Dimension reduction and associated industrial challenges

One functional input treatment in an optimization context

Miguel Munoz Zuniga[1] and **Yann Richet**[1]

[1] Institut de Radioprotection et de Sûreté Nucléaire

Initial model

Consider some input X define on a given space : $\mathbb{R}^p$, $L^2([0,1])$, $L^2(\Omega \times [0,1])$...

Eventually this input is linked to an output Y via an unknown function f.

$$Y = f(X)$$

How and what for

Goal :

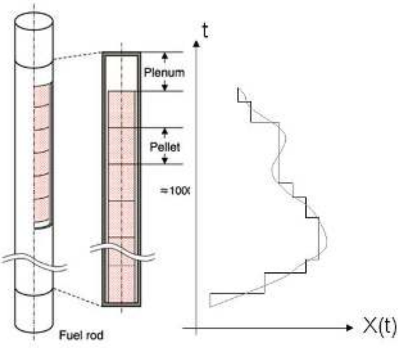
- find a sub-space on which the initial input can be projected without or with a minimum loss of information
  - independently of the output (unsupervised learning).
  - guided by the output (supervised learning).

Methods :

- Selection : eliminate the input dimensions which do not explain the phenomenon under study (sensitivity analysis, LASSO...)
- Extraction : optimally reorganized the information carried by the input with regard or not to an observable output (ACP, SIR,...)

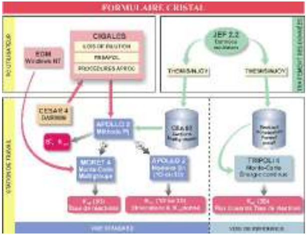
Context

Crayon après cycle de vie dans le réacteur nucleaire. On considère le taux de combustion mesuré en Megawatt-jour/tonnes sur l'axe verticale du crayon



Context

In nuclear power technology, burnup (also known as fuel utilization) is a measure of how much energy is extracted from a nuclear fuel assembly. It is measured both as the actual energy released per mass of initial fuel in gigawatt-days/metric ton of heavy metal (GWd/tHM)



The optimization problem

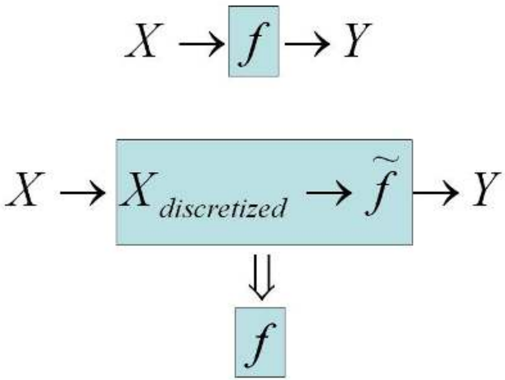
As the criticality coefficient depends on the combustion rate, identify the most penalizing rate leads to solve the following inverse optimization problem :

$$x^* = \operatorname*{argmax}_{x \in V} \; f(x). \tag{1}$$

where $V \subset L^2(\Omega \times [0,1])$ is a normed linear functional space.

To ensure the existence of the maximum and for the problem (1) to be well posed, some suitable hypothesis on the functional f and the set V should be verified (semicontinuity, compacity...).

The discretized model



Optimal constant piecewise approximation

We need to convert a smooth curve into an optimal piecewise constant function ($L = 12$ intervals). What is the optimal time partition for the following criterion

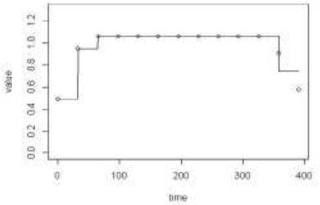
$$\min_{t_2,\dots,t_{L-1}} \left\| f - \sum_{j=1}^{L-1} m_i \mathbb{1}_{[t_j, t_{j+1}]} \right\|_{L^2[0,1]}$$

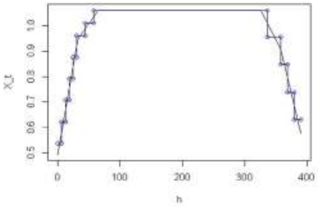
where

$$m_i = \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} f(t) dt$$

(Why L_2 ?)

Optimal constant piecewise approximation





Splines (with optimal knots) are better

de Boor, C. (1973).
Good approximation by spline with variable knots.
Internat. Ser. Numer. Math., 21 :57–72

- Nasty non linear problem
- Any computational scheme has to be content to find, by some descent method a locally best approximate (even quite expensive)

Let's look to bounds on

$$dist_\infty(f, S_n^k) = \inf_{g \in S_n^k} \|f - g\|_\infty$$

$$S_n^k = \bigcup_{t \in T} S_t^n \text{ and } S_t^n = \{\text{spline of degree} < n \text{ with knots } t\}$$

Splines (with optimal knots) are better

Buchard, H. (1974).
Splines (with optimal knots) are better.
Appl. Anal., 3 :309–319

If $f \in C^n[0,1]$, $N \geq N_f$, $1 \leq p \leq \infty$ then

$$dist_p(f, S_n^k) \leq d_n k^{-n} \|f^{(n)}\|_\sigma \text{ with } \sigma = \frac{1}{n + 1/p}$$

Pence, D. and Smith, P. (1982).
Asymptotic properties of best $L_p[0,1]$ approximation by splines.
SIAM J. Math. Anal., 13(3) :409–420

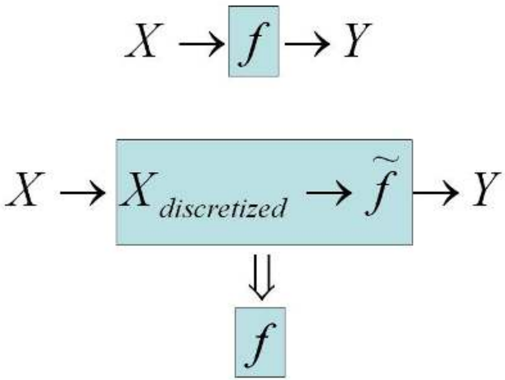
$\Rightarrow$ A functional proof of optimality.

Optimal constant piecewise approximation

Under monotony hypothesis the problem has a unique solution. We don't have monotony hypothesis, the solution can be multiple (symmetry). We run a BFGS algorithm with starting point $\{t_i\}_{i=2,\ldots,L-1}$

$$t_i = H^{-1}(\frac{i}{L}) \quad \text{and} \quad H(t) = \int_0^t |f'(s)|^{2/3} ds / \int_0^1 |f'(s)|^{2/3} ds$$

The discretized model



The stochastic process

Let $X(t)$ be a mean square continuous process defined on [0,1], i.e.

$$X(t) \in L^2(\Omega \times [0,1]) = \{S(t) : \Omega \to \mathbb{R}, t \in [0,1] \,|\, \mathbb{E}\int_0^1 S(t)^2 dt < +\infty\}$$

We suppose given a functional sample of size n

$$\{X_i(t), t \in [0,1]\}_{i=1,\dots,n} \quad \text{and} \quad \{Y_i\}_{i=1,\dots,n}$$

known on $\{t_1, \dots, t_p\}$ a uniform discretization of [0,1] :

$$\{X_i(t_j)\}_{\substack{i=1,\dots,n \\ j=1,\dots,p}}$$

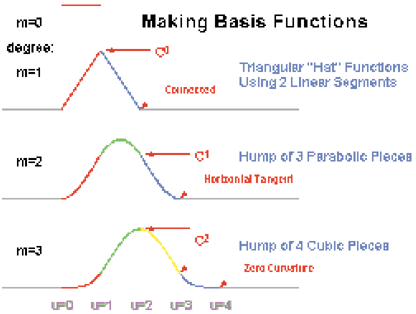
Smoothing/Projecting Data

The value of the functional samples on $\{t_1, ..., t_p\}$:

$$\{X_i(t_j)\}_{\substack{i=1,...,n \\ j=1,...,p}}$$

are use to obtain the smooth functions $X_i(t)$ via either B-spline or wavelet decompositions.

Smoothing/Pojecting Data : B-spline



$$\boldsymbol{X}(t) = \left(X_1(t), ..., X_n(t)\right)^t$$

$$= \left(\sum_{k=1}^{K} c_{1k}\phi_k(t) + \epsilon_1(t), ..., \sum_{k=1}^{K} c_{nk}\phi_k(t) + \epsilon_n(t)\right)^t$$

Smoothing/Pojecting Data : B-spline

$$\begin{aligned}
\boldsymbol{X}(t) &= \left(X_1(t), \ldots, X_n(t)\right)^t \\
&= \left(\sum_{k=1}^{K} c_{1k}\phi_k(t) + \epsilon_1(t), \ldots, \sum_{k=1}^{K} c_{nk}\phi_k(t) + \epsilon_n(t)\right)^t \\
&= \mathbf{C}\phi(t) + \boldsymbol{\epsilon}(t) \\
&\simeq \tilde{\mathbf{C}}\phi(t)
\end{aligned}$$

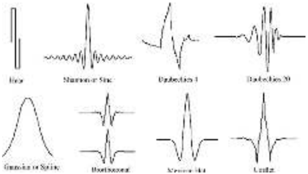
Where the ϵ's are zero mean, uncorrelated and $\tilde{\mathbf{C}}$ is obtained by least square or penalized least square on the discretized sample $\left\{X_i(t_j)\right\}_{\substack{i=1,\ldots,n \\ j=1,\ldots,p}}$

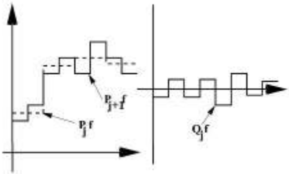
Smoothing/Projecting Data : Wavelet

Under the considered hypothesis the stochastic process has the following wavelet decomposition

$$X(t) = \sum_{k=0}^{2^{j_0}-1} \xi_{j_0,k}\phi_{j_0,k} + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \eta_{j,k}\psi_{j,k}$$

where $\xi_{j_0,k} = \langle \phi_{j_0,k}, X \rangle$ and $\eta_{j,k} = \langle \psi_{j,k}, X \rangle$ are sequences of random variables.





Smoothing/Projecting Data : Wavelet

Under the considered hypothesis the stochastic process has the following wavelet decomposition

$$X(t) = \sum_{k=0}^{2^{j_0}-1} \xi_{j_0,k}\phi_{j_0,k} + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \eta_{j,k}\psi_{j,k}$$

where $\xi_{j_0,k} = \langle \phi_{j_0,k}, X \rangle$ and $\eta_{j,k} = \langle \psi_{j,k}, X \rangle$ are sequences of random variables. Approximated by it's 'empirical' projection onto V_J

$$X^J(t) = \sum_{k=0}^{2^{j_0}-1} \tilde{\xi}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \tilde{\eta}_{j,k}\psi_{j,k}$$

with $\tilde{\xi}_{j_0,k} = p^{-1}\sum_{j=1}^{p-1} X(t_j)\phi_{j_0,k}(t_j)$, $\tilde{\eta}_{j,k} = p^{-1}\sum_{j=1}^{p-1} X(t_j)\psi_{j,k}(t_j)$ and $p = 2^J$.

The model

$$\begin{aligned} Y &= f\big(X(t)\big) + \epsilon \\ &= g\big(\langle \beta_1, X\rangle, ..., \langle \beta_K, X\rangle, \epsilon'\big) + \epsilon \\ &= g\big(\theta_1, ..., \theta_K, \epsilon'\big) + \epsilon \\ &= \tau(\boldsymbol{\theta}) + Z(\boldsymbol{\theta}) \end{aligned} \quad (2)$$

where ϵ, ϵ' are centered random variable independent of $X(t), \{\beta_i(t), i = 1...K\}$ are K orthonormal functions in $L^2\big([0,1]\big)$, f, g smooth functions, $\langle .,. \rangle$ the standard inner product on $L^2\big([0,1]\big)$, Z a centered Gaussian process and τ a trend.

The algorithm

- Data as discretized functional data
- Smoothing data : B-spline or Wavelet, estimate decomposition coefficients and parameters
- Find, β's, the reduced space base functions : PCA, MAVE, SIR
- Sample a number of curves in the previous basis, i.e sample a number of coordinates θ's
- Run EGO algorithm on the model with inputs the θ's + $(*)$
- Transpose the found optimal θ^* onto the 'smooth' functional space

$(*)$ Transform the obtained smooth function into the considered data (piecewise)

The empirical approximations

For any $s, t \in [0,1]$ and $v \in L^2[0,1]$, we suppose and denote :

- $\mu(t) = \mathbb{E}\big(X(t)\big) = 0$
- $\sigma^2(t) = \mathbb{E}\big(X(t)^2\big)$
- $\gamma(t,s) = \mathbb{E}\big(X(t)X(s)\big)$
- $(\Gamma v)(t) = \int_0^1 \gamma(t,s)v(s)ds$

We consider the empirical estimators :

- $\hat{\sigma}^2(t) = (n-1)^{-1}\sum_{i=1}^n X_i(t)^2 = (n-1)^{-1}\mathbf{X}(t)^2$
- $\hat{\gamma}(t,s) = (n-1)^{-1}\sum_{i=1}^n X_i(t)X_i(s) = (n-1)^{-1}\mathbf{X}'(t)\mathbf{X}(s)$
- $(\hat{\Gamma} v)(t) = \int_0^1 \hat{\gamma}(t,s)v(s)ds$

Functional Principal Component analysis (1/2)

$$X(t) = \sum_{i=1}^{+\infty} \langle \beta_i, X \rangle \, \beta_i(t) \text{ (Karhunen-Loeve)}$$

- Find the β's functions minimizing $\mathbb{E}\left(\|X - \sum_{i=1}^{K} \langle \beta_i, X \rangle \, \beta_i\|^2\right)$
- The solutions are the K eigenfunctions with largest eigenvalues of the following eigenequation :

$$(\Gamma \beta)(t) = \lambda \beta(t) \sim (\hat{\Gamma} \beta)(t) = \lambda \beta(t)$$

- Keep the eigenfunction for which the eigenvalue correspond to more than 1% of the total variance

Functional Principal Component analysis (2/2)

Considering a B-spline expansion of the functional inputs $\boldsymbol{X}(t) = \tilde{\mathbf{C}}\phi$ and the basis eigenfunction $\beta(t) = \phi(t)b'$ solve the eigenequation

$$(\hat{\Gamma}\beta)(t) = \lambda\beta(t)$$

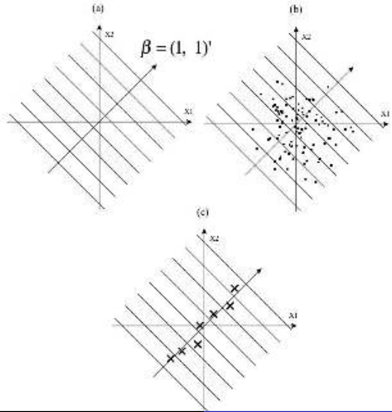
is equivalent to solve

$$(n-1)^{-1}\tilde{\mathbf{C}}'\tilde{\mathbf{C}}\mathbf{W}\mathbf{b} = \lambda\mathbf{b}$$

where $\mathbf{W} = \int \phi(t)\phi'(t)dt$ and $W_{ij} = \int \phi_i(t)\phi_j(t)dt$

Multidimensional Sliced Inverse Regression

Find β's minimizing $\mathbb{E}\left(\|\mathbb{E}(X|Y) - \sum_{i=1}^{K} \langle \beta_i, \mathbb{E}(X|Y)\rangle \beta_i\|^2\right)$



Wavelet Sliced Inverse Regression (1/2)

Under Generalized Linear Design Condition

- Find β's minimizing $\mathbb{E}\left(\|\mathbb{E}(X|Y) - \sum_{i=1}^{K} \langle \beta_i, \mathbb{E}(X|Y) \rangle \beta_i\|^2\right)$
- For any t, consider the BINWAV wavelet estimator of $\mathbb{E}(X(t)|Y = y)$ with design points the Y_i's : $\hat{M}_y(t)$
- Consider $\hat{\Gamma}$ and estimate the covariance operator of $\mathbb{E}(X|Y)$: $\hat{\Gamma}_e = n^{-1} \sum_{i=1}^{n} \hat{M}_{Y_i} \otimes \hat{M}_{Y_i}$

Wavelet Sliced Inverse Regression (2/2)

Under Generalized Linear Design Condition

- Evalute spectral decomposition of $\hat{\Gamma}_e$ and it's projection $\hat{\Gamma}_e^{k_n}$
- Evaluate spectral decomposition of $\hat{\Gamma}^{1/2}(\hat{\Gamma}_e^{k_n})^{+}\hat{\Gamma}^{1/2} : (\alpha_i, \eta_i)_{i=1,\dots,K}$
- EDR directions are given by $\beta_i = \alpha_i^{-1}(\hat{\Gamma}_e^{k_n})^{+}\hat{\Gamma}^{1/2}\eta_i$
- Use chi-square sequential test for determining the number of effective dimension

Wavelet Minimum Average Variance Estimation (1/4)

- Find β's minimizing $\mathbb{E}\Big(\big[Y - \mathbb{E}(Y|\langle\beta_1, X\rangle, ..., \langle\beta_K, X\rangle)\big]^2\Big)$
- Consider a wavelet expansion of the functional inputs

$$X^J(t) = \sum_{k=0}^{2^{j_0}-1} \tilde{\xi}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \tilde{\eta}_{j,k}\psi_{j,k}$$

and the basis

$$\beta_i^J(t) = \sum_{k=0}^{2^{j_0}-1} c^i_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} d^i_{j,k}\psi_{j,k}$$

- Let $\tilde{\gamma}^J$ gather $X^J(t)$ expansion coefficient and β_i^J gather $\beta_i^J(t)$ expansion coefficients then

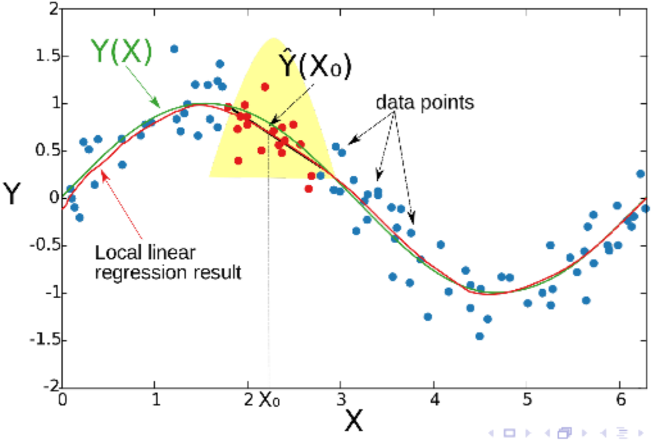
Wavelet Minimum Average Variance Estimation (2/4)

$$\begin{aligned}
\mathbb{E}\left(Y | \langle\beta_1, X\rangle, \ldots, \langle\beta_K, X\rangle\right) &= \mathbb{E}\left(Y | \left\langle \boldsymbol{\beta}_1^J, \tilde{\gamma}^J \right\rangle, \ldots, \left\langle \boldsymbol{\beta}_K^J, \tilde{\gamma}^J \right\rangle\right) \\
&= \mathbb{E}\left(Y | B' \tilde{\gamma}^J\right) \\
&\simeq a + \mathbf{b}' B' (\tilde{\gamma}^J - \tilde{\gamma}_0^J)
\end{aligned}$$

Solve the minimization problem

$$\min_{\substack{B : B'B = Id \\ a_l, \mathbf{b}_l}} \sum_{i=1}^{n} \sum_{l=1}^{n} w_{il} \left[Y_i - (a_l + \mathbf{b}_l' B' (\tilde{\gamma}_i^J - \tilde{\gamma}_l^J)) \right]^2$$

Wavelet Minimum Average Variance Estimation (3/4)



Wavelet Minimum Average Variance Estimation (4/4)

- Estimate weigths w_{il} depending on the distance $\tilde{\gamma}_i^J - \tilde{\gamma}_l^J$ by kernel approximation. Can be adaptative : RWMAVE
- Find the dimension of the dimension reduction space : Cross-validation

Kriging (1/2)

$$Y = g(\boldsymbol{\theta}) + \epsilon = \tau(\boldsymbol{\theta}) + Z(\boldsymbol{\theta})$$

- Consider the trend of the form

$$\tau(\boldsymbol{\theta}) = \sum_{j=1}^{l} \delta_j f_j(\boldsymbol{\theta})$$

- Consider Gaussian (hyp.) sample (Transform it into uniform)

$$\{\boldsymbol{\theta}_i\}_{i=1,\ldots,N} \text{ or } \{\boldsymbol{\theta}_i\}_{i=1,\ldots,n}$$

- Consider covariance kernel

$$C(\mathbf{u},\mathbf{v}) = \sigma^2 R(\mathbf{u}-\mathbf{v},\psi)$$

Kriging (1/2)

- Then $Y(\theta)|\theta_i's \sim \mathcal{N}\left(m_{UK}(\theta), s^2_{UK}(\theta)\right)$ where

$$
\begin{aligned}
m_{UK}(\theta) &= \mathbf{f}(\theta)'\hat{\delta} + \mathbf{c}(\theta)'C^{-1}(\mathbf{y} - F\hat{\delta}) \\
s^2_{UK}(\theta) &= s^2_{SK}(\theta) + (\mathbf{f}'(\theta) - \mathbf{c}(\theta)'C^{-1}F)'(F'C^{-1}F)^{-1}(\mathbf{f}'(\theta) - \mathbf{c}(\theta)'C^{-1}F) \\
s^2_{SK}(\theta) &= C(\hat{\theta}, \hat{\theta}) - \mathbf{c}(\theta)'C^{-1}\mathbf{c}(\theta) \\
\hat{\delta} &= (F'C^{-1}F)^{-1}F'C^{-1}\mathbf{y}
\end{aligned}
$$

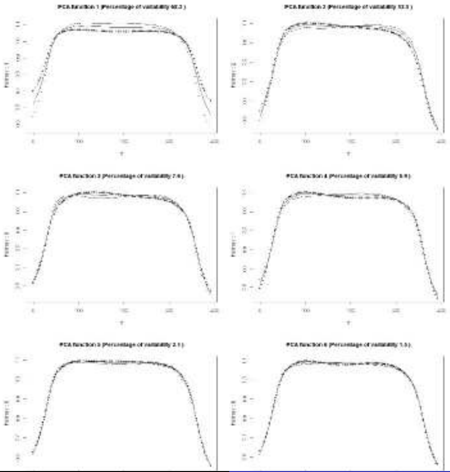
Expected improvement criterion

$$
\begin{aligned}
EI(\theta) &= \mathbb{E}\Big[\big(\min(Y_i) - Y(\theta)\big)^+ | Y_i\Big] \\
&= \big(\min(Y_i) - m(\theta)\big)\Phi\Big(\frac{\min(Y_i) - m(\theta)}{s(\theta)}\Big) + s(\theta)\phi\Big(\frac{\min(Y_i) - m(\theta)}{s(\theta)}\Big)
\end{aligned}
$$

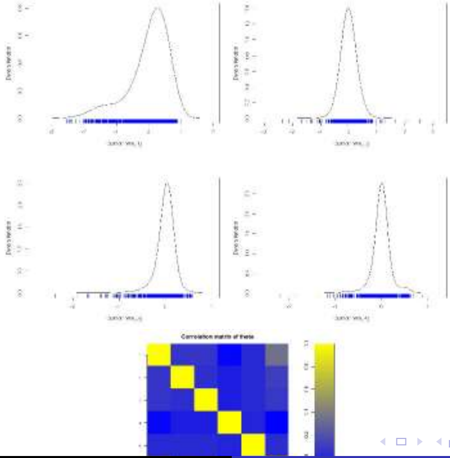
Expected Improvement

- Evaluate $Y_i = Y(\theta_i)$ and estimate covariance parameters by maximum likelihood.
- Compute $\theta_{new} = argmaxEI(\theta)$
- Evaluate $Y(\theta_{new})$
- Re-estimate covariance parameters and update kriging taking into account the new design point

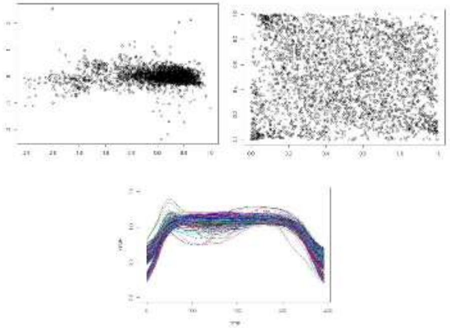
Functional principal components



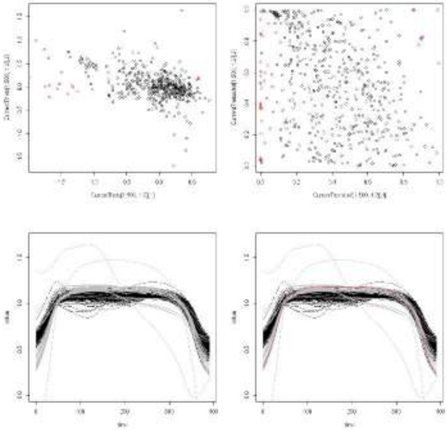
Coordinates distributions



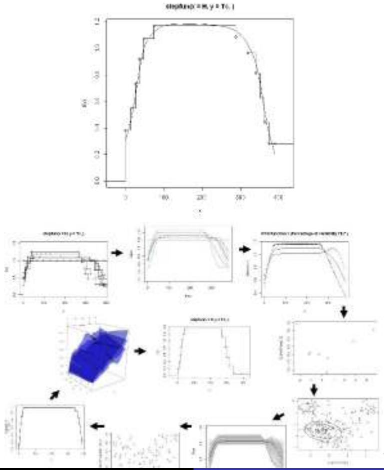
Projected functions



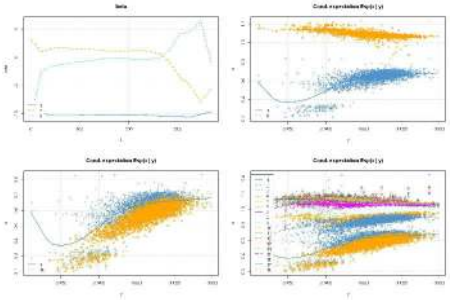
EGO results



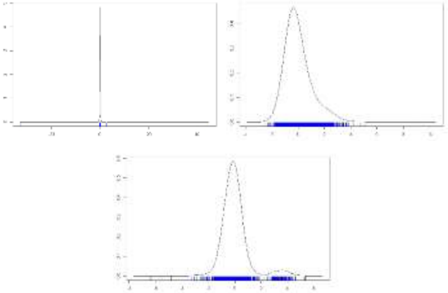
Optimum



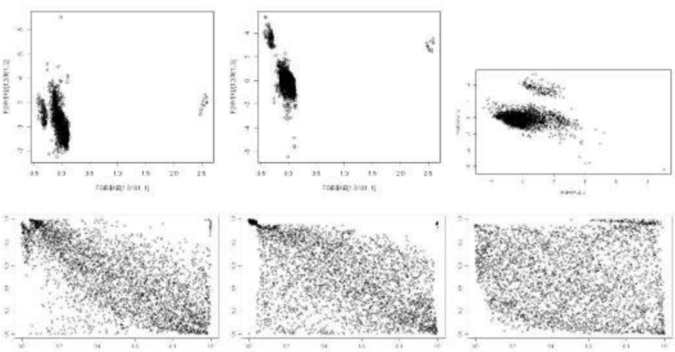
EDR basis and conditional expectations



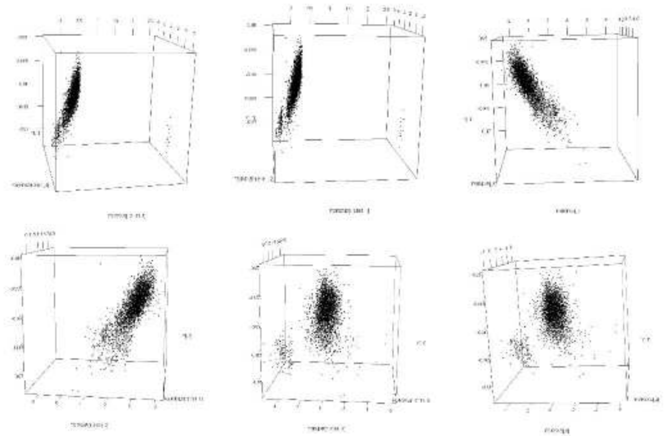
Coordinates distributions



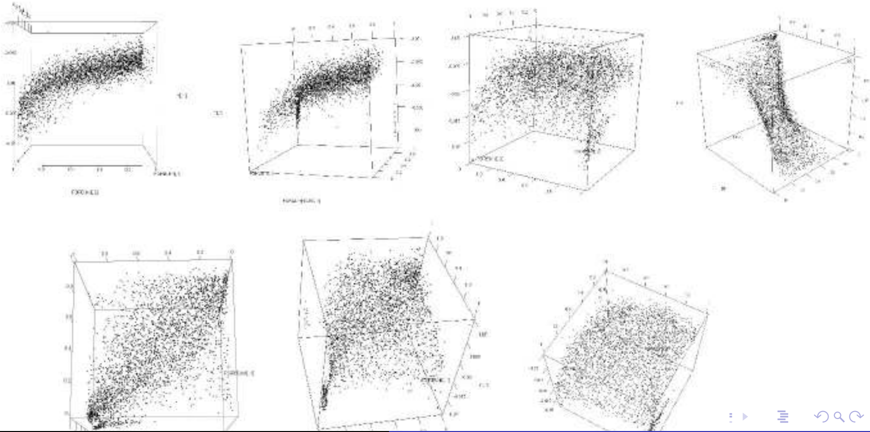
Coordinates in EDR basis



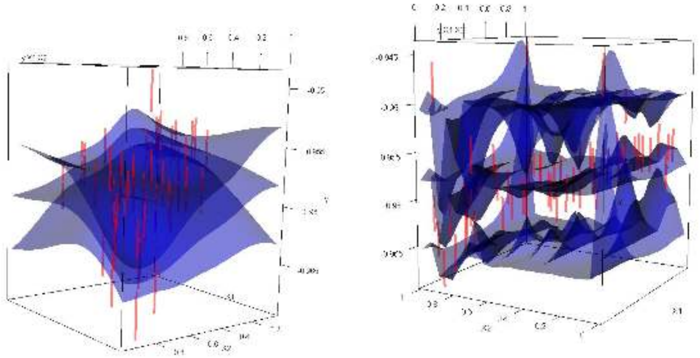
Coordinates against output



Coordinates in the uniform space



Kriging and EGO



Coordinates of EDR basis