A novel approach for estimating functions in the multivariate setting based on an adaptive knot selection for B-splines

StatMathAppli 2023

presented by Mary Savino, PhD student [1,2]

supervised by Céline Levy-Leduc [1]
and by Benoit Cochepin [2] and Marc Leconte [2]

[1]MIA-Paris-Saclay, AgroParisTech

[2]Andra

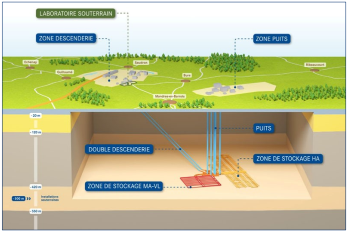
General context

Andra : French National Agency for Radioactive Waste Management

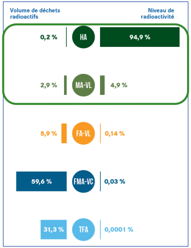
"Taking charge of radioactive waste produced by past and current generations to render it secure for future generations "

Location of Andra sites*

Cigéo Project*



Radioactive waste classification*



*Sources Andra

General context

To model the evolution of Cigéo components and the different interactions with their environment

Solving chemical reactions with **PHREEQC**, a computer program which takes x as inputs and gives output y such that $y = f(x)$ where f is an unknown function

Problem : time-consuming and costly simulations

IDEA : to estimate f (« black box »)

First solution : deep learning with **a deep neural network** exploiting a set of observations generated with PHREEQC

Problem : complex calibration of parameters, not yet optimized

Second solution : Implementation of a **statistical method** for estimating functions in a **multivariate nonparametric regression model** based on an **adaptive knot selection for B-splines.**

Definition of B-splines of order M

Let $\mathbf{t} = (t_1, \ldots, t_K)$ be a set of K points called **knots**. We define the augmented knot sequence τ such that:

$$\tau_1 = \ldots = \tau_M = x_{min},$$

$$\tau_{j+M} = t_j, \quad j = 1, \ldots, K,$$

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

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

Definition of B-splines of order M

Let $\mathbf{t} = (t_1, \ldots, t_K)$ be a set of K points called **knots**. We define the augmented knot sequence τ such that:

$$\tau_1 = \ldots = \tau_M = x_{min},$$
$$\tau_{j+M} = t_j, \quad j = 1, \ldots, K,$$
$$x_{max} = \tau_{K+M+1} = \ldots = \tau_{K+2M},$$
$$\tau = (\tau_1, \ldots, \tau_{K+2M}) = \big(\underbrace{x_{min}, \ldots, x_{min}}_{\text{M times}}, \underbrace{t_1, \ldots, t_K}_{\mathbf{t}}, \underbrace{x_{max}, \ldots, x_{max}}_{\text{M times}}\big),$$

B-splines are defined by De Boor (1978) by the following recursion: Denoting by $B_{i,m}(x)$ the ith B-spline basis function of order m for the knot sequence τ with $m \leq M$:

Definition of B-splines by recursion

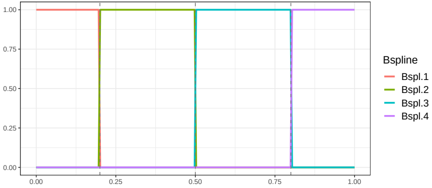
$$B_{i,1}(x) = \begin{cases} 1 & \text{if } \tau_i \leq x < \tau_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad \text{for} \quad i = 1, \ldots, K + 2M - 1,$$

and for $m \leq M$,

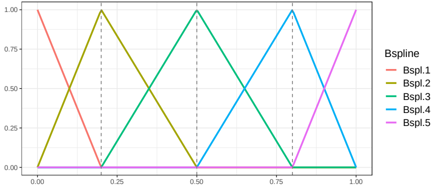
$$B_{i,m}(x) = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_{i+m} - x}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}(x),$$

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

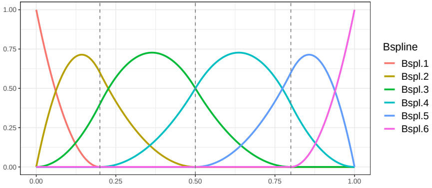
Visualization of B-splines of order M



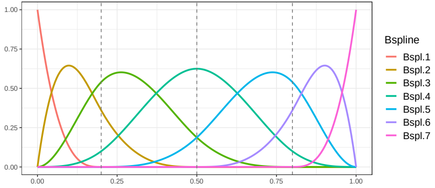
(a) B-spline basis of order $M = 1$



(b) B-spline basis of order $M = 2$



(c) B-spline basis of order $M = 3$



(d) B-spline basis of order $M = 4$

with $\mathbf{t} = (0.2, 0.5, 0.8)$

Nonparametric method to estimate function of one or two variables (1)

$$Y_i = f(x_i) + \varepsilon_i, \quad 1 \leq i \leq n, \quad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

where the x_i are observation points which belong to a compact set of $\mathbb{R}^d$, $d \geq 1$.
Approach : GLOBER inspired by MARS method introduced by Friedman (1991),
***d** $=$ **1** :

1. From the observation points, selection of specific points called **knots** by using the $(q+1)$th order generalized lasso defined by Tibshirani and Taylor (2011),
2. Definition of a B-spline basis of a certain order M,
3. Estimation of a one-dimensional function $(d = 1)$

$$\sum_{i=1}^{K+M} \gamma_i B_{i,M}(x), \tag{1}$$

where K is the number of knots defining the B-spline basis.

Nonparametric method to estimate function of one or two variables (2)

*$\mathbf{d = 2}$:

1. From the observation points, selection of **knots** for each dimension by fixing one dimension at a time so can be rewritten as an estimation problem in the one-dimensional framework ($d = 1$),
2. Definition of a B-spline basis for each dimension,
3. Estimation of a two-variable function ($d = 2$)

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

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

Selection of the knot set ($d = 1$)

Generalized lasso (Tibshirani et al, 2011)

$$\widehat{\beta}(\lambda) = \underset{\beta \in \mathbb{R}^n}{\text{argmin}} \{ ||\mathbf{Y} - \beta||_2^2 + \lambda ||D\beta||_1 \} \quad (3)$$

where $||y||_2^2 = \sum_{i=1}^n y_i^2$ for $y = (y_1, \ldots, y_n)$ and $||u||_1 = \sum_{i=1}^m |u_i|$ for $u = (u_1, \ldots, u_m)$, $\lambda > 0$ and $D \in \mathbb{R}^{m \times n}$ is a specified penalty matrix depending on the **order of differentiation** $(q + 1)$.

Selection of the knot set ($d = 1$)

Generalized lasso (Tibshirani et al, 2011)

$$\widehat{\beta}(\lambda) = \underset{\beta \in \mathbb{R}^n}{\text{argmin}}\{||\mathbf{Y} - \beta||_2^2 + \lambda ||D\beta||_1\} \quad (3)$$

where $||y||_2^2 = \sum_{i=1}^n y_i^2$ for $y = (y_1, \ldots, y_n)$ and $||u||_1 = \sum_{i=1}^m |u_i|$ for $u = (u_1, \ldots, u_m)$, $\lambda > 0$ and $D \in \mathbb{R}^{m \times n}$ is a specified penalty matrix depending on the **order of differentiation** $(q+1)$.

Let $\Lambda = (\lambda_1, \ldots, \lambda_k)$ be a grid of penalization parameters λ_i. We define $\mathbf{a}(\lambda)$ by:

$$\boldsymbol{a}(\lambda) = D \cdot \widehat{\beta}(\lambda), \quad \lambda \in \Lambda$$

Selection of the knot set $(d = 1)$

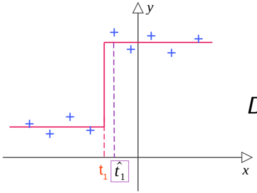
Generalized lasso (Tibshirani et al, 2011)

$$\widehat{\beta}(\lambda) = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \{ ||\mathbf{Y} - \beta||_2^2 + \lambda ||D\beta||_1 \} \quad (3)$$

where $||y||_2^2 = \sum_{i=1}^n y_i^2$ for $y = (y_1, \ldots, y_n)$ and $||u||_1 = \sum_{i=1}^m |u_i|$ for $u = (u_1, \ldots, u_m)$, $\lambda > 0$ and $D \in \mathbb{R}^{m \times n}$ is a specified penalty matrix depending on the **order of differentiation** $(q+1)$.

Let $\Lambda = (\lambda_1, \ldots, \lambda_k)$ be a grid of penalization parameters λ_i. We define $\mathbf{a}(\lambda)$ by:

$$\mathbf{a}(\lambda) = D \cdot \widehat{\beta}(\lambda), \quad \lambda \in \Lambda$$



$$D = \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -1 & 1 \end{bmatrix}, \ \widehat{\beta}(\lambda_3) = \begin{bmatrix} 1.1 \\ 1.1 \\ 1.1 \\ 1.1 \\ 3.6 \\ 3.6 \\ 3.6 \\ 3.6 \\ 3.6 \end{bmatrix}, \ \mathbf{a}(\lambda_3) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Selection of the knot set $(d = 1)$

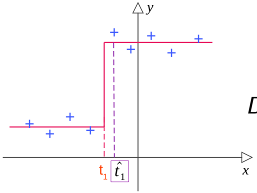
Generalized lasso (Tibshirani et al, 2011)

$$\widehat{\beta}(\lambda) = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \{ ||\mathbf{Y} - \beta||_2^2 + \lambda ||D\beta||_1 \} \qquad (3)$$

where $||y||_2^2 = \sum_{i=1}^n y_i^2$ for $y = (y_1, \ldots, y_n)$ and $||u||_1 = \sum_{i=1}^m |u_i|$ for $u = (u_1, \ldots, u_m)$, $\lambda > 0$ and $D \in \mathbb{R}^{m \times n}$ is a specified penalty matrix depending on the **order of differentiation** $(q + 1)$.

Let $\Lambda = (\lambda_1, \ldots, \lambda_k)$ be a grid of penalization parameters λ_i. We define $\mathbf{a}(\lambda)$ by:

$$\mathbf{a}(\lambda) = D \cdot \widehat{\beta}(\lambda), \quad \lambda \in \Lambda$$



$$D = \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -1 & 1 \end{bmatrix}, \ \widehat{\beta}(\lambda_3) = \begin{bmatrix} 1.1 \\ 1.1 \\ 1.1 \\ 1.1 \\ 3.6 \\ 3.6 \\ 3.6 \\ 3.6 \\ 3.6 \end{bmatrix}, \ \mathbf{a}(\lambda_3) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Which penalization parameter λ to choose to get an optimal estimator of f?

Selection criterion of the parameter λ ($d = 1$)

EBIC criterion defined by Chen and Chen (2008)

$$\mathrm{EBIC}(\lambda) = \mathrm{SS}(\lambda) + (q + K_\lambda + 1)\log n + 2\log\binom{q + K_{\max} + 1}{q + K_\lambda + 1}, \quad (4)$$

where $K_{\max} = n$ and $\mathrm{SS}(\lambda)$ is the sum of squares defined by:

$$\mathrm{SS}(\lambda) = \|\mathbf{Y} - \widehat{\mathbf{Y}}(\lambda)\|_2^2\,, \quad (5)$$

where

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

with $\widehat{\gamma}$ and $\mathbf{B}(\lambda)$ a $n \times (q + K_\lambda + 1)$ matrix having as ith column $(B_{i,M}(x_k))_{1 \le k \le n}$, i belonging to $\{1, \dots, q + K_\lambda + 1\}$.

Final estimator of f

$$\widehat{f}(x) = \widehat{f}_{\lambda_{\mathrm{EBIC}}}(x), \quad (6)$$

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

$$\lambda_{\mathrm{EBIC}} = \underset{\lambda \in \Lambda}{\mathrm{argmin}}\{\mathrm{EBIC}(\lambda)\}. \quad (7)$$

One-dimensional framework for the knot selection

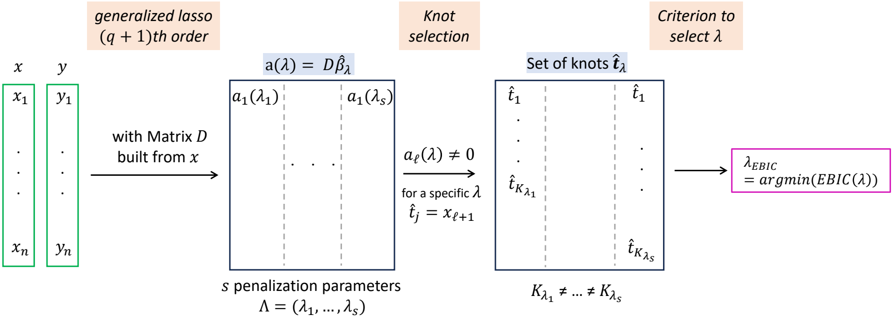


Figure 2: One-dimensional framework

Two-dimensional framework for the knot selection

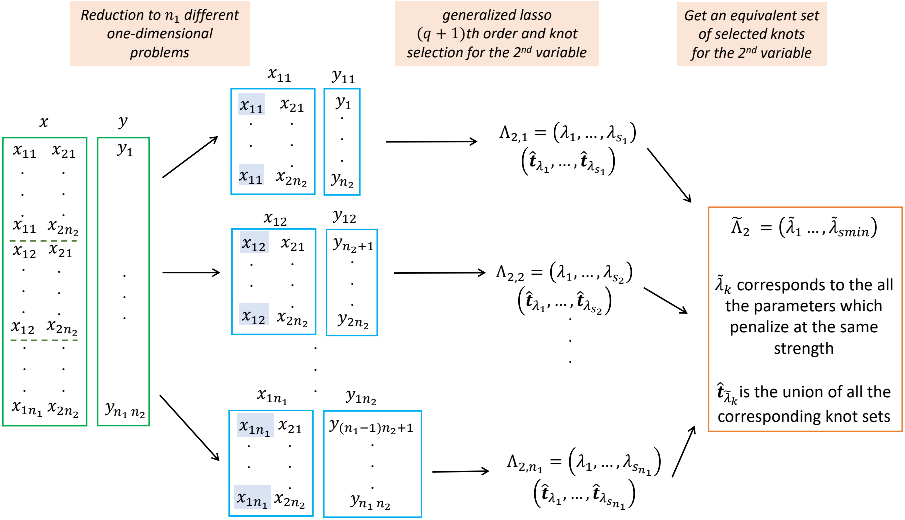


Figure 3: Two-dimensional framework

Selection of knot sets ($d = 2$)

Equivalent sets of knots – First dimension

$$\widetilde{\Lambda}_1 = \left\{\widetilde{\lambda}_{1,1}, \ldots, \widetilde{\lambda}_{1,s_{min_1}}\right\} \quad \text{and} \quad s_{min_1} = \min_{1 \leq i \leq n_2} s_i \ , \tag{8}$$

$$\widetilde{\lambda}_{1,k} = \left(\lambda_{(1,i),k}\right)_{1 \leq i \leq n_2}, \quad 1 \leq k \leq s_{min_1}. \tag{9}$$

In (9), $\widetilde{\lambda}_{1,k}$ can be seen as the vector of parameters which penalize (3) at an equivalent strength for each fixed value of x_2.

Equivalent sets of knots – Second dimension

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

Let us consider two generic penalization parameters $\widetilde{\lambda}_1$ belonging to $\widetilde{\Lambda}_1$ and $\widetilde{\lambda}_2$ belonging to $\widetilde{\Lambda}_2$.

Selection of knot sets ($d = 2$)

Equivalent sets of knots – First dimension

$$\widetilde{\Lambda}_1 = \left\{ \widetilde{\lambda}_{1,1}, \dots, \widetilde{\lambda}_{1,s_{min_1}} \right\} \quad \text{and} \quad s_{min_1} = \min_{1 \le i \le n_2} s_i \ , \tag{8}$$

$$\widetilde{\lambda}_{1,k} = \left(\lambda_{(1,i),k} \right)_{1 \le i \le n_2} , \quad 1 \le k \le s_{min_1} . \tag{9}$$

In (9), $\widetilde{\lambda}_{1,k}$ can be seen as the vector of parameters which penalize (3) at an equivalent strength for each fixed value of x_2.

Equivalent sets of knots – Second dimension

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

Let us consider two generic penalization parameters $\widetilde{\lambda}_1$ belonging to $\widetilde{\Lambda}_1$ and $\widetilde{\lambda}_2$ belonging to $\widetilde{\Lambda}_2$.

Which combination of penalization parameters $(\widetilde{\lambda}_1, \widetilde{\lambda}_2)$ to choose to get an optimal estimator of f ?

Selection criterion of the parameters $(\widetilde{\lambda}_1, \widetilde{\lambda}_2)$ $(d = 2)$

EBIC criterion

$$\text{EBIC}\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right) = \text{SS}\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right) + \widetilde{Q}_1 \widetilde{Q}_2 \log n + 2\log\left(\frac{(q+n_1+1)(q+n_2+1)}{\widetilde{Q}_1 \widetilde{Q}_2}\right). \tag{10}$$

where $\widetilde{Q}_1 = q + K_{\widetilde{\lambda}_1} + 1$ and $\widetilde{Q}_2 = q + K_{\widetilde{\lambda}_2} + 1$ and $\text{SS}\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right)$ is the sum of squares.

Final estimator of f

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

with $\widehat{f}_{\widetilde{\lambda}_1, \widetilde{\lambda}_2}$ defined as:

$$\widehat{f}_{\widetilde{\lambda}_1, \widetilde{\lambda}_2}(x) = \widehat{f}_{\widetilde{\lambda}_1, \widetilde{\lambda}_2}(x_1, x_2) = \sum_{i=1}^{\widetilde{Q}_1} \sum_{j=1}^{\widetilde{Q}_2} \widehat{\gamma}_{ij} B_{1,i,M}(x_1) B_{2,j,M}(x_2). \tag{11}$$

Metrics

One-dimensional form

$$\text{Normalized MAE}(\lambda) = \frac{1}{N}\sum_{k=1}^{N} \frac{\left|f(x_k) - \widehat{f}_\lambda(x_k)\right|}{f_{max} - f_{min}}. \quad (12)$$

$$\text{Normalized sup norm}(\lambda) = \max_{1\leq k\leq N} \frac{\left|f(x_k) - \widehat{f}_\lambda(x_k)\right|}{f_{max} - f_{min}}, \quad (13)$$

where $\widehat{f}_\lambda$ is defined in (1). In (13), 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.
f_{min} and f_{max} denote the minimum and maximum values of f evaluated on $\{x_1, \dots, x_N\}$, respectively.

Two-dimensional form

(12) and (13) with λ becomes $\widetilde{\lambda}_1$ and $\widetilde{\lambda}_2$ and $\widehat{f}_\lambda$ is replaced by $\widehat{f}_{\widetilde{\lambda}_1,\widetilde{\lambda}_2}$.

Results on geochemical applications ($d = 1$)

Function to estimate: Simple case of precipitation, we consider here one input (Spa) and one output (Amount of Salt) Savino et al. (2022). Real evaluations of f have been obtained with PHREEQC.

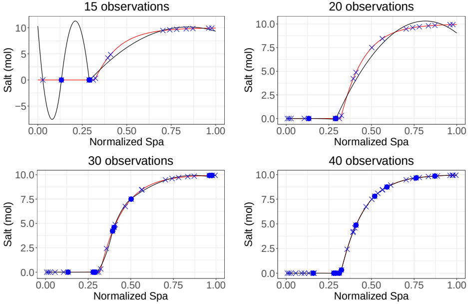


Figure 4: Illustration of the method over an increasing number of observations

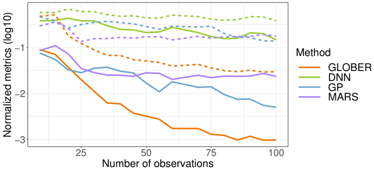


Figure 5: Statistic performance of our method (GLOBER) and of the state-of-the-art methods.The dashed (resp. solid) line displays the average of the Normalized Sup Norm (resp. Normalized MAE) values obtained from 10 replications.

Results on geochemical applications ($d = 2$)

Function to estimate: Simple case of precipitation, we consider here two inputs (Ca and Mg) and one output (Amount of Dolomite). Real evaluations of f have been obtained with PHREEQC.

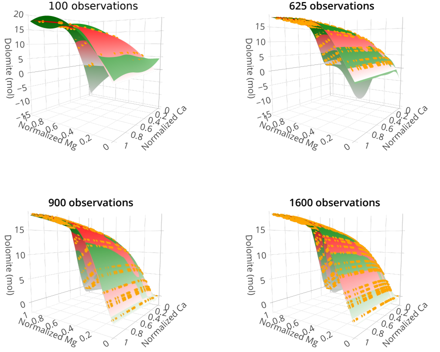


Figure 6: Illustration of the method over an increasing number of observations

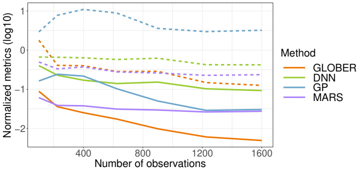


Figure 7: Statistic performance of our method (GLOBER) and of the state-of-the-art methods.The dashed (resp. solid) line displays the average of the Normalized Sup Norm (resp. Normalized MAE) values obtained from 10 replications.

Discussion and perspectives

- New way of estimating univariate and bivariate functions with B-splines
- Application to the one and two-dimentional settings
- Submitted article: M. E. Savino, C. 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**, *arxiv:2306.00686, 2023.*
- Implementation of the method: R package `glober` available on the CRAN, by using the `genlasso` R package (Arnold and Tibshirani, 2016).
- Extension to higher dimensional settings and to general grid ongoing

References

Arnold, T. B. and R. J. Tibshirani (2016). Efficient implementations of the generalized lasso dual path algorithm. Journal of Computational and Graphical Statistics 25(1), 1–27.

Chen, J. and Z. Chen (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.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. The Annals of Statistics 19(1), 1–67.

Savino, M., C. Lévy-Leduc, M. Leconte, and B. Cochepin (2022). An active learning approach for improving the performance of equilibrium based chemical simulations. Computational Geosciences 26(2), 365–380.

Tibshirani, R. J. and J. Taylor (2011). The solution path of the generalized lasso. The Annals of Statistics 39(3), 1335 – 1371.

Back-up slides

Definition of penalty matrix D

Case of evenly-spaced observations

$$D = D_{tf,q+1} = D_0 \cdot D_{tf,q} \quad q \geq 0, \tag{14}$$

with $D_{tf,0} = \mathrm{Id}_{\mathbb{R}^n}$, the identity matrix of $\mathbb{R}^n$

D_0 is the penalty matrix for the one-dimensional fused Lasso:

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

Case of unevenly-spaced observations

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

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

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

Selection of the knot set $(d = 1)$

Let $\Lambda = (\lambda_1, \dots, \lambda_k)$ be a grid of penalization parameters λ_i. We define $\mathbf{a}(\lambda)$ by:

$$\boldsymbol{a}(\lambda) = D \cdot \widehat{\beta}(\lambda), \quad \lambda \in \Lambda$$

Approach to find the selected knots associated to λ

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

where

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

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

Sum of square detailed for two-dimensional case

Definition of SS $\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right)$

$$\mathrm{SS}\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right) = \left\| \mathbf{Y} - \widehat{\mathbf{Y}}\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right) \right\|_2^2,$$

where

$$\widehat{\mathbf{Y}}\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right) = \mathbf{B}\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right) \widehat{\gamma}, \tag{15}$$

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

$$\mathbf{B}\left(\widetilde{\lambda}_1, \widetilde{\lambda}_2\right) = \mathbf{B}\left(\widetilde{\lambda}_1\right) \otimes \mathbf{B}\left(\widetilde{\lambda}_2\right), \tag{16}$$

$E \otimes F$ denoting the Kronecker product of the matrices E and F. In (16), $\mathbf{B}\left(\widetilde{\lambda}_1\right)$ is a $n_1 \times \widetilde{Q}_1$ matrix having as ith column $(B_{1,i,M}(x_{1k}))_{1 \leq k \leq n_1}$, i belonging to $\{1, \ldots, \widetilde{Q}_1\}$ and $\mathbf{B}\left(\widetilde{\lambda}_2\right)$ is a $n_2 \times \widetilde{Q}_2$ matrix having as jth column $(B_{2,j,M}(x_{2\ell}))_{1 \leq \ell \leq n_2}$, j belonging to $\{1, \ldots, \widetilde{Q}_2\}$.

State-of-the-art methods

- **Gaussian Processes (GP)**: squared exponential covariance function, implementation by using `scikit-learn` Python package,
- **Multivariate Adaptive Regression Splines (MARS):** interaction terms are included, implementation by using `earth` R package,
- **Deep Neural Networks (DNNs):** arbitrarily chosen since our goal is not to optimize it:
  - 2-hidden-layered structure composed of 10 neurons per layer
  - Activation function of the hidden layers: RELU function since it is one of the most used functions.
  - Optimizer: stochastic gradient descent method Adam
  - Loss function: the Mean Squared Error (MSE).
  - Number of epochs: 300 epochs for functions of $d = 1$ and 50 epochs for functions of $d = 2$ to avoid overfitting,

  implementation by using `keras` R package.

Suppplementary application for the two-dimensional framework

Function to estimate: Simple case of precipitation, we consider here two inputs (Spa and Spb) and one output (Amount of Halite). Real evaluations of f have been obtained with PHREEQC.

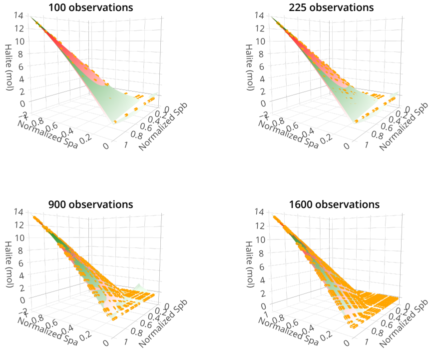


Figure 8: Illustration of the method over an increasing number of observations

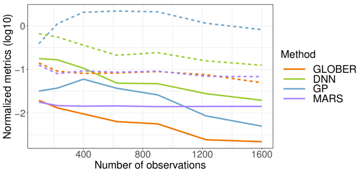


Figure 9: Statistic performance of our method (GLOBER) and of the state-of-the-art methods.The dashed (resp. solid) line displays the average of the Normalized Sup Norm (resp. Normalized MAE) values obtained from 10 replications.