Agnostic Multi-Fidelity Gaussian Process Regression for Modelling in Aerospace Applications

(and future perspectives)

Giulio Gori[1,3], Olivier Le Maître[2,3] and Pietro M. Congedo[3]

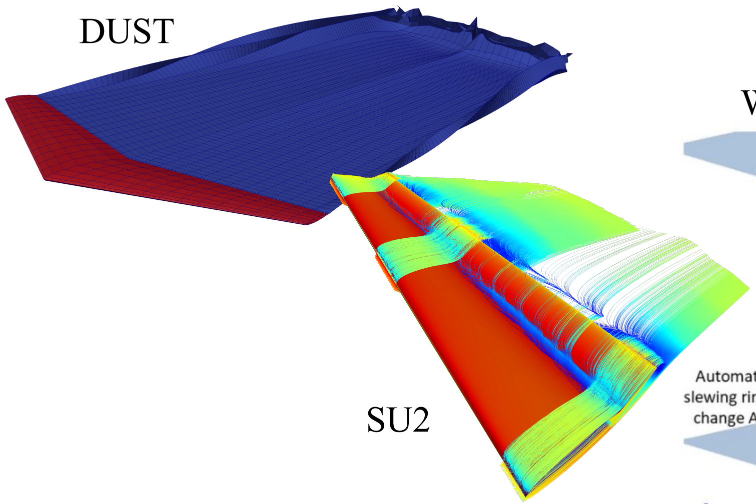
September, 10th 2021

[1] Post-doc, Department of Aerospace Science and Technology, Politecnico di Milano, Via La Masa 34, 20156, Milano, Italy
[2] CNRS
[3] Platon team, INRIA/CMAP, École Polytechnique, IPP, Route de Saclay, 91128, Palaiseau, France

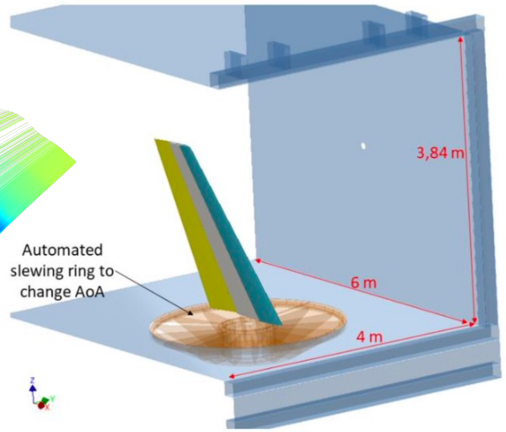
The MONNALISA Project

MOdelling **N**o**N**linear **A**erodynamics of **LI**fting **S**urf**A**ces

Focused on the development of a novel methodology to generate informative data, both from experiments and numerical simulations, and use them to increase the accuracy of low-order models for an extended range of conditions.



Wind tunnel experiments



Develop Uncertainty Quantification techniques to **select the most critical configurations** on which expensive investigation methods will be employed, the remaining configurations being evaluated by means of cheaper tools.

To produce a reliable database concerning the aerodynamic characteristics of unconventional tail-plane surfaces.

Research question

In **multi-fidelity** approaches, information of diverse fidelity and complexity complement each other, leading to **improved estimate accuracy** and to a **minimization of the cost** associated with parametrization.

Require **establishing of a fidelity hierarchy** among the available models (data), introducing **inherent modelling biases!**

Examples:
Typically (but not always correctly), people assume that resources requirements are proportional to accuracy.

The **more** expensive, the **better** it is!

People also generally assume that experiments are more accurate than computations

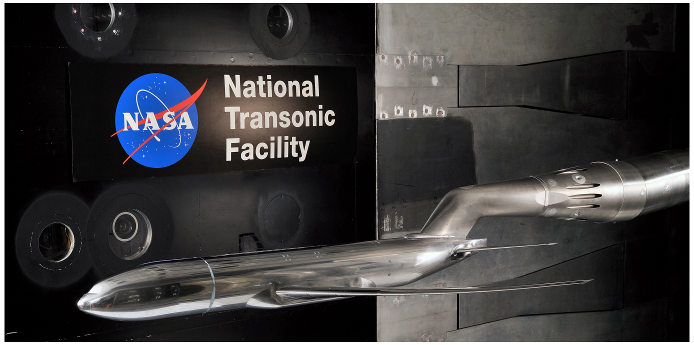
Computations **must** predict the experiment!

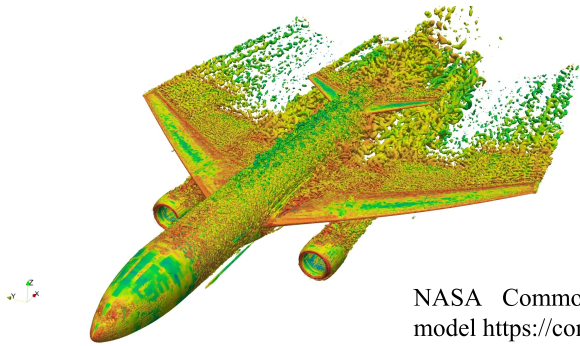
Research question

In **multi-fidelity** approaches, information of diverse fidelity and complexity complement each other, leading to **improved estimate accuracy** and to a **minimization of the cost** associated with parametrization.

Require **establishing of a fidelity hierarchy** among the available models, introducing **inherent modelling biases!**

A resounding example is the NASA Common Research Model (CRM). The CRM test cases were devised for the **purpose of validating specific applications of Computational Fluid Dynamics (CFD).**





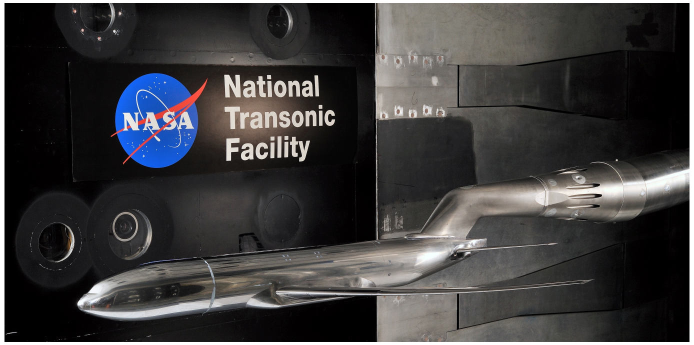
NASA Common Research Model (CRM) test model https://commonresearchmodel.larc.nasa.gov

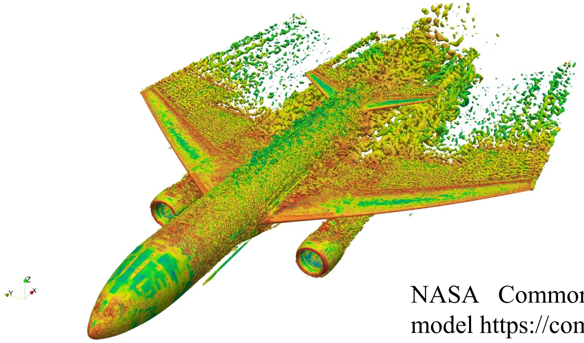
Research question

In **multi-fidelity** approaches, information of diverse fidelity and complexity complement each other, leading to **improved estimate accuracy** and to a **minimization of the cost** associated with parametrization.

Require **establishing of a fidelity hierarchy** among the available models, introducing **inherent modelling biases!**

A resounding example is the NASA Common Research Model (CRM). The CRM test cases were devised for the **purpose of validating specific applications of Computational Fluid Dynamics (CFD).**





NASA Common Research Model (CRM) test model https://commonresearchmodel.larc.nasa.gov

Is it possible to develop an **agnostic** multi-fidelity framework for improving prediction accuracy?

Part I:
developing an agnostic multi-fidelity modelling framework

Setting the mindset

We refer to the **reality of interest** as a physical process described by an unknown model

$$f : \Omega \subset \mathbb{R}^d \longmapsto \mathbb{R}$$

The reality of interest can possibly be measured or approximated by models M, either computational or experimental (we will make no distinction), to collect data of different fidelity

Being $l = 1, \ldots, L$ and L our reality of interest

$X^l = \{x_{1,l}, \ldots, x_{N^l,l}\}$ be the set of N^l training points s.t. $x_{n,l} \in \Omega \subset \mathbb{R}^d$

$Y^l = \{y_{1,l}, \ldots, y_{N^l,l}\}$ be a set of noisy observations, assuming that $y_{n,l} = M^l(x_{n,l}) + \epsilon_{n,l}$

Note that the indexing $l = 1, \ldots, L\text{-}1$ does not indicate any particular ordering or preferences among the available models

Gaussian Process

A **Gaussian (or stochastic) process** is a collection of indexed normal random variables generalizing the concept of a probability distribution to functions

GPs are supervised lazy Machine Learning (ML) models that can be employed, among many other scopes, to approximate multidimensional functions and to ultimately make predictions

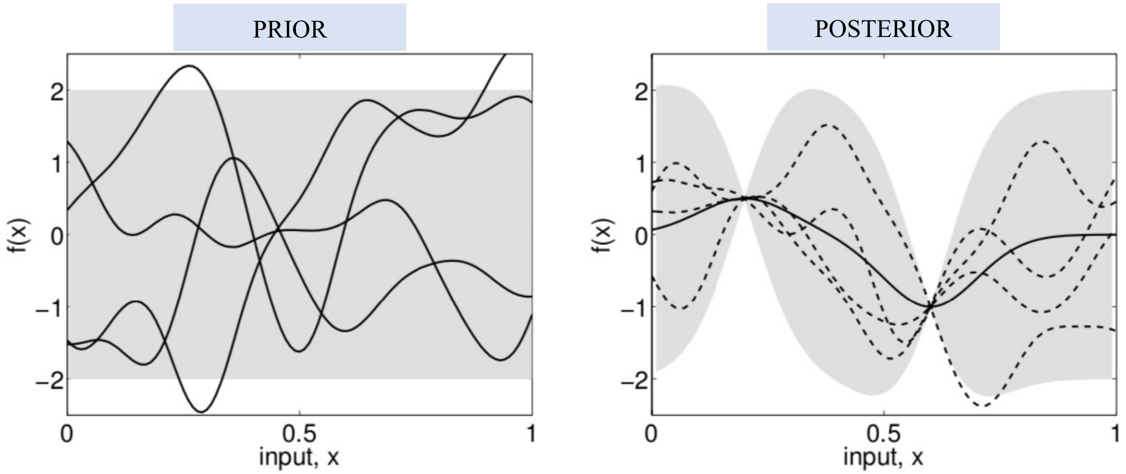


Image from C.E. Rasmussen and C.K.I. Williams, "Gaussian Processes for Machine Learning", *the MIT Press,* 2006 ISBN 026218253X

In regression problems, a GP is defined as a **function approximator** $\mathcal{GP}\left(f_l^T(x)\beta_l, K\left(x,x';\Theta^l\right)\right)$

The prediction is not just an estimate for that point, but also has **uncertainty**.

The ***kernel function K*** measures the similarity between observations and allows for **predicting** the value at an unseen point.
A **parameterised kernel** is typically used to fit a Gaussian process to data (**model selection**).

Gaussian Process

A **Gaussian (or stochastic) process** is a collection of indexed normal random variables generalizing the concept of a probability distribution to functions

GPs are supervised lazy Machine Learning (ML) models that can be employed, among many other scopes, to approximate multidimensional functions and to ultimately make predictions

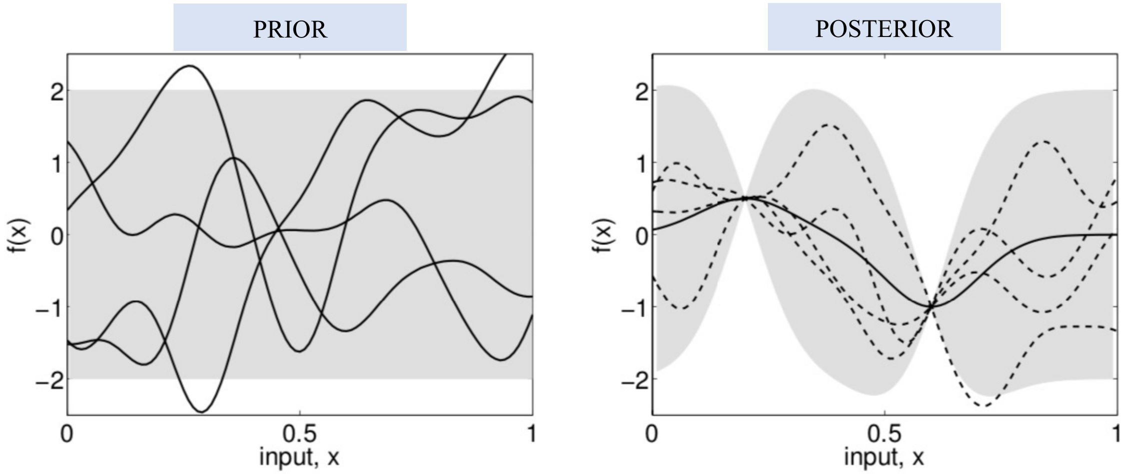


Image from C.E. Rasmussen and C.K.I. Williams, "Gaussian Processes for Machine Learning", *the MIT Press,* 2006 ISBN 026218253X

The kernel is the crucial ingredient in a Gaussian process predictor, as it encodes our assumptions about the function which we wish to learn

$$K(\boldsymbol{x}, \boldsymbol{x}'|\Theta) \coloneqq \theta_1 \exp\left(\frac{-(\boldsymbol{x}-\boldsymbol{x}')^2}{2\theta_2^2}\right) + \theta_3 \tilde{\delta}(\boldsymbol{x}-\boldsymbol{x}')$$

Gaussian Process

A **Gaussian (or stochastic) process** is a collection of indexed normal random variables generalizing the concept of a probability distribution to functions

GPs are supervised lazy Machine Learning (ML) models that can be employed, among many other scopes, to approximate multidimensional functions and to ultimately make predictions

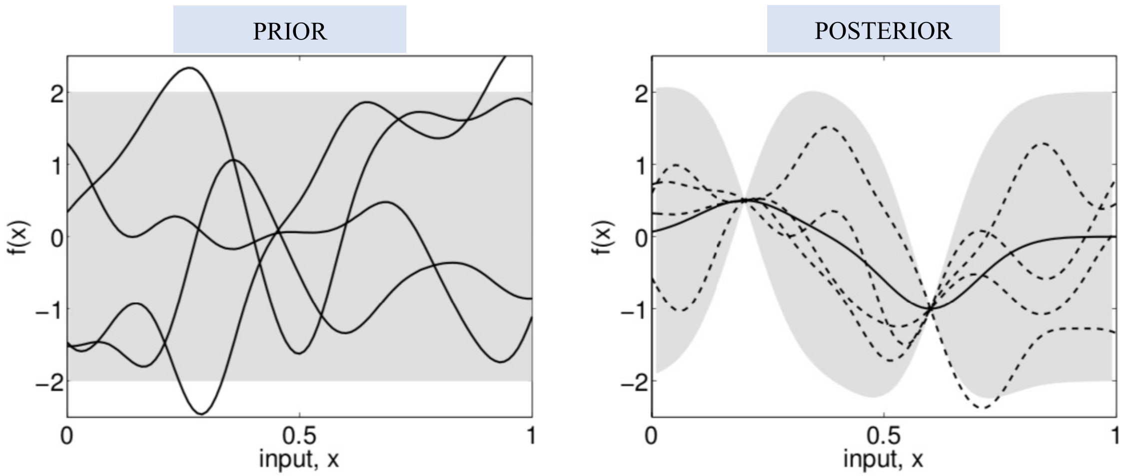


Image from C.E. Rasmussen and C.K.I. Williams, “Gaussian Processes for Machine Learning”, *the MIT Press,* 2006 ISBN 026218253X

Models can be fitted according to different methods e.g., a Maximum Likelihood approach

$$\mathcal{L}(Y|X,\Theta,\sigma_n) = -\frac{1}{2}Y^T\left[K(\boldsymbol{x},\boldsymbol{x}'|\Theta) + \sigma_n^2\underline{\mathrm{I}}\right]^{-1}Y - \frac{1}{2}\log\left|K(\boldsymbol{x},\boldsymbol{x}'|\Theta) + \sigma_n^2\underline{\mathrm{I}}\right|$$

Gaussian Process

A **Gaussian (or stochastic) process** is a collection of indexed normal random variables generalizing the concept of a probability distribution to functions

GPs are supervised lazy Machine Learning (ML) models that can be employed, among many other scopes, to approximate multidimensional functions and to ultimately make predictions

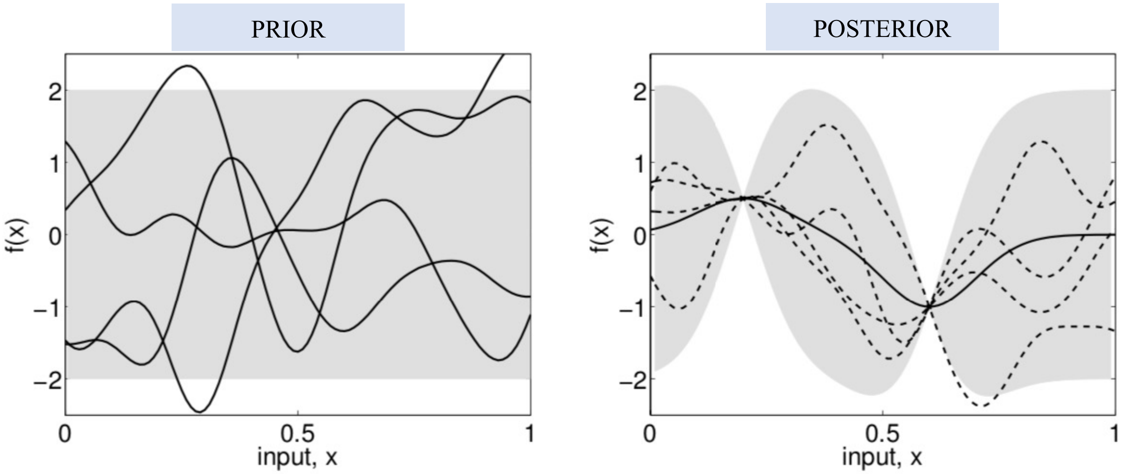


Image from C.E. Rasmussen and C.K.I. Williams, "Gaussian Processes for Machine Learning", *the MIT Press,* 2006 ISBN 026218253X

The mean prediction and the associated uncertainty can be evaluated according to:

$$\mu(x^*) \triangleq \mathbb{E}(f|Y, X, x^*) = K(x^*, X\,|\Theta)[K(X, X\,|\Theta) + \sigma_n^2 I]^{-1} Y$$

$$C(x^*) = K(x^*, x^*|\Theta) - K(x^*, X\,|\Theta)[K(X, X\,|\Theta) + \sigma_n^2 I]^{-1} K(X, x^*|\Theta)$$

Multi-fidelity surrogates: Kennedy-O'Hagan formulation[1]

They considered predictions and uncertainty analysis from complex computer codes which can be run at different level of sophistication, defining the following autoregressive model

$$\begin{cases} M^{l}(x) = \rho_{(l-1)}(x) M^{(l-1)}(x) + \delta_l, \\ M^{(l-1)}(x) \perp \delta_l(x), \\ \rho_{(l-1)}(x) = g_{l-1}^{T}(x) \beta_{\rho_{(l-1)}} \end{cases}$$

being

$$\delta_l \sim \mathcal{GP}\left(f_l^T(x)\beta_l, K\left(x, x'; \Theta^l\right) \right)$$

The training of this model is **quite expensive** because it requires the inversion of the covariance matrix arising from the construction of an L-level co-kriging model.

[1] Kennedy, M.C. and O'Hagan, A., *Predicting the output from a complex computer code when fast approximations are available*, Biometrika, 87, 1, pp.1-13, 2000

Multi-fidelity surrogates: Le Gratiet formulation[2]

The recursive formulation proposed in [2] substitutes the full model $M^{(l-1)}$ with its approximation $\mathcal{M}^{(l-1)}$ by means of a Gaussian process modelling the response of the lower level

$$\begin{cases} \mathcal{M}^{l}(x) = \rho_{(l-1)}(x)\mathcal{M}^{(l-1)}(x) + \delta_l, \\ \mathcal{M}^{(l-1)}(x) \perp \delta_l(x), \\ \rho_{(l-1)}(x) = g_{l-1}^{T}(x)\beta_{\rho_{(l-1)}} \end{cases}$$

In Ref.[2], the authors show that **building *L* independent kriging models is equivalent to building a *L*-level co-kriging model**.

This allows a model **complexity reduction** since the inversion of *L* “smaller” covariance matrices is less expensive than inverting a single large covariance matrix.

[2] Le Gratiet, L., Multi-fidelity Gaussian process regression for computer experiments, Ph.D. Thesis, Université Paris-Diderot - Paris VII, 2013. Français.

Multi-fidelity surrogates: Le Gratiet formulation[2]

The recursive formulation proposed in [2] substitutes the full model $M^{(l-1)}$ with its approximation $\mathcal{M}^{(l-1)}$ by means of a Gaussian process modelling the response of the lower level

$$\begin{cases} \mathcal{M}^{l}(x) = \rho_{(l-1)}(x)\mathcal{M}^{(l-1)}(x) + \delta_l, \\ \mathcal{M}^{(l-1)}(x) \perp \delta_l(x), \\ \rho_{(l-1)}(x) = g_{l-1}^{T}(x)\beta_{\rho_{(l-1)}} \end{cases}$$

In the following we will assume that

$$\delta_l \sim \mathcal{GP}\left(f_l^T(x)\beta_l, K\left(x, x'; \Theta^l\right) \right) \text{ with } \; f_l^T = \mathbf{0} \;\; \forall l$$

$$g_{(l-1)}^T = [1] \text{ which reduces the length of } \beta_{\rho_{(l-1)}} \text{ to one}$$

In the following we will aim at inferring $\rho_{(l-1)}$ rather than $\beta_{\rho_{(l-1)}}$

[2] Le Gratiet, L., Multi-fidelity Gaussian process regression for computer experiments, Ph.D. Thesis, Université Paris-Diderot - Paris VII, 2013. Français.

Multi-fidelity surrogates: Le Gratiet formulation[2]

The recursive formulation proposed in [2] substitutes the full model $M^{(l-1)}$ with its approximation $\mathcal{M}^{(l-1)}$ by means of a Gaussian process modelling the response of the lower level

$$\mu^{l}(x) = \rho_{(l-1)}\mu_{(l-1)}(x) + \mathcal{K}^{l}\left(x, X^{l} \middle| \rho_{(l-1)}, \Theta^{l}\right)\left[\mathcal{K}^{l}\left(X^{l}, X^{l} \middle| \rho_{(l-1)}, \Theta^{l}\right) + \sigma_{l}^{2}\underline{\mathrm{I}}\right]^{-1}\left(Y^{l} - \rho_{(l-1)}\mu_{(l-1)}(x)\right),$$

And

$$C^{l}(x, x') = \mathcal{K}^{l}\left(x, x' \middle| \rho_{(l-1)}, \Theta^{l}\right) - \left[\mathcal{K}^{l}\left(x, X^{l} \middle| \rho_{(l-1)}, \Theta^{l}\right)\right]\left[\mathcal{K}^{l}\left(X^{l}, X^{l} \middle| \rho_{(l-1)}, \Theta^{l}\right) + \sigma_{l}^{2}\underline{\mathrm{I}}\right]^{-1}\left[\mathcal{K}^{l}\left(X^{l}, x' \middle| \rho_{(l-1)}, \Theta^{l}\right)\right],$$

Where, being $\odot$ the element by element matrix product, $\mathcal{K}$

$$\mathcal{K}^{l}\left(A, B \middle| \rho_{(l-1)}, \Theta^{l}\right) = \rho_{(l-1)}^{2} \odot C^{(l-1)}(A, B) + K^{l}\left(A, B \middle| \Theta^{l}\right).$$

[2] Le Gratiet, L., Multi-fidelity Gaussian process regression for computer experiments, Ph.D. Thesis, Université Paris-Diderot - Paris VII, 2013. Français.

Multi-fidelity surrogates: Increasingly Including Recursive Strategy

We propose an increasingly including strategy in which all the *(l-1)* models are included at once.

That is, we seek an extension of [2] where the lower predictor consists in a linear combination of all previous levels, with coefficients $\rho^{l}_{l'<l}$
For convenience we write

$$\mu^{<l}(x|\rho^l) \equiv \sum_{l'<l} \rho^{l}_{l'} \mu^{l'}(x), \quad C^{<l}(x,x'|\rho^l) \equiv \sum_{l'<l} (\rho^{l}_{l'})^2 \odot C^{l'}(x,x').$$

With these notations, ρ^l becomes a vector and the expression for the mean and the covariance become:

$$\mu^l(x) = \mu^{<l}(x|\rho^l) + \mathcal{K}^l(x,X^l|\rho^l,\Theta^l)\left[\mathcal{K}^l(X^l,X^l|\rho^l,\Theta^l) + \sigma_l^2 \underline{\mathrm{I}}\right]^{-1}\left(Y^l - \mu^{<l}(x|\rho^l)\right),$$

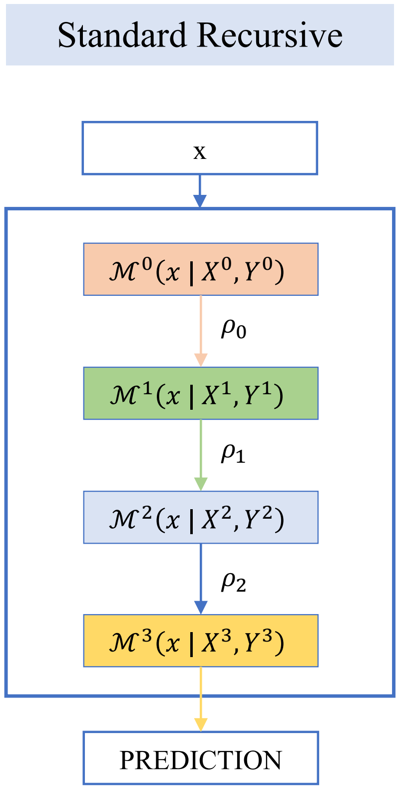
$$C^l(x,x') = \mathcal{K}^l(x,x'|\rho_{(l-1)},\Theta^l) - \left[\mathcal{K}^l(x,X^l|\rho^l,\Theta^l)\right]\left[\mathcal{K}^l(X^l,X^l|\rho^l,\Theta^l) + \sigma_l^2 \underline{\mathrm{I}}\right]^{-1}\left[\mathcal{K}^l(X^l,x'|\rho^l,\Theta^l)\right],$$

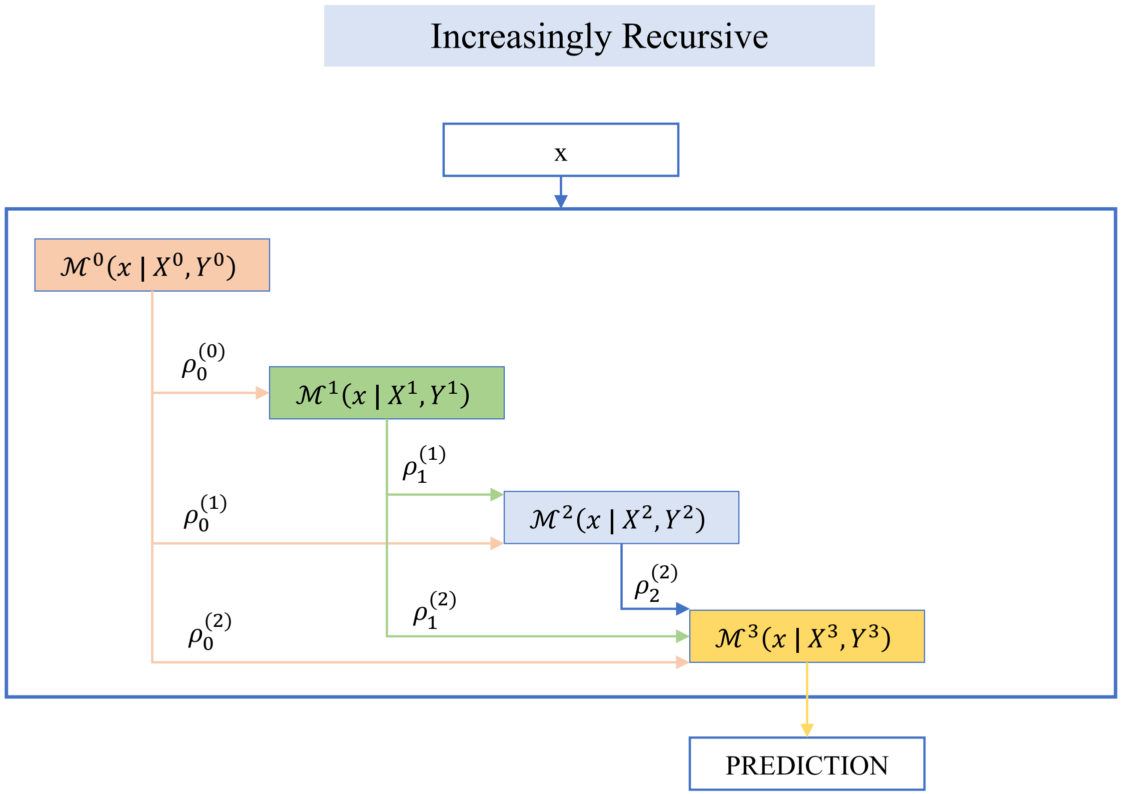
where $\mathcal{K}$

$$\mathcal{K}^l(A,B|\rho^l,\Theta^l) = C^{<l}(A,B|\rho^l) + K^l(A,B|\Theta^l).$$

[2] Le Gratiet, L., Multi-fidelity Gaussian process regression for computer experiments, Ph.D. Thesis, Université Paris-Diderot - Paris VII, 2013. Français.

Multi-fidelity surrogates: graphical comparison





Training multi-fidelity surrogates

At each level, the multi-fidelity regression problem requires the estimation of the ρ^l (or $\rho_{(l-1)}$ in the standard formulation), Θ^l and σ_l^2 parameters.

Thanks to the recursive formulation, the estimation of these parameters can be **done sequentially, from the lowest to the highest level**.

We apply a **Maximum Likelihood method** (ML): we aim at maximizing the negative marginal log-likelihood

$$\mathcal{L}\left(Y^l \middle| X^l, \rho^l, \Theta^l, \sigma_l\right) = -\frac{1}{2}\left(Y^l - \mu^{<l}\left(X^l \middle| \rho^l\right)\right)\left[\mathcal{K}^l\left(X^l, X^l \middle| \rho^l, \Theta^l\right) + \sigma_l^2 \underline{\mathrm{I}}\right]^{-1}\left(Y^l - \mu^{<l}\left(X^l \middle| \rho^l\right)\right) - \frac{1}{2}\log\left|\mathcal{K}^l\left(X^l, X^l \middle| \rho^l, \Theta^l\right) + \sigma_l^2 \underline{\mathrm{I}}\right|$$

Being $\mathcal{K}^l\left(A, B \middle| \rho^l, \Theta^l\right) = C^{<l}\left(A, B \middle| \rho^l\right) + K^l\left(A, B \middle| \Theta^l\right)$

Other approaches are possible e.g., Leave-One-Out (LOO) cross validation, **Monte-Carlo Markov-Chain** (MCMC)

Evaluating the performances

We will be comparing different approaches, the Single-Fidelity (**SF**), the Standard Recursive strategy from Le Gratiet (**SR**) and our Increasingly Recursive approach (**IR**)

For a fair comparison, we randomize the training data by artificially creating *K*=100 training sets for each level.

Surrogates models are trained *K* times and performances are compared in averaged terms

$$\text{score}_{\text{AVG}} = \frac{1}{K}\sum_k \text{score}_k \quad \text{with} \quad \text{score}_k \left(1 - \frac{\int_\Omega \left(T(x) - \mu_k^L(x)\right)}{\int_\Omega \left(T(x) - \frac{1}{\Omega}\int_\Omega T(x)\right)}\right),$$

$$\text{score}_{\text{STD}} = \frac{1}{K}\sum_k \left(\text{score}_k - \text{score}_{\text{AVG}}\right),$$

The score performance is included in (-∞, 1]

Test A

We assume a 1D mapping $T: \Omega \subset \mathbb{R}^1 \longmapsto \mathbb{R}$ to be the true model underlying the reality of interest i.e., the target model we want to approximate $\mathcal{M}^L(x) \simeq T(x)$

$$T(x) = x\sin(8\pi x) + x, \qquad \text{with } x \in [0.0, 1.0]$$

We then assume that four models of different fidelity are at our disposal to approximate $T(x)$

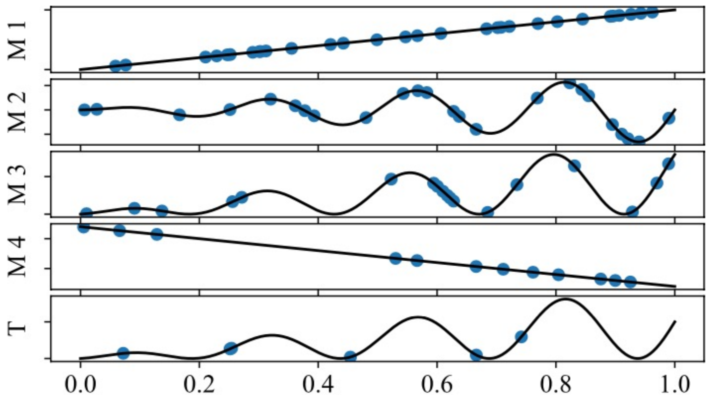
$$\begin{aligned} M1(x) &= x, \\ M2(x) &= 0.7x\sin(8\pi x), \\ M3(x) &= x\sin(8.2\pi x) + x, \\ M4(x) &= -5x + 1, \end{aligned}$$

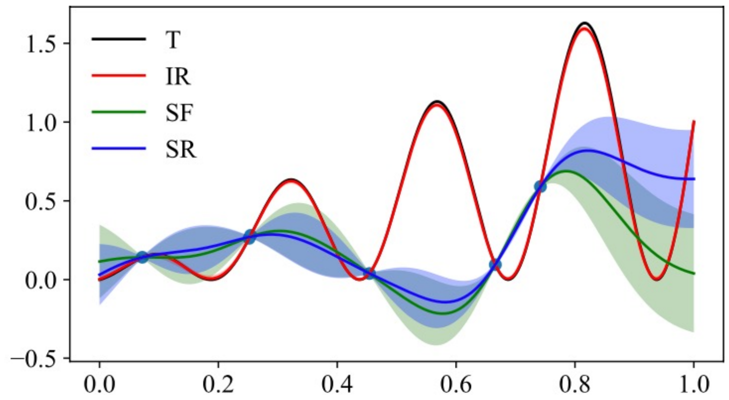
The following equivalence holds $T = M1 + 1.429M2$

We consider two arbitrarily ordered sequences $S_A = \{M1, M2, M3, M4, T\}$ and $S_B = \{M1, M2, M4, M3, T\}$

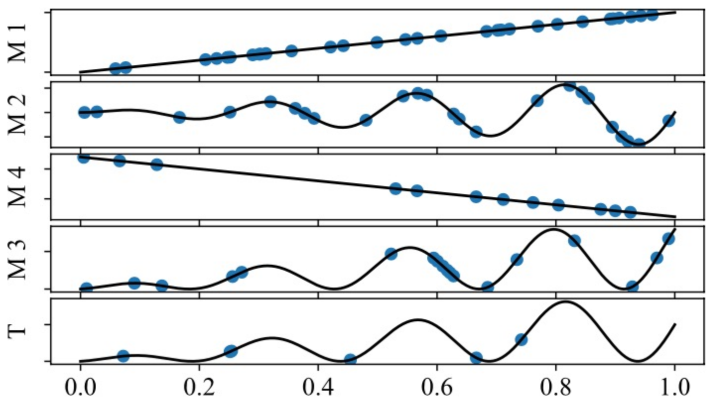
Test A

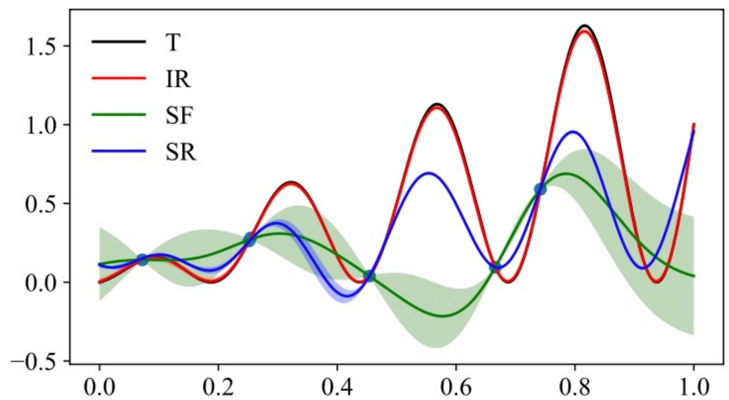
S_A


M 1
M 2
M 3
M 4
T
0.0
0.2
0.4
0.6
0.8
1.0


T
IR
SF
SR
1.5
1.0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0

S_B


M 1
M 2
M 4
M 3
T
0.0
0.2
0.4
0.6
0.8
1.0


T
IR
SF
SR
1.5
1.0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0

Test A

		$\rho_{AVG} \pm \rho_{STD}$				$\text{score}_{AVG} \pm \text{score}_{STD}$
		M1	M2	M3	M4	
S_A	IR	0.875 ± 0.219	1.278 ± 0.309	0.120 ± 0.250	-0.004 ± 0.036	0.959 ± 0.123
	SR	-	-	-	-0.246 ± 0.138	-0.355 ± 0.665
S_B	IR	0.884 ± 0.186	1.292 ± 0.282	0.110 ± 0.234	0.004 ± 0.035	0.961 ± 0.122
	SR	-	-	0.914 ± 0.293	-	0.608 ± 0.371
	SF	-	-	-	-	-0.264 ± 0.481

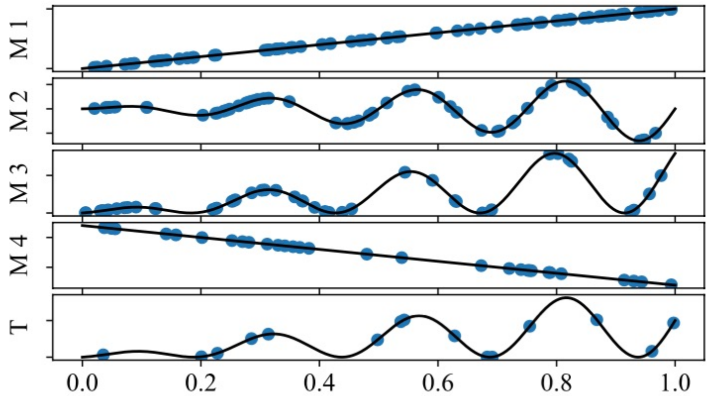
Recall that

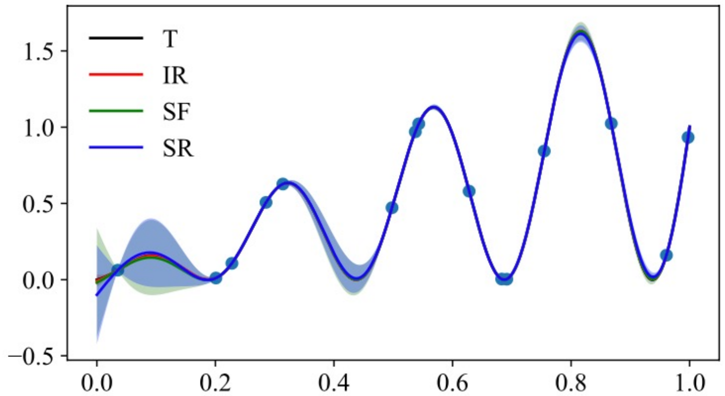
$$T = M1 + 1.424 M2$$

$$M1(x) = x,$$
$$M2(x) = 0.7x \sin(8\pi x),$$
$$M3(x) = x \sin(8.2\pi x) + x,$$
$$M4(x) = -5x + 1,$$
$$T(x) = x \sin(8\pi x) + x,$$

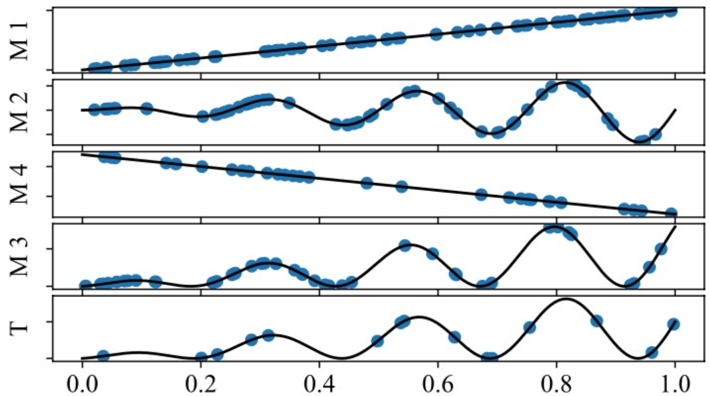
Test A

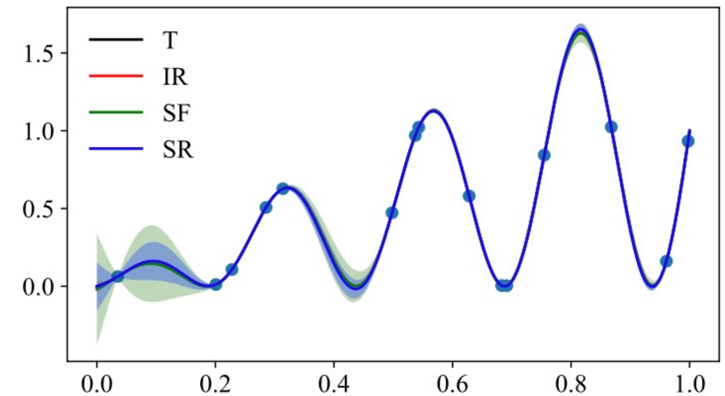
S_A


M 1
M 2
M 3
M 4
T
0.0
0.2
0.4
0.6
0.8
1.0


T
IR
SF
SR
1.5
1.0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0

S_B


M 1
M 2
M 4
M 3
T
0.0
0.2
0.4
0.6
0.8
1.0


T
IR
SF
SR
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Test A

		$\rho_{AVG} \pm \rho_{STD}$ M1	M2	M3	M4	$score_{AVG} \pm score_{STD}$
S_A	IR	0.968 ± 0.008	1.426 ± 0.004	0.002 ± 0.003	-0.006 ± 0.002	0.999 ± 0.000
	SR	-	-	-	-0.272 ± 0.111	0.730 ± 0.515
S_B	IR	0.968 ± 0.008	1.425 ± 0.004	0.002 ± 0.003	-0.006 ± 0.002	0.999 ± 0.000
	SR	-	-	0.846 ± 0.058	-	0.928 ± 0.097
	SF	-	-	-	-	0.803 ± 0.243

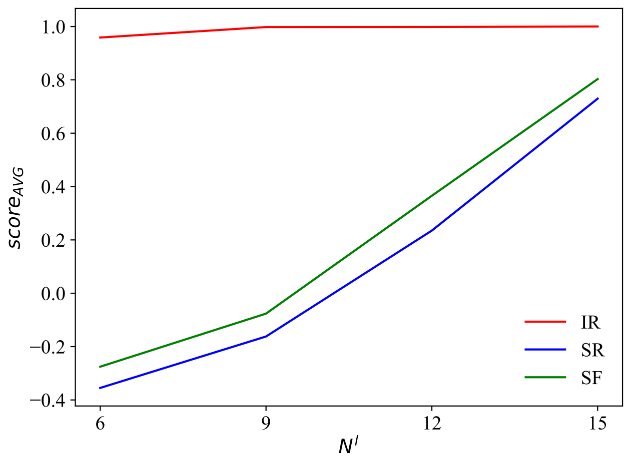
Recall that

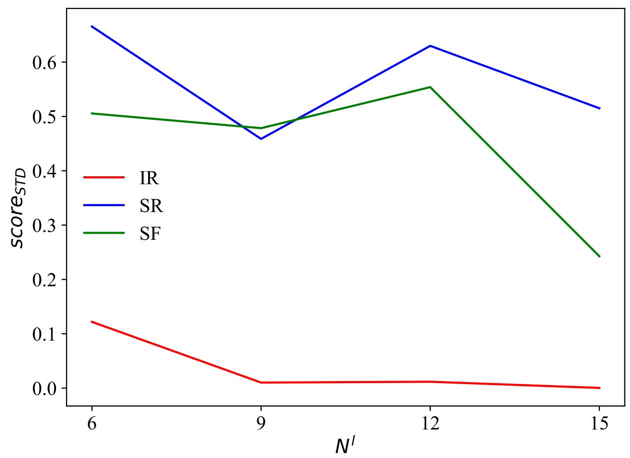
$$T = M1 + 1.424M2$$

$$M1(x) = x,$$
$$M2(x) = 0.7x\sin(8\pi x),$$
$$M3(x) = x\sin(8.2\pi x) + x,$$
$$M4(x) = -5x + 1,$$
$$T(x) = x\sin(8\pi x) + x,$$

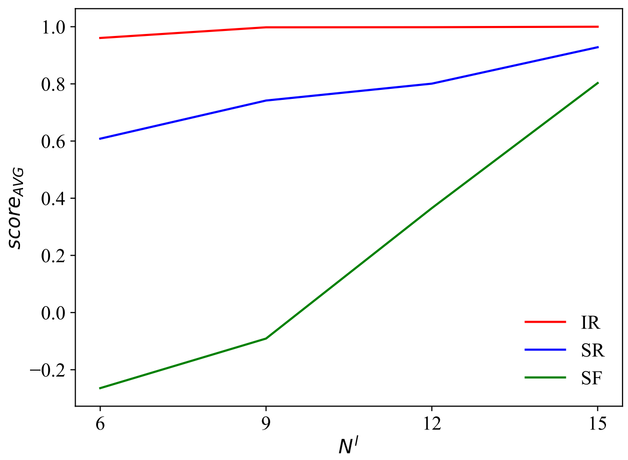
Test A

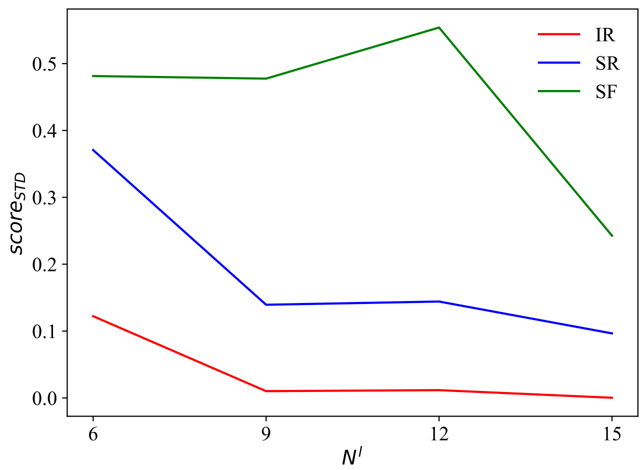
S_A





S_B





Computational effort for training 100 surrogates

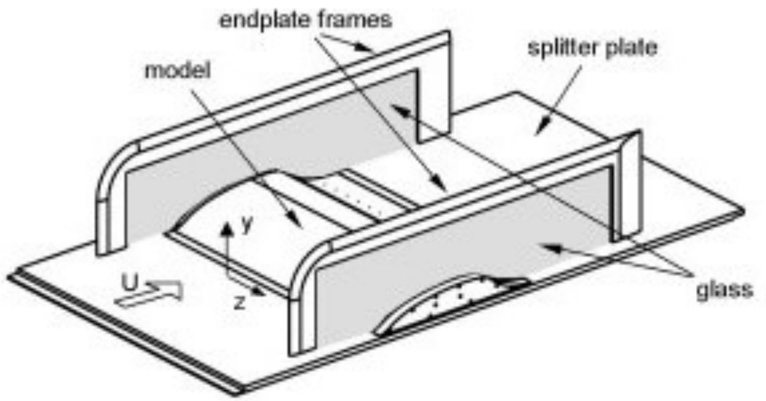
	$\mathfrak{c}(Y_T)$			
	6	9	12	15
IR	3148 [s]	3084 [s]	3388 [s]	3694 [s]
SR	1898 [s]	1948 [s]	2350 [s]	2623 [s]
SF	147 [s]	160 [s]	163 [s]	166 [s]

Test B

Ercoftac Classic Collection Database:

Wall-Mounted 2-D Hump with Oscillatory Zero-Mass-Flux Jet or Suction through a Slot
By Greenblatt Paschal, Yao, Harris, Schaeffler and Washburn

Chosen because of a large experimental database (including both CFD simulations and experiments)



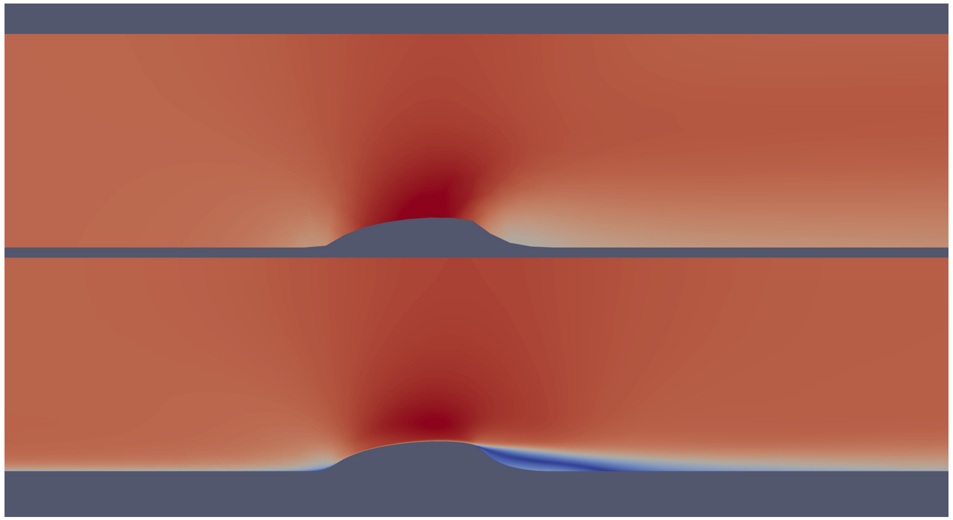
Test B

Ercoftac Classic Collection Database:

Wall-Mounted 2-D Hump with Oscillatory Zero-Mass-Flux Jet or Suction through a Slot
By Greenblatt Paschal, Yao, Harris, Schaeffler and Washburn

The experiment is our reality of interest

Euler model
M1 and M2 with increasing grid resolution
(37k and 83k elements)

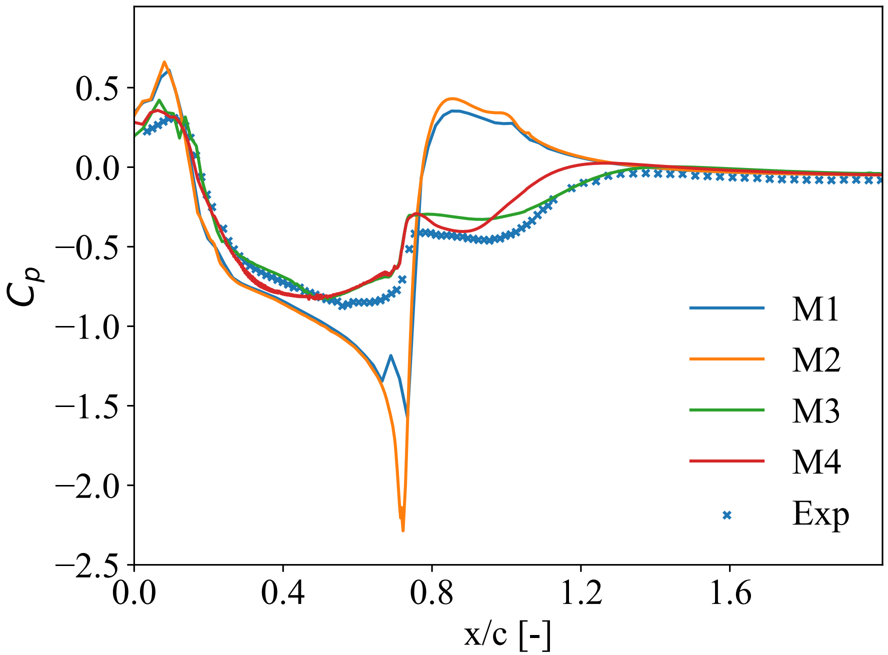


Reynolds-Averaged Navier-Stokes model
M3 and M4 with increasing grid resolution
(100k and 132k elements)

Test B

Ercoftac Classic Collection Database:

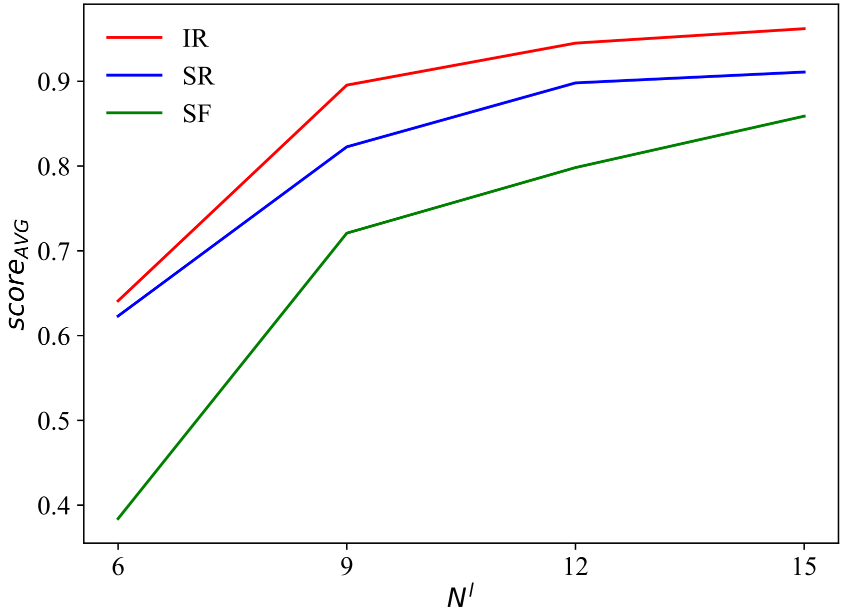
Wall-Mounted 2-D Hump with Oscillatory Zero-Mass-Flux Jet or Suction through a Slot
By Greenblatt Paschal, Yao, Harris, Schaeffler and Washburn

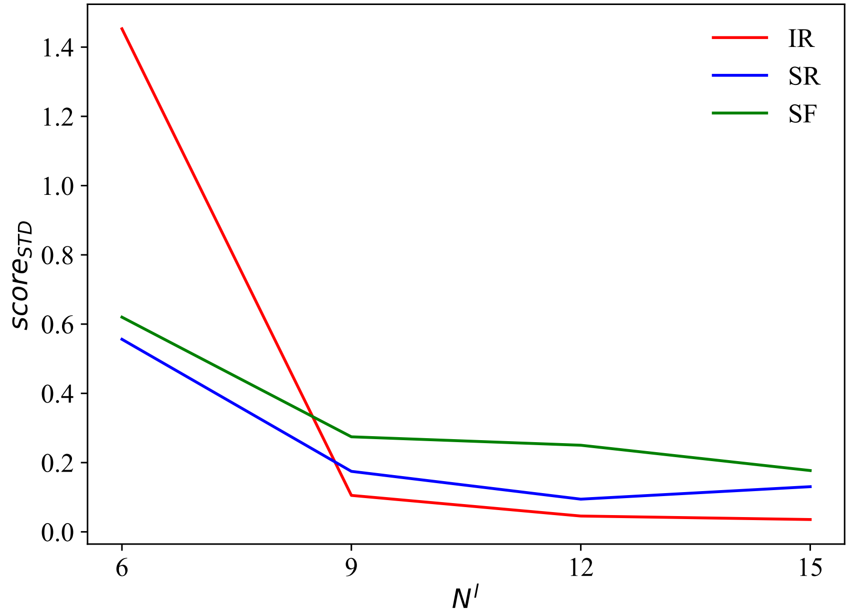


Test B

Ercoftac Classic Collection Database:

Wall-Mounted 2-D Hump with Oscillatory Zero-Mass-Flux Jet or Suction through a Slot
By Greenblatt Paschal, Yao, Harris, Schaeffler and Washburn





Part I: conclusions

Achievements:

- We exposed a weakness of the current multi-fidelity recursive sequential approach
- We proposed a solution at an acceptable computational cost
- Preliminary experiments show that the IR approach has superior averaged performances (even for a multi-dimensional input space)
- The IR approach can be exploited to obtain physics insights about the reality of interest

Next steps:

- Restoring the space varying weighting of the low levels predictions $g^T(x)$
- Investigate relevance of the multi-level sequence ordering in a more detailed manner

Part II:
future perspectives

Bayesian analysis for physics inference

We assume a 1D mapping $T\colon \Omega \subset \mathbb{R}^1 \longmapsto \mathbb{R}$ to be the true model underlying the reality of interest i.e., the target model we want to approximate $\mathcal{M}^L(x) \simeq T(x)$

$$T(x) = x\sin(8\pi x) + x, \qquad \text{with } x \in [0.0, 1.0]$$

We then assume that four models of different fidelity are at our disposal to approximate $T(x)$

$$M1(x) = x,$$
$$M2(x) = 0.7x\sin(8\pi x),$$
$$M3(x) = x\sin(8.2\pi x) + x,$$
$$M4(x) = -5x + 1,$$

We carry out the training of the surrogate using a Bayesian inference approach (Monte-Carlo Markov-Chain, MCMC), to obtain a probabilistic characterization of the regression parameters

Bayesian analysis for physics inference

We rely on a classical Bayesian approach for inferring the unknown parameters (ρ^l, Θ^l and σ_l^2)

$$\mathcal{P}\left(\rho^l, \Theta^l, \sigma_l \middle| Y^l, X^l\right) \propto \mathcal{P}\left(Y^l \middle| \rho^l, \Theta^l, \sigma_l, X^l\right) \mathcal{P}\left(\rho^l, \Theta^l, \sigma_l \middle| X^l\right)$$

The likelihood was defined earlier and it reads

$$\mathcal{P}\left(Y^l \middle| \rho^l, \Theta^l, \sigma_l, X^l\right) = -\frac{1}{2}\left(Y^l - \mu^{<l}\left(X^l \middle| \rho^l\right)\right)\left[\mathcal{K}^l\left(X^l, X^l \middle| \rho^l, \Theta^l\right) + \sigma_l^2 \underline{\mathrm{I}}\right]^{-1}\left(Y^l - \mu^{<l}\left(X^l \middle| \rho^l\right)\right) - \frac{1}{2}\log\left|\mathcal{K}^l\left(X^l, X^l \middle| \rho^l, \Theta^l\right) + \sigma_l^2 \underline{\mathrm{I}}\right|$$

Conveniently, we assume a uniform prior distributions $\mathcal{P}\left(\rho^l, \Theta^l, \sigma_l\right) \sim \mathcal{U}$ (min, max)

Bayesian analysis for physics inference

We rely on a classical Bayesian approach for inferring the unknown parameters (ρ^l, Θ^l and σ_l^2)

$$\mathcal{P}\left(\rho^l, \Theta^l, \sigma_l \middle| Y^l, X^l\right) \ \propto \ \mathcal{P}\left(Y^l \middle| \rho^l, \Theta^l, \sigma_l, X^l\right) \mathcal{P}\left(\rho^l, \Theta^l, \sigma_l \middle| X^l\right)$$

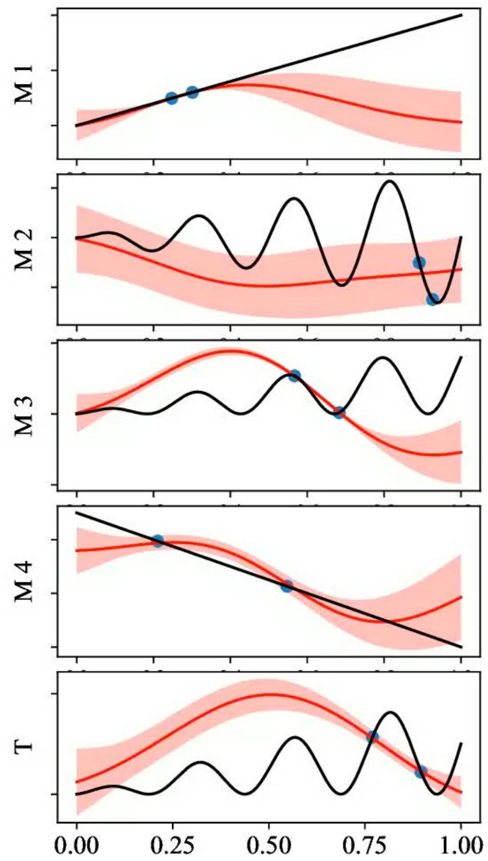
The likelihood was defined earlier and it reads

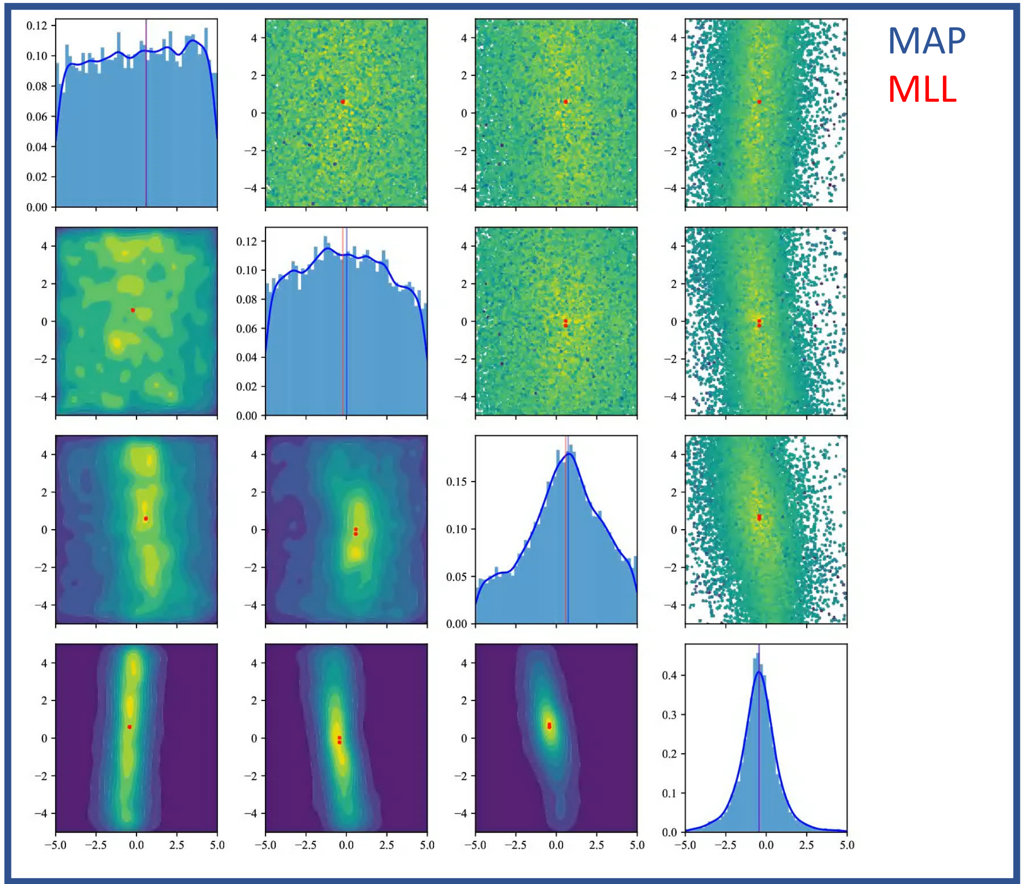
$$\mathcal{P}\left(Y^l \middle| \rho^l, \Theta^l, \sigma_l, X^l\right) = -\frac{1}{2}\left(Y^l - \mu^{<l}\left(X^l \middle| \rho^l\right)\right) \left[\mathcal{K}^l\left(X^l, X^l \middle| \rho^l, \Theta^l\right) + \ \sigma_l^2 \underline{\mathrm{I}}\right]^{-1} \left(Y^l - \mu^{<l}\left(X^l \middle| \rho^l\right)\right) - \frac{1}{2}\log\left|\mathcal{K}^l\left(X^l, X^l \middle| \rho^l, \Theta^l\right) + \ \sigma_l^2 \underline{\mathrm{I}}\right|$$

Conveniently, we assume a uniform prior distributions $\mathcal{P}\left(\rho^l, \Theta^l, \sigma_l\right) \sim \mathcal{U}$ (min, max)

Bayesian methods allows for introducing some scientific knowledge into the training process e.g., employ physics-informed priors.
Theory Guided Data Science (TGDS)

Bayesian analysis for physics inference





Bayesian optimization

We target the lift coefficient of a morphing airfoil and no constraint applies, therefore we seek for

$$\boldsymbol{x} = \underset{\boldsymbol{x}\in\Omega}{\operatorname{argmax}}\, C_L(\boldsymbol{x})$$

The 4 design parameter: max camber (x_1), position of the max camber point (x_2), maximum thickness (x_3) and the AoA (x_4)

- Maximum computational budget: 100 full computational model evaluations
- No early convergence criteria applied
- The low fidelity data base includes 720 points evaluated using the potential flow solver from XFOIL (just few minutes required for building the database on a single core desktop machine)
- The mid fidelity database (RANS model, coarse mesh 3.5k) including 30 data points
- One single high fidelity RANS evaluation (57k elements) requires 10 to 40 min on 6 cores

Bayesian optimization

We target the lift coefficient of a morphing airfoil and no constraint applies, therefore we seek for

$$\boldsymbol{x} = \underset{\boldsymbol{x}\in\Omega}{\operatorname{argmax}}\, C_L(\boldsymbol{x})$$

The 4 design parameter: max camber (x_1), position of the max camber point (x_2), maximum thickness (x_3) and the AoA (x_4)

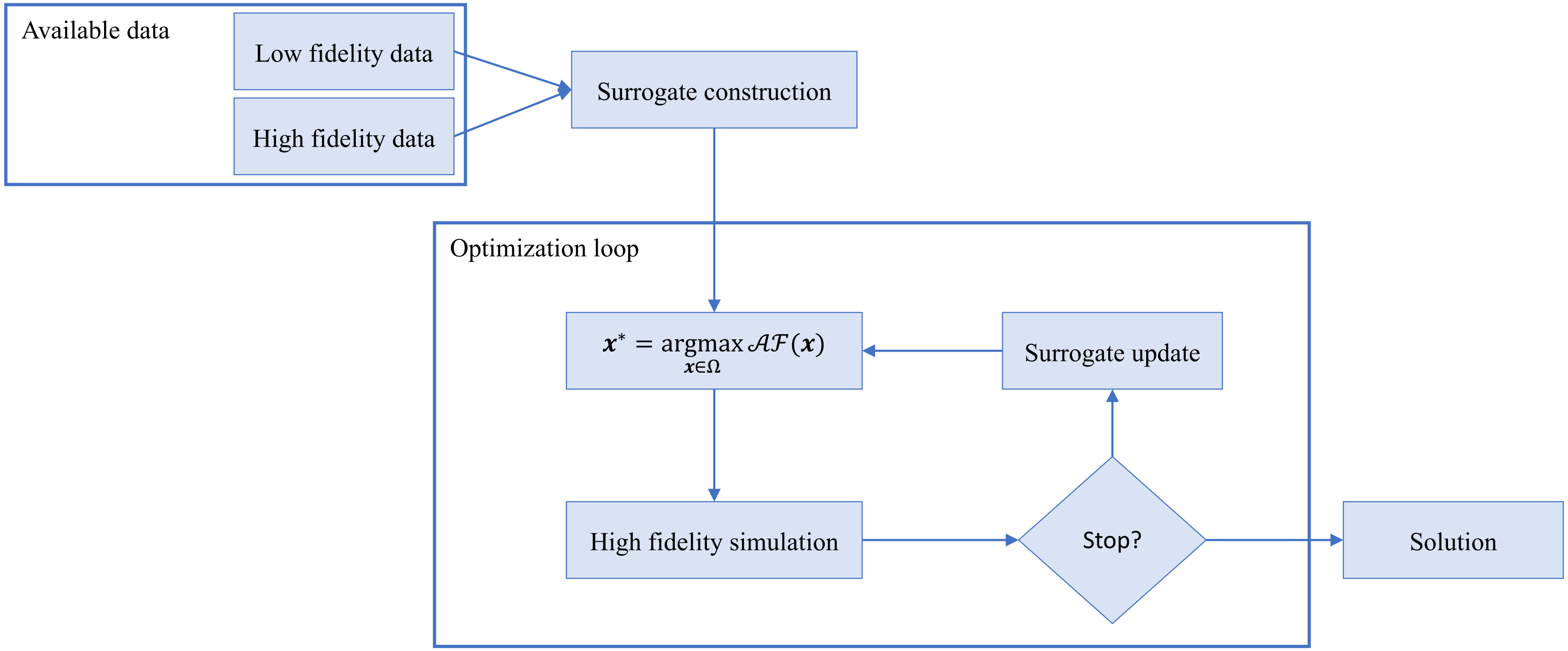
- Maximum computational budget: 100 full computational model evaluations
- No early convergence criteria applied
- The low fidelity data base includes 720 points evaluated using the potential flow solver from XFOIL (just few minutes required for building the database on a single core desktop machine)
- The mid fidelity database (RANS model, coarse mesh 3.5k) including 30 data points
- One single high fidelity RANS evaluation (57k elements) requires 10 to 40 min on 6 cores

We implement a sequential approach where a GP surrogate is built to approximate the (**unknown**) objective function $C_L(\boldsymbol{x})$

GPs provide a probabilistic characterization of the prediction: allow exploitation/exploration by means of an *acquisition function* $\mathcal{AF}(\boldsymbol{x})$

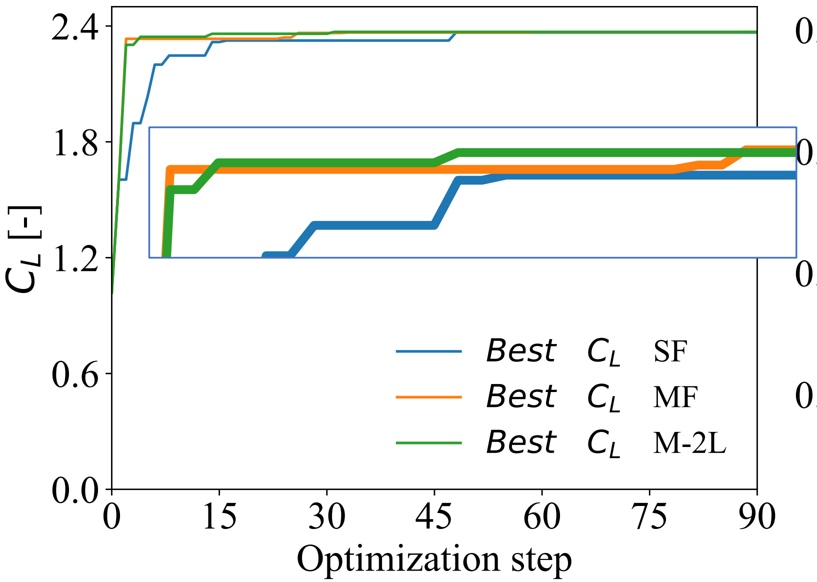
Bayesian optimization

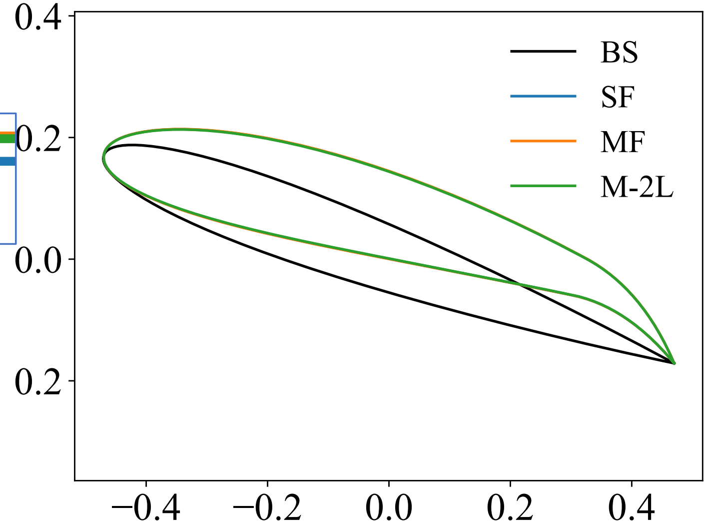
Efficient Global Optimization (EGO): Bayesian methods take advantage of prior knowledge and data as they became available during the design space exploration/exploitation



Bayesian optimization

Optimal designs comparison





	SF	MF	MF-2
x_1	8.0	8.0	8.0
x_2	8.0	8.0	8.0
x_3	1.5336	1.5395	1.5520
x_4	19.2730	19.5553	19.5925
C_L	2.3691	2.3663	2.3690

Infilling strategies for enriching a database

An acquisition function can be defined to enrich the database according to arbitrary criteria

Since we are targeting an improved knowledge of reality, we aim at reducing prediction variance at level L

$$x_* = \underset{x_* \in \Omega}{\operatorname{argmax}} \, \sigma_L^2(x)$$

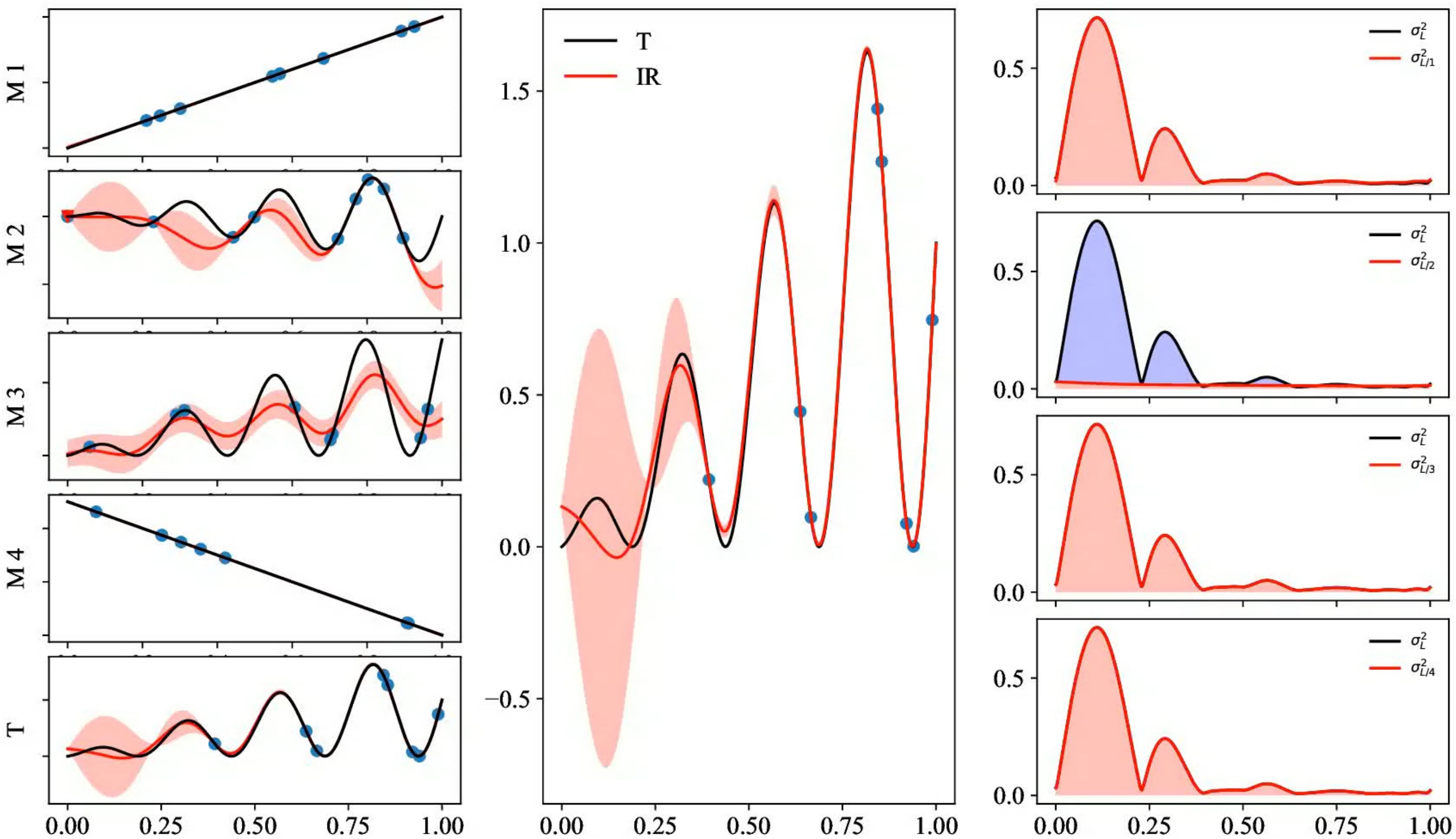
Once x* is defined, we can select the level to sample according to the maximum of

$$\Delta^i_{\sigma_L^2(x)} = \sigma_L^2(x) - \sigma_{L/i}^2(x) \qquad \text{with } i = 1, \ldots, L-1.$$

$\Delta^i_{\sigma_L^2(x)}$ is the *i*-th level contribution to $\sigma_L^2(x)$.

$\sigma_{L/i}^2(x)$ is obtained by zeroing the *i*-th regression coefficient.

Infill strategies for enriching the database



Part II: conclusions

Perspectives:

- The IR approach can be exploited to obtain physics insights about the reality of interest. LOO-CV/MCMC-based methods are already available but they still needs a thorough assessment

- Multi-fidelity models bring clear advantages to design problems. We expect to demonstrate the further advantages of our agnostic formulation for a real application (aerodynamic optimization of the tail of a commercial aircraft).

- The proposed agnostic formulation opens the path to the development of novel strategies for efficiently building multi-fidelity databases.

Questions?

This project has received funding from the Clean Sky 2 Joint Undertaking (JU) under grant agreement No 101008257 (MONNALISA project, **www.monnalisa-project.eu**).

The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Clean Sky 2 JU members other than the Union. All the material reported in the present work reflects the view of the project partners only. The JU is not responsible for any use that may be made or the information it contains.