Intrinsic Kriging and Prior Information

Emmanuel Vazquez

Eric Walter

– Laboratoire des Signaux et Systèmes *CNRS* –

– Supélec –

– Université Paris Sud –

Part One: Kriging

- named after D.G. Krige (1951)
- classical theory – (Matheron, 1963)
- intrinsic theory – (Matheron, 1973)
- well-established in Geostatistics
- connections with time-series prediction
- also known under the name of Gaussian processes

Objective: predict system output from observed data using black-box modelling

System modeled by a random process $F(\mathbf{x})$

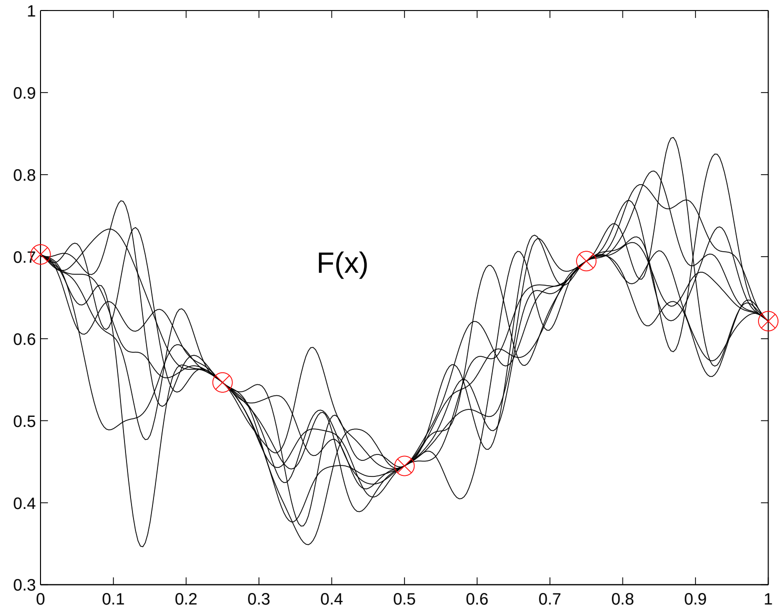


Figure 1: Conditional simulations of $F(x)$ given $F(x_i)$

Linear prediction on $\mathcal{H}_S = \text{vect}\{F(\mathbf{x}_1), \cdots, F(\mathbf{x}_n)\}$

$$\hat{F}(\mathbf{x}) = \sum_{i=1}^{n} \lambda_{i,\mathbf{x}} F(\mathbf{x}_i), \tag{1}$$

Best linear prediction

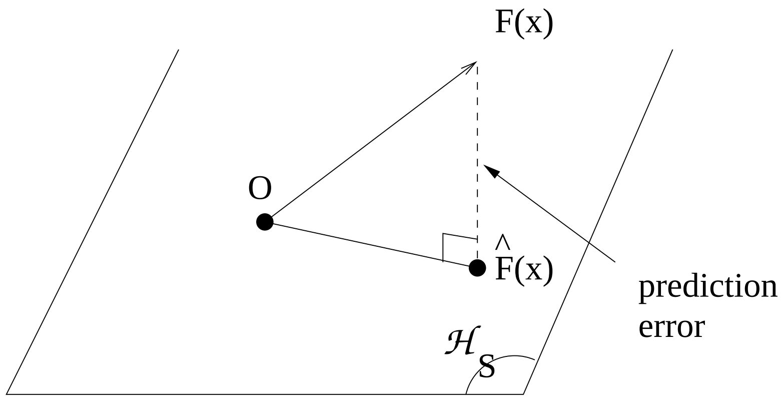


Figure 2: Orthogonal projection

Three points of view, one structure

(Assume $\mathsf{E}F(\mathbf{x}) = 0$)

1. $F(\mathbf{x})$ generates a Hilbert space $\mathcal{H}$, whose elements are limits of linear combinations

$$\sum_{i=1}^{n} \lambda_{i,\mathbf{x}} F(\mathbf{x}_i) \tag{2}$$

$\mathcal{H}$ endowed with scalar product

$$(F(\mathbf{x}), F(\mathbf{y}))_{\mathcal{H}} = \mathsf{E}[F(\mathbf{x})F(\mathbf{y})] = k(\mathbf{x}, \mathbf{y})$$

2. Define the finite support measure

$$\lambda = \sum_{i=1}^{n} \lambda_{i,\mathbf{x}} \delta_{\mathbf{x}_i} \tag{3}$$

The set of these measures generates another Hilbert space Λ

– Λ can be identified to $\mathcal{H}$

– inherits the scalar product of $\mathcal{H}$

$$(\lambda, \mu)_\Lambda = \sum_{i,j} \lambda_i k(\mathbf{x}_i, \mathbf{x}_j) \mu_j \tag{4}$$

3. Λ can also be viewed as the dual of an *rkhs*, or a *feature space* $\mathcal{F}$ in SVM theory.

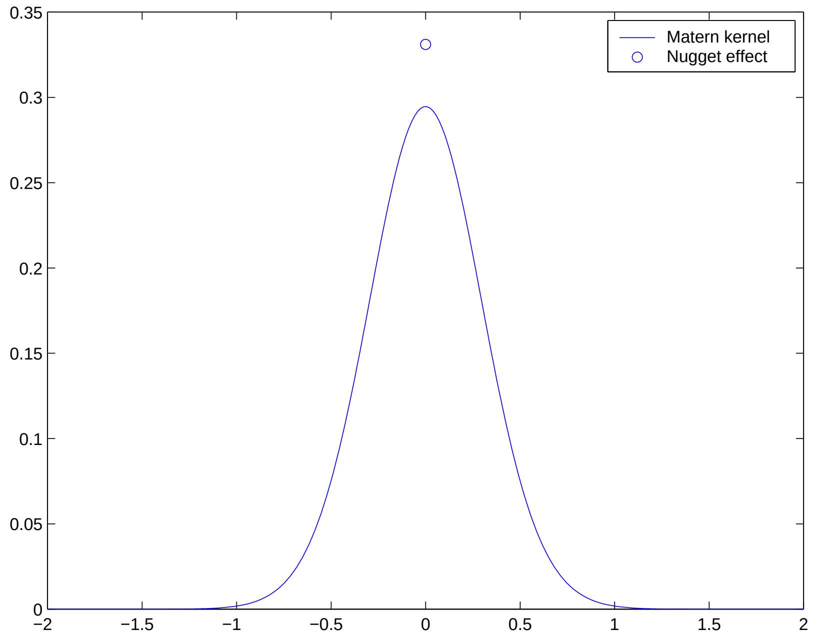
The covariance $k(\mathbf{x}, \mathbf{y})$ is the reproducing kernel of $\mathcal{F}$

$\mathcal{F}$ and $\mathcal{H}$ share the same structure.

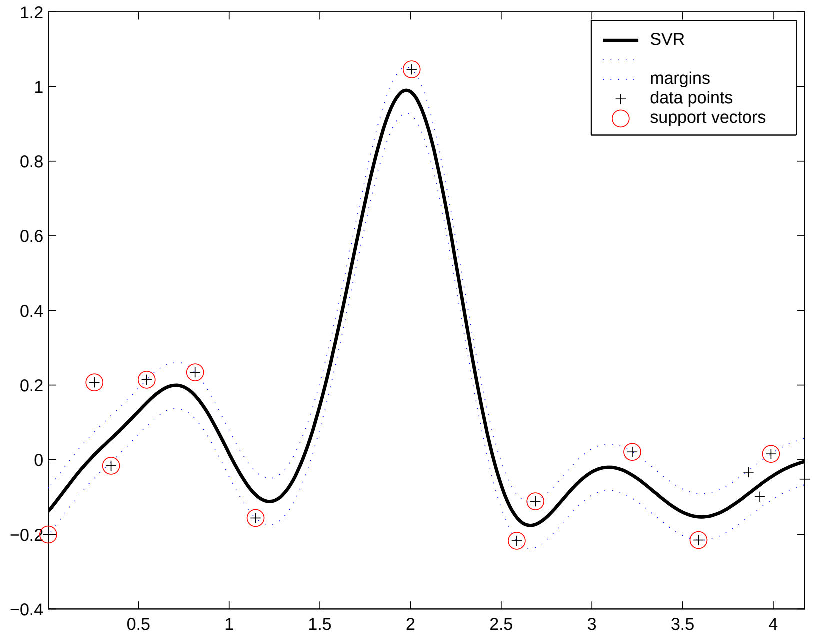
$\Rightarrow$ Kriging can be used to build SVM

Example

Kriging lore



applied in SVR



Regularization and intrinsic Kriging

Regularization minimizes

$$\underbrace{||\hat{f}||^2_{\mathcal{F}}}_{\text{smoothness}} \quad + \quad \underbrace{C\sum_i l(\hat{f}(\mathbf{x}_i) - f_{\mathbf{x}_i})}_{\text{data fidelity}} \tag{5}$$

- Strongly connected with Kriging (Wahba and Kimeldorf, 1970) and (Matheron, 1981)
- Intrinsic Kriging (I.K.) facilitates incorporation of prior knowledge
- In I.K., $||.||_{\mathcal{F}}$ replaced by a seminorm $\Rightarrow$ null-space N of the seminorm not regularized $\rightarrow$ N should correspond to prior information (Smola et al., 1999).

For prediction by I.K., error of prediction required to be be orthogonal to N.

- Define $\Lambda' \subset \Lambda$ such that $\Lambda' \perp N$ (more rigourously $\Lambda' = \{\lambda, \langle \lambda, g \rangle_{\mathcal{F}^*,\mathcal{F}} = 0 \quad \forall g \in N\}$).
- Λ' corresponds to $\mathcal{H}' \subset \mathcal{H}$, with $\mathcal{H}' \perp N$.
- Thus, in I.K. prediction error required to be in $\mathcal{H}'$.

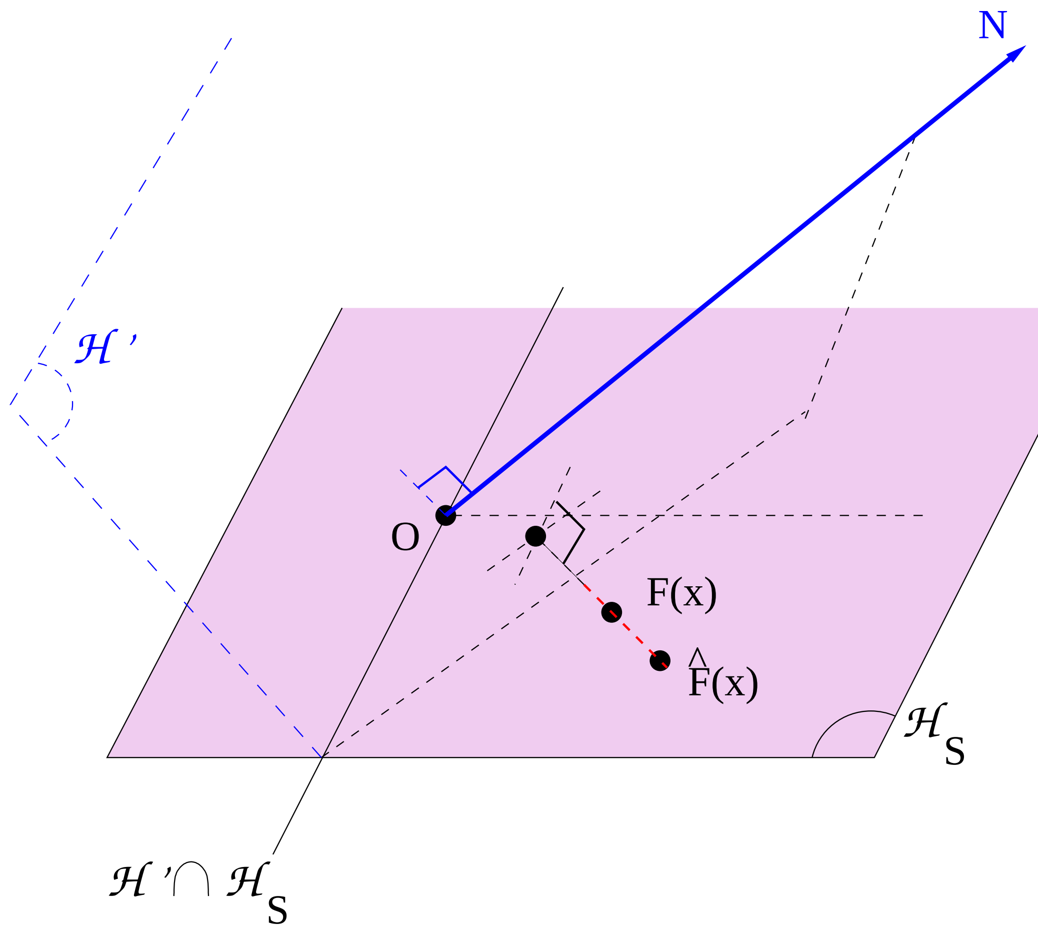


Figure 3: Best approximation given $\hat{F}(\mathbf{x}) - F(\mathbf{x}) \perp N$

Simple example of I.K.

Assume variance of $F(\mathbf{x}_1) - F(\mathbf{x}_2)$ is stationary, *i.e.*,

$$\mathsf{Var}(F(\mathbf{x}_1) - F(\mathbf{x}_2)) = 2\gamma(||\mathbf{x}_1 - \mathbf{x}_2||) = 2\gamma(h) \tag{6}$$

Then

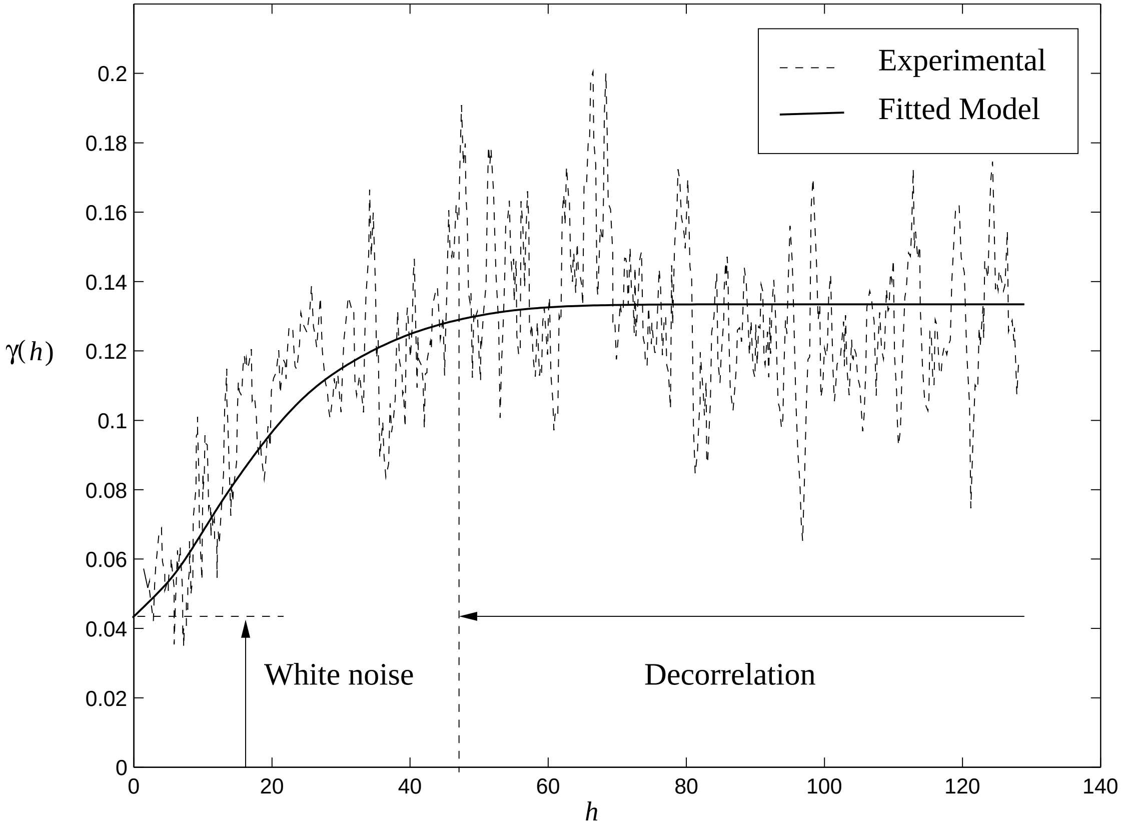
- $\gamma(h)$ called variogram
- Nullspace N made of constant functions

Canonical decomposition:

$$F(\mathbf{x}) = G(\mathbf{x}) + b \text{ with } \mathsf{E}G(\mathbf{x}) = 0 \tag{7}$$

I.K. is a semi-parametric formulation of Kriging

Example of variogram

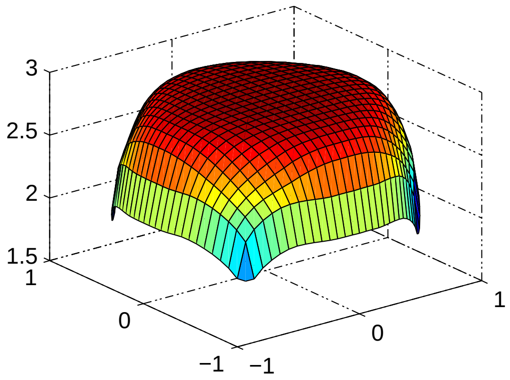


Part Two: Incorporating prior information

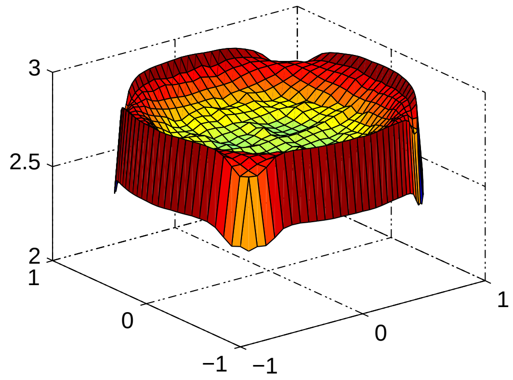
Application to flow measurement

collaboration with Services des Mesures Supélec

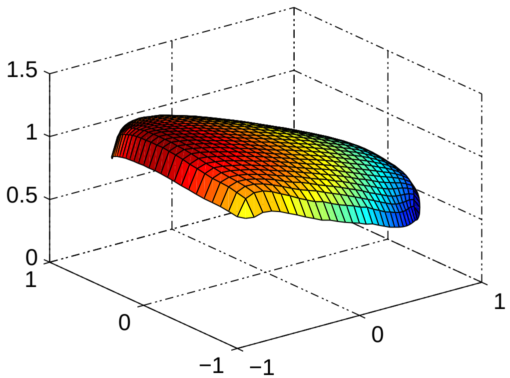
- Estimation of flow in a pipe from punctual observations of speed of fluid;
- Desired performance: relative error $< 1\%$;
- Many fluid speed profiles have been simulated, for different types of pipes.



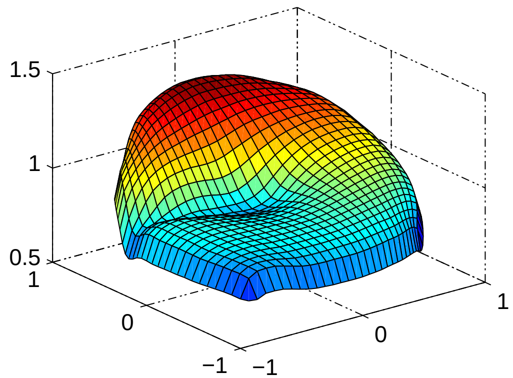
(A)



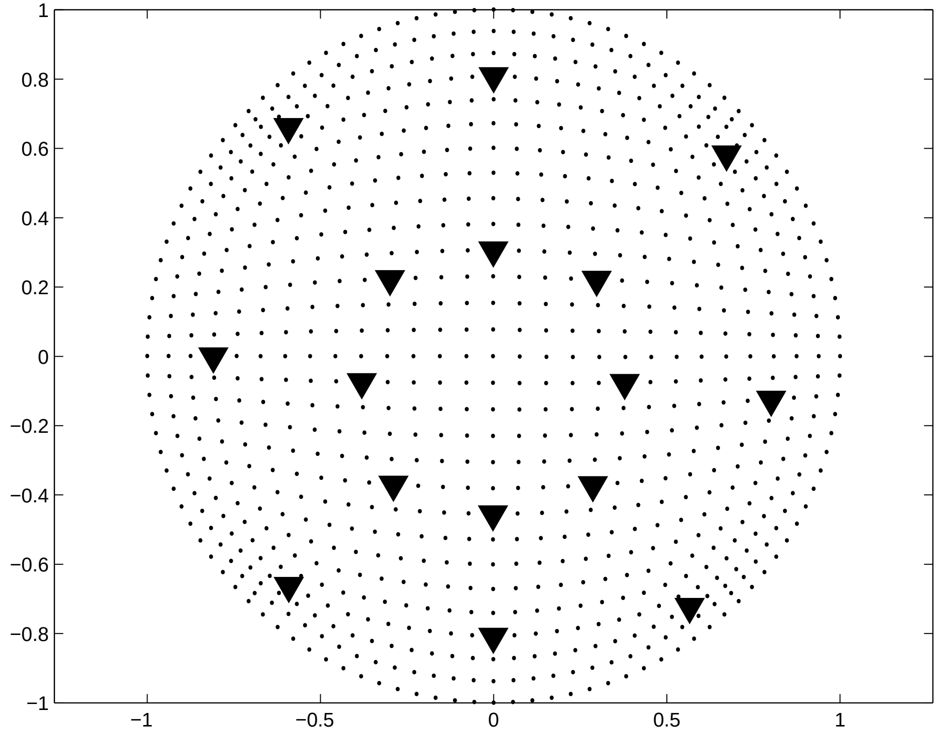
(B)



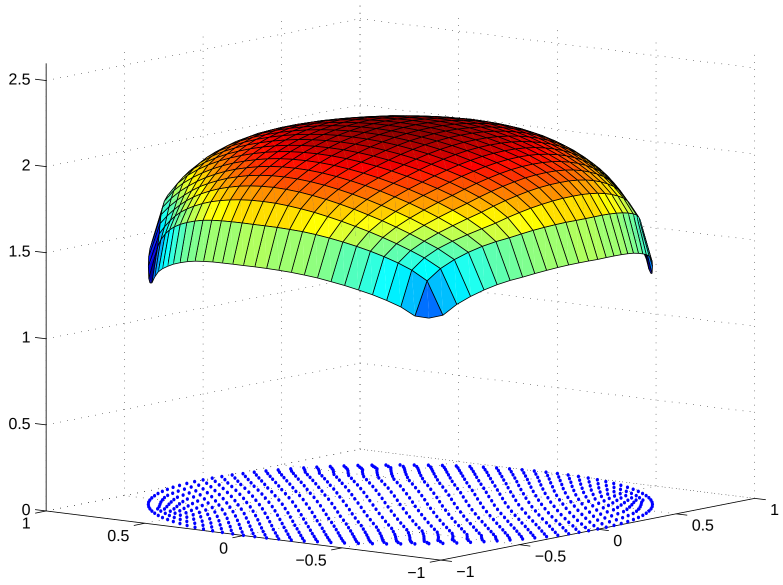
(C)



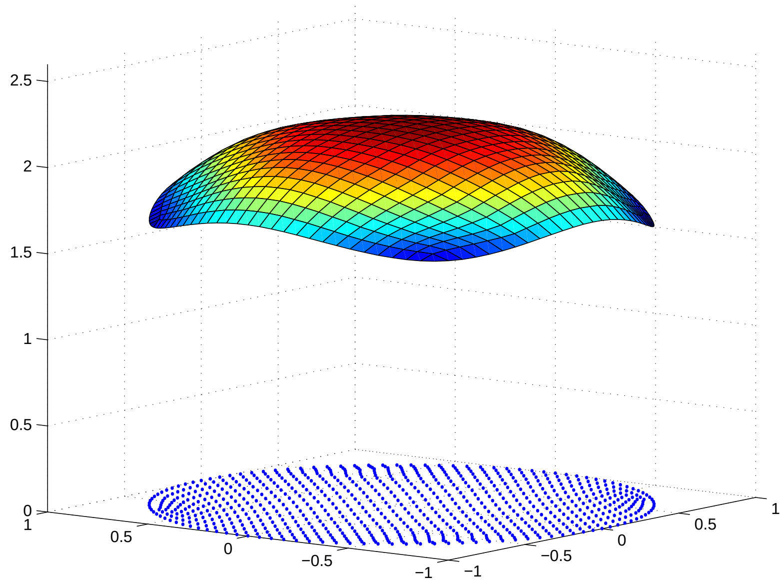
(D)


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1
−0.5
0
0.5
1

Direct estimation of speed profile is not sufficient.

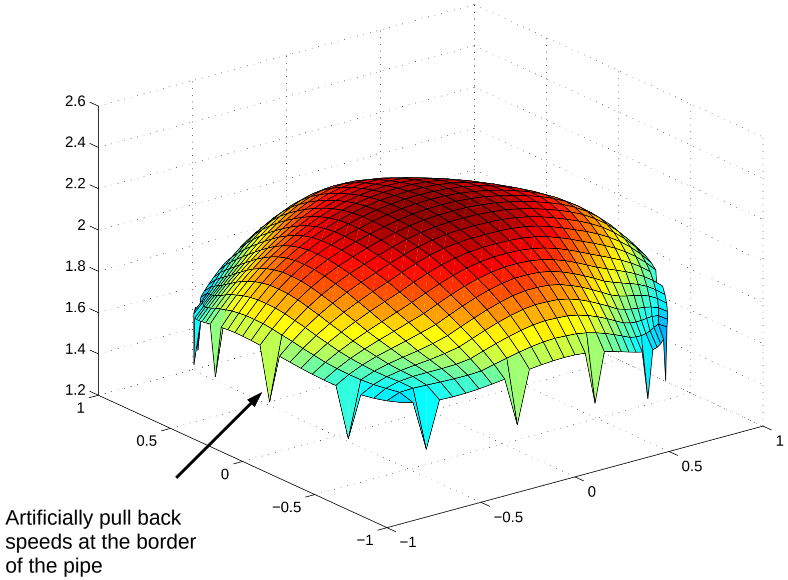


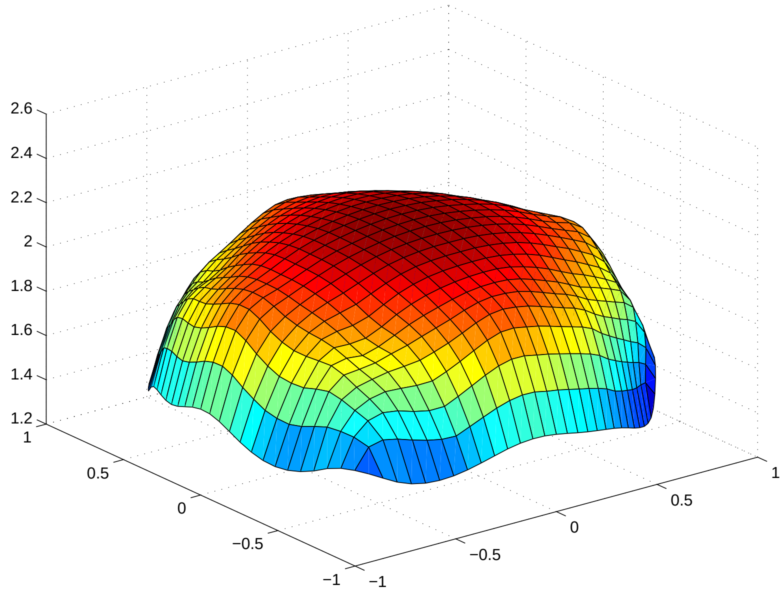
Actual profile



Predicted profile

First idea: artificially add the trend





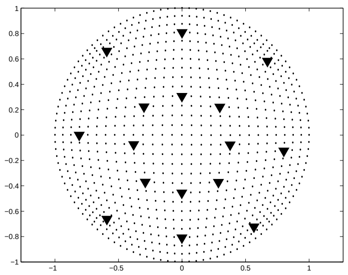
Including prior information via I.K.

Principle:

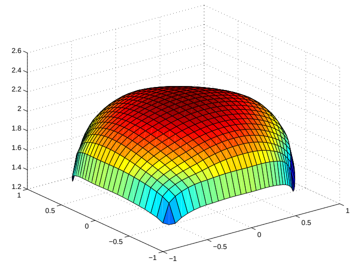
- Add new factors $\mathbf{x}^*$ corresponding to prior information. For instance, at each position $\mathbf{x}$ of the cross section of the pipe, add a scalar factor x^* indicating nominal speed there.
- For prior knowledge to be preserved by regularization it should be introduced in the nullspace N. This is done under the form

$$\mathbf{x}, \mathbf{x}^* \mapsto \mathbf{a}^\top \mathbf{x}^* + \mathbf{b} \tag{8}$$

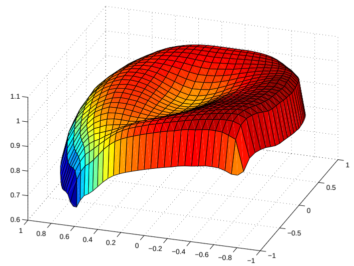
- $\mathbf{a}$ estimated by I.K. Prediction error orthogonal to $\mathbf{x}^*$.



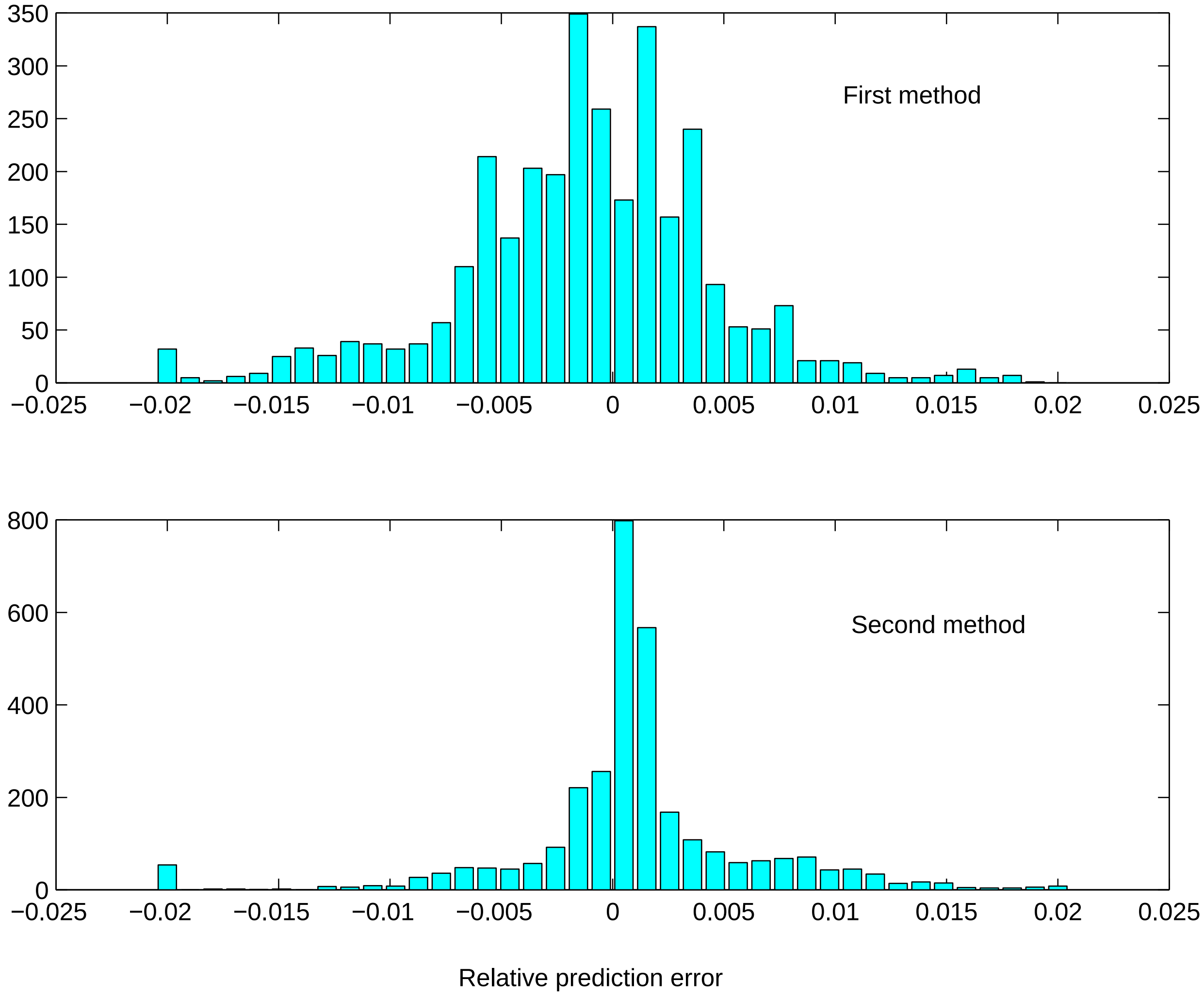
$+$



$\longrightarrow$



observations	prior knowkedge	estimation
$f(\mathbf{x}_i)$	x^*	$\hat{f}(\mathbf{x}, x^*)$


First method
Second method
Relative prediction error
350
300
250
200
150
100
50
0
800
600
400
200
−0.025
−0.02
−0.015
−0.01
−0.005
0.005
0.01
0.015
0.02
0.025

Conclusions

- including prior knowledge in black-box modelling made easy via a semi-parametric approach
- good results obtained in a real application
- I.K. theory allows to interpret regularization methods (Splines, RBF, SVR) in a probabilistic framework
- interpretation of a kernel as the covariance of a random process helps choosing a kernel for a given application

References

Matheron, G., 1963. Principles of geostatistics. Economic Geology 58, 1246–1266.

Matheron, G., 1973. The intrinsic random functions, and their applications. Adv. Appl. Prob. 5, 439–468.

Matheron, G., 1981. Splines and Kriging: their formal equivalence. In: D.F., M. (Ed.), Down-to-Earth Statistics: Solutions Looking for Geological Problems, syracuse univ. of geology contributions Edition. pp. 77–95.

Smola, A., Friess, T., Schölkopf, B., 1999. Semiparametric support vector and linear programming machines. In: Kearns, M., Solla, S., Cohn, D. (Eds.), Advances in Neural Information Processing Systems. 11. MIT Press, Cambridge, pp. 585 – 591.

Wahba, G., Kimeldorf, G., 1970. Spline functions and stochastic processes. Sankhyä: the Indian Journal of Statistics: Series A 32 (2), 173–180.