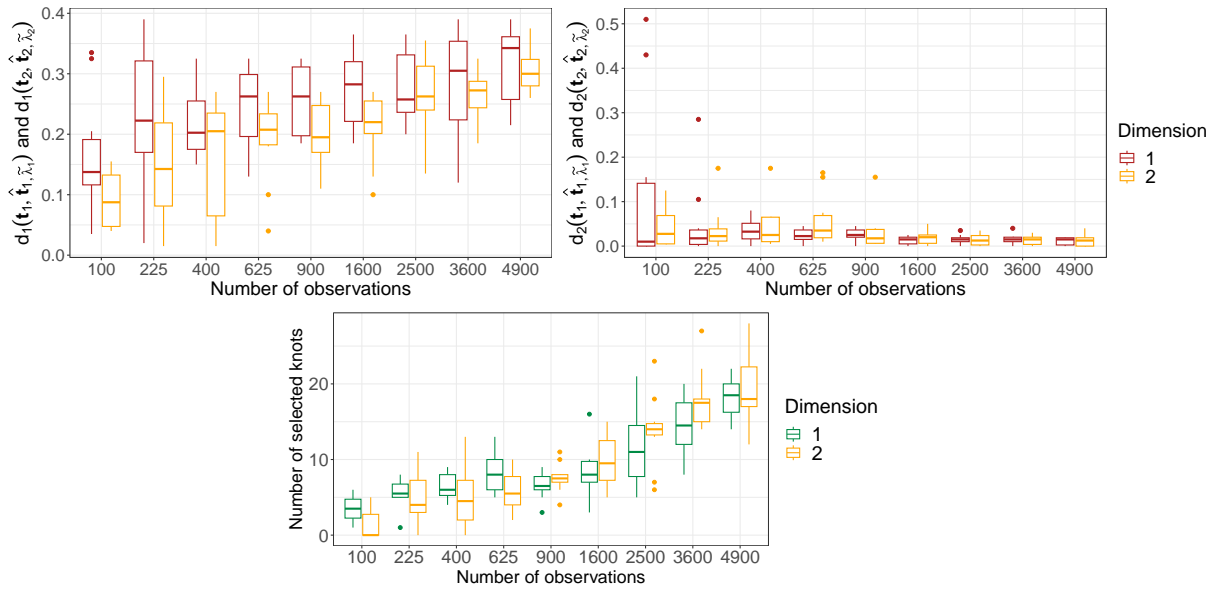


FIGURE 19. Similar to Figure 18 by choosing $\tilde{\lambda}_1 = \tilde{\lambda}_{1,\text{opt}}$ and $\tilde{\lambda}_2 = \tilde{\lambda}_{2,\text{opt}}$ for the estimation of f_2 with $\sigma = 0.01$.

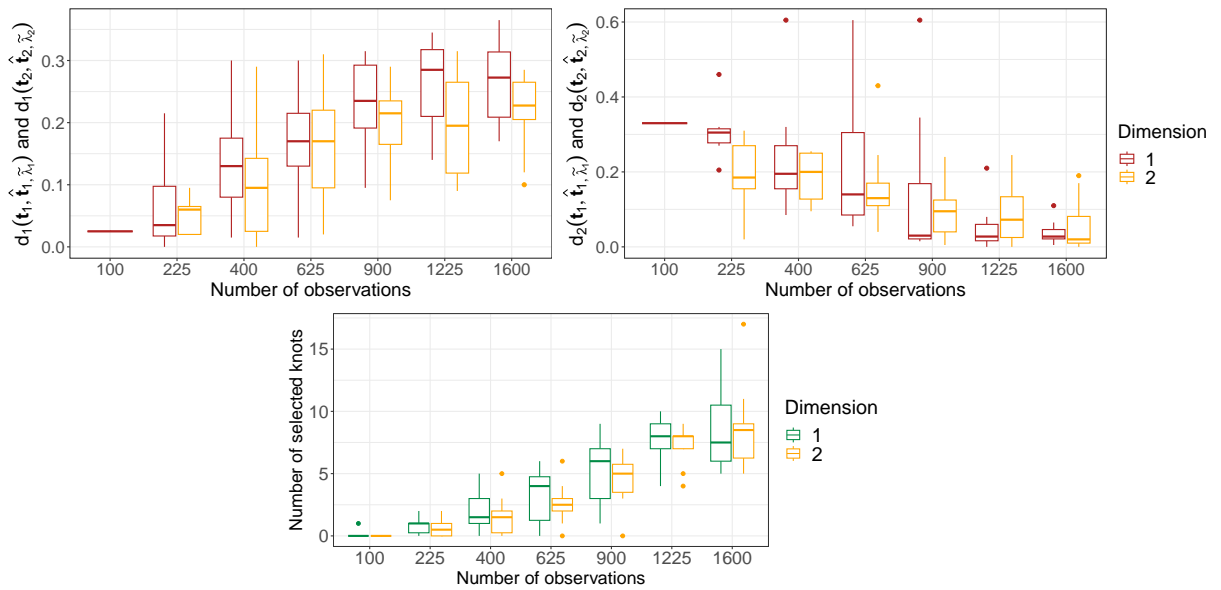


FIGURE 20. Similar to Figure 18 by choosing $\tilde{\lambda}_1 = \tilde{\lambda}_{1,\text{EBIC}}$ and $\tilde{\lambda}_2 = \tilde{\lambda}_{2,\text{EBIC}}$ for the estimation of f_2 with $\sigma = 0.05$.

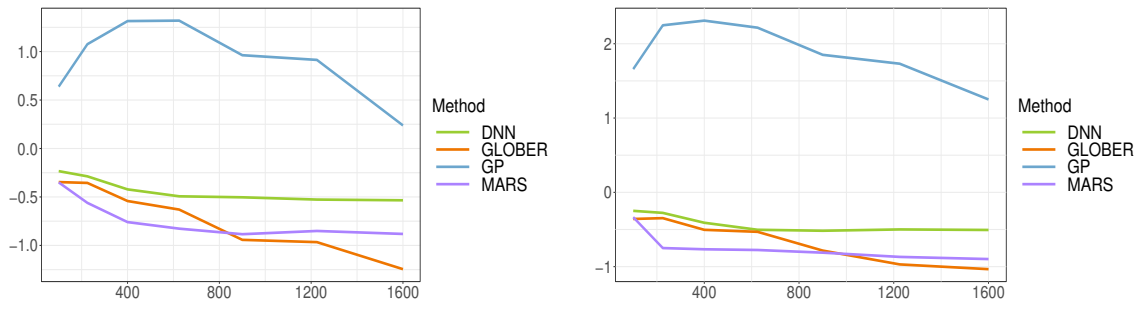


FIGURE 21. Statistical performance (Normalized Sup Norm) of our method using the EBIC criterion for $\sigma = 0.005$ (left) and $\sigma = 0.05$ (right) and of the state-of-the-art methods obtained from 10 replications.

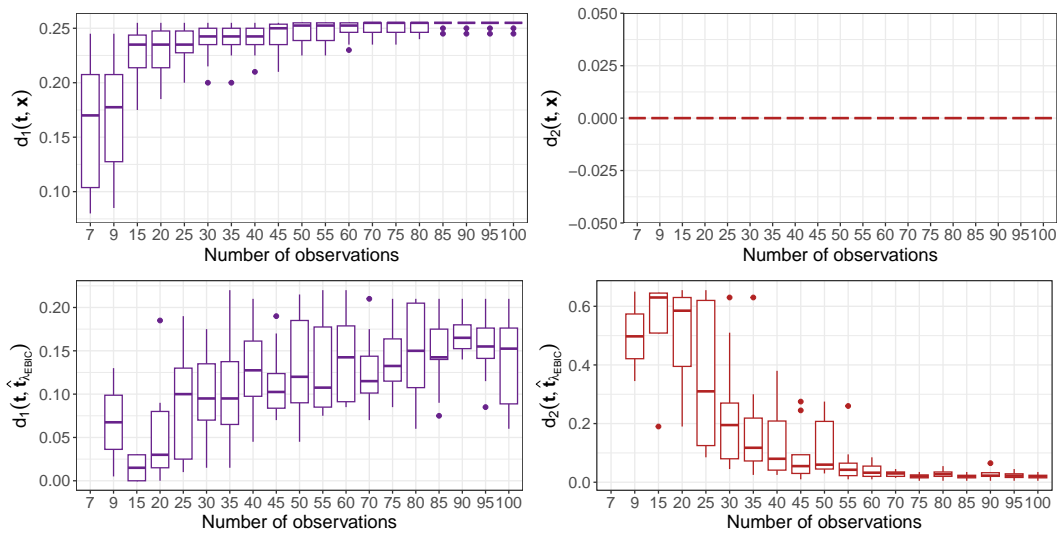


FIGURE 22. Similar to Figure 10 for the estimation of f_1 with \mathbf{t} belonging to \mathbf{x} and $\sigma = 0.05$.

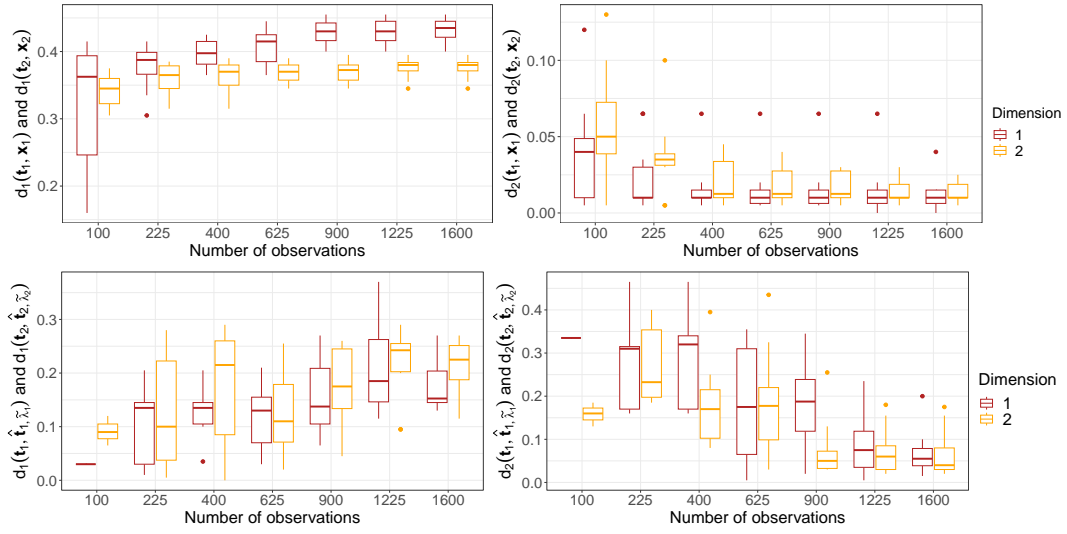


FIGURE 23. Boxplots of the first part of the Hausdorff distance as a function of n : $d_1(\mathbf{t}_1, \mathbf{x}_1)$ (resp. $d_1(\mathbf{t}_2, \mathbf{x}_2)$) (top left) and $d_1(\mathbf{t}_1, \hat{\mathbf{t}}_1, \tilde{\lambda}_1)$ (resp. $d_1(\mathbf{t}_2, \hat{\mathbf{t}}_2, \tilde{\lambda}_2)$) (bottom left) and for the second part of the Hausdorff distance as a function of n : $d_2(\mathbf{t}_1, \mathbf{x}_1)$ (resp. $d_2(\mathbf{t}_2, \mathbf{x}_2)$) (top right) and $d_2(\mathbf{t}_1, \hat{\mathbf{t}}_1, \tilde{\lambda}_1)$ (resp. $d_2(\mathbf{t}_2, \hat{\mathbf{t}}_2, \tilde{\lambda}_2)$) (bottom right) for the first (resp. second) dimension, for the estimation of f_2 by choosing $\tilde{\lambda}_1 = \tilde{\lambda}_{1,\text{EBIC}}$ and $\tilde{\lambda}_2 = \tilde{\lambda}_{2,\text{EBIC}}$ with a random sampling of the observation set with $\sigma = 0.01$.

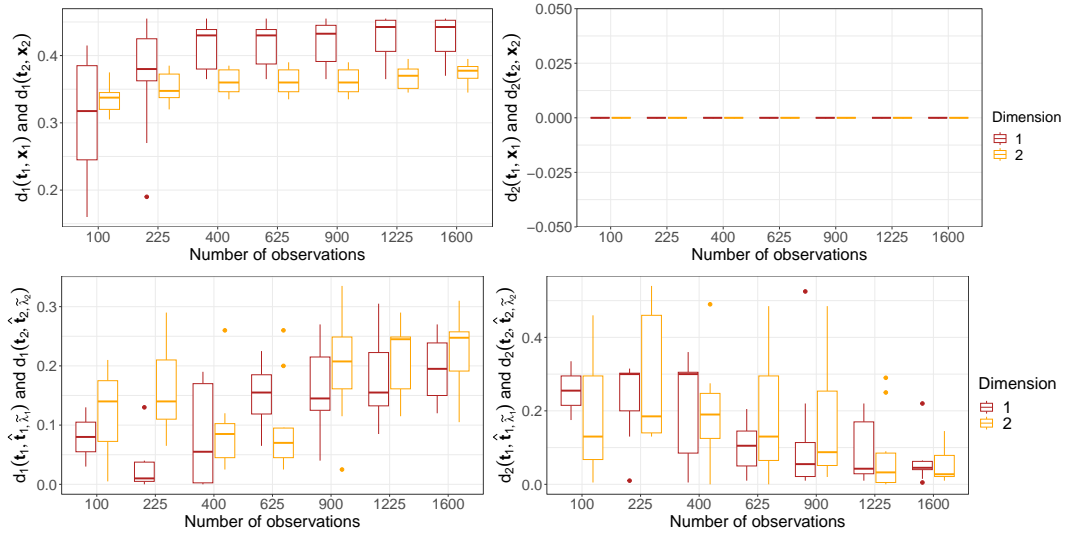


FIGURE 24. Similar to Figure 23 for the estimation of f_2 with \mathbf{t}_1 belonging to \mathbf{x}_1 (resp. \mathbf{t}_2 belonging to \mathbf{x}_2) and $\sigma = 0.01$.

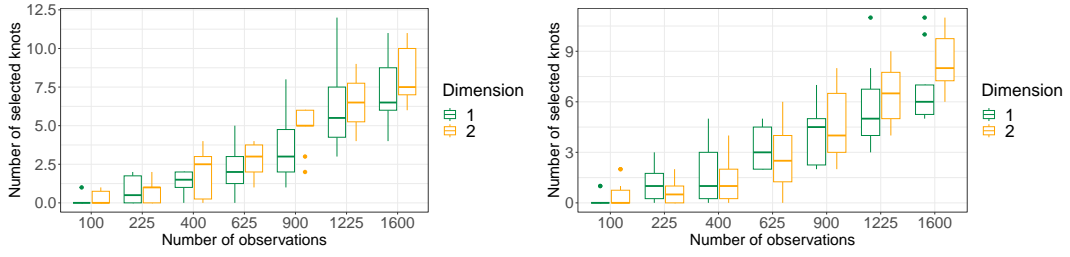


FIGURE 25. Left: number of estimated knots as a function of n for the estimation of f_2 with GLOBER from a random sampling of observations (left) and when \mathbf{t}_1 and \mathbf{t}_2 belong to \mathbf{x}_1 and \mathbf{x}_2 , respectively (right) with $\sigma = 0.01$.

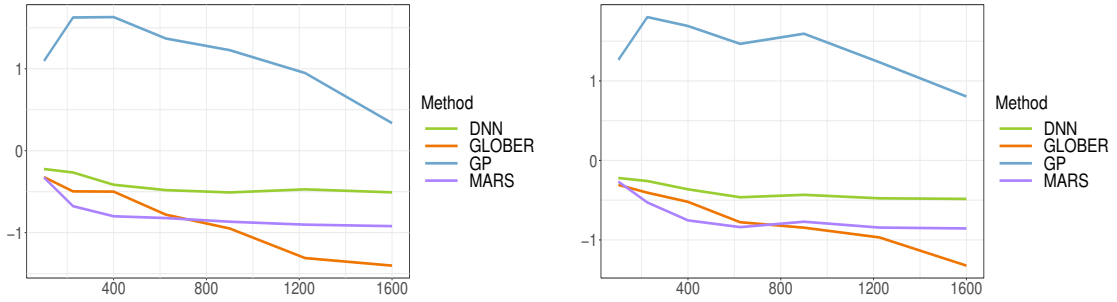


FIGURE 26. Statistical performance (Normalized Sup Norm) of GLOBER from a random sampling of the noisy observation set (left) and with \mathbf{t}_1 and \mathbf{t}_2 belonging to the observation set (right) with $\sigma = 0.01$. Comparison to the performance of state-of-the-art methods obtained from 10 replications.

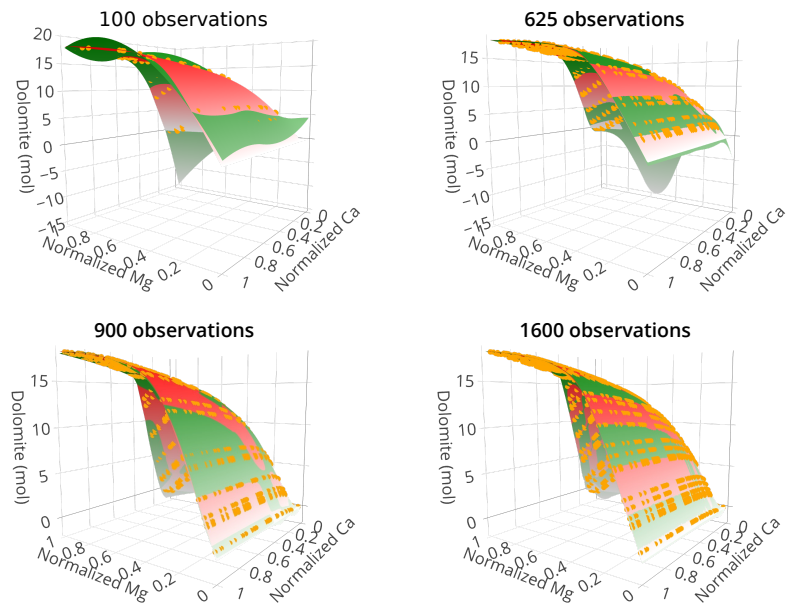


FIGURE 27. Illustration of the estimation of the amount of dolomite depending on the normalized concentration of Ca and Mg for 100 (top left) 625 (top right) 900 (bottom left) and 1600 observations (bottom right). The red surface describes the true underlying function f_4 to estimate, the green surface corresponds to the estimation with GLOBER and the orange bullets are the observation points.