Deep Gaussian processes for the analysis and optimization of complex systems

Ali Hebbal [1,2]

Co-directors : E-G. Talbi[2], N. Melab[2]
Co-advisors : M. Balesdent[1], L. Brevault [1]

[1]ONERA - The French Aerospace Lab
[2]Université de Lille, CNRS/CRIStAL, Inria Lille

Ph.D Thesis Defense 21/01/2020

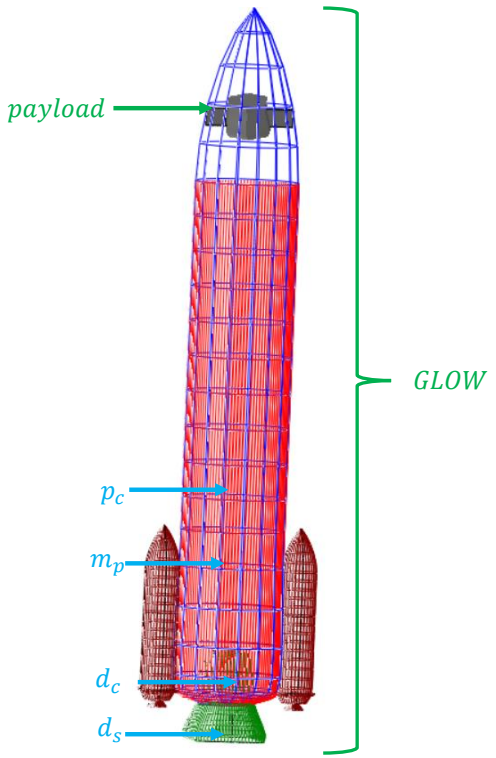
ONERA THE FRENCH AEROSPACE LAB

Inria INVENTEURS DU MONDE NUMÉRIQUE

Université de Lille 1 SCIENCES ET TECHNOLOGIES

Design of complex systems

Design of complex systems



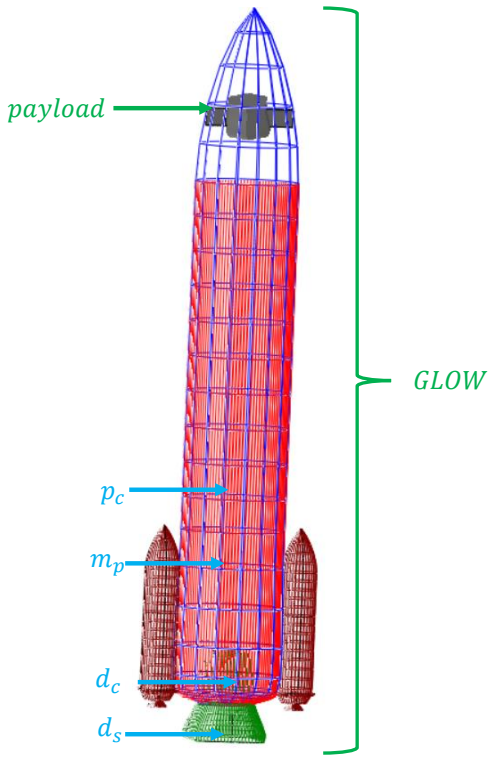
Design variables → **Physical model** → Response of interest

The design of complex systems :

Analysis and **optimization** approaches on a **physical model** in order to :

- Specify the design variables
  - propellant mass m_p,
  - combustion pressure p_c,
  - nozzle diameters d_s, d_c, *etc.*
- most adequate to the purpose of the design system
  - maximize payload mass,
  - minimize Gross Lift Off-Weight (GLOW), *etc.*
- while respecting the system feasibility
  - structural stresses,
  - injection constraint, *etc.*

Design of complex systems



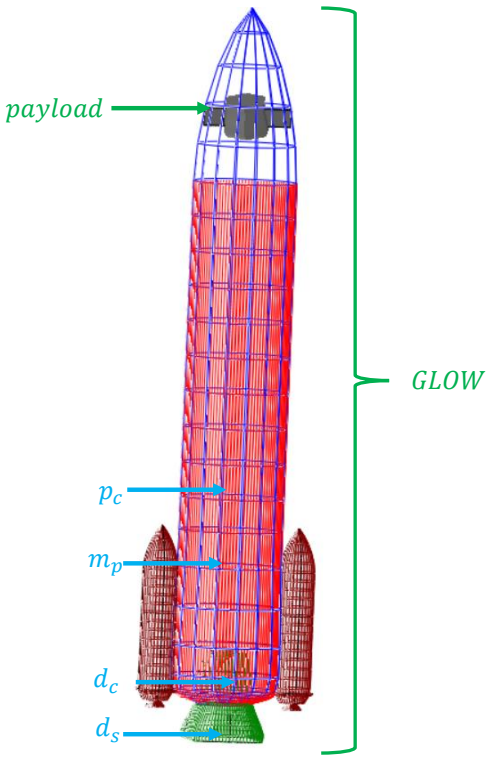
Design variables → **Physical model** → Response of interest

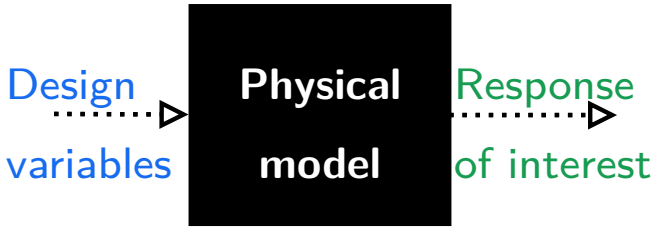
The design of complex systems :

Analysis and **optimization** approaches on a **physical model** in order to :

- Specify the design variables
  - propellant mass m_p,
  - combustion pressure p_c,
  - nozzle diameters d_s, d_c, *etc.*
- most adequate to the purpose of the design system
  - maximize payload mass,
  - minimize Gross Lift Off-Weight (GLOW), *etc.*
- while respecting the system feasibility
  - structural stresses,
  - injection constraint, *etc.*

Design of complex systems



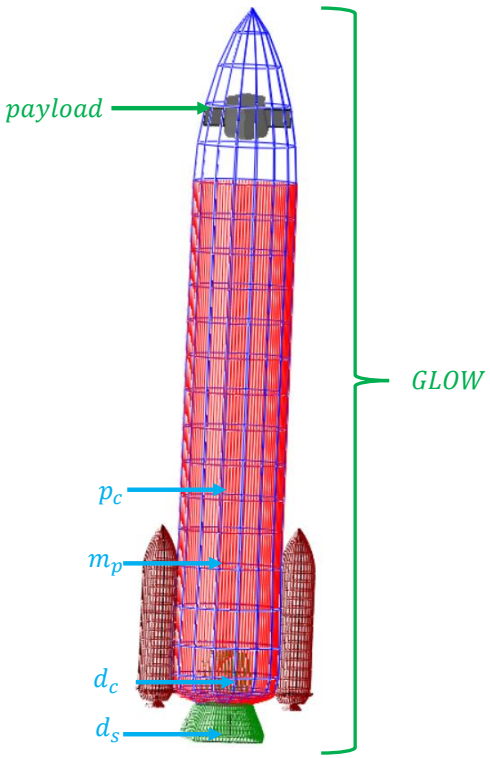


The design of complex systems :

Analysis and **optimization** approaches on a **physical model** in order to :

- Specify the design variables
  - propellant mass m_p,
  - combustion pressure p_c,
  - nozzle diameters d_s, d_c, *etc.*
- most adequate to the purpose of the design system
  - maximize payload mass,
  - minimize Gross Lift Off-Weight (GLOW), *etc.*
- while respecting the system feasibility
  - structural stresses,
  - injection constraint, *etc.*

Design of complex systems



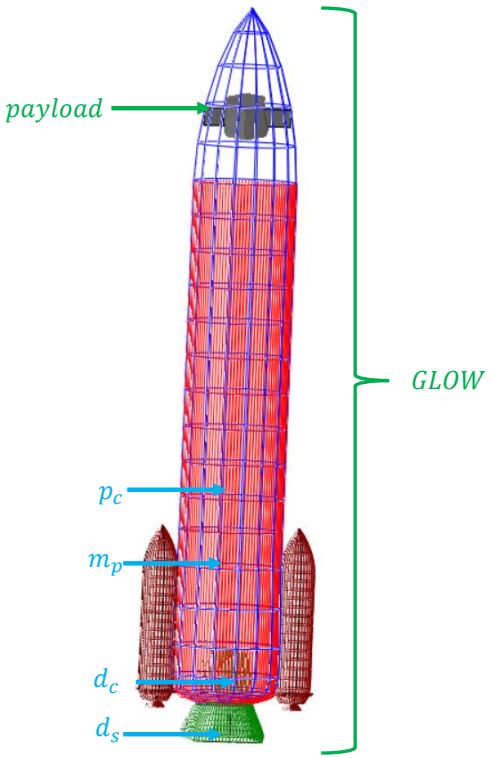
Design variables → **Physical model** → Response of interest

The design of complex systems :

Analysis and **optimization** approaches on a **physical model** in order to :

- Specify the design variables
  - propellant mass m_p,
  - combustion pressure p_c,
  - nozzle diameters d_s, d_c, *etc.*
- most adequate to the purpose of the design system
  - maximize payload mass,
  - minimize Gross Lift Off-Weight (GLOW), *etc.*
- while respecting the system feasibility
  - structural stresses,
  - injection constraint, *etc.*

Design of complex systems



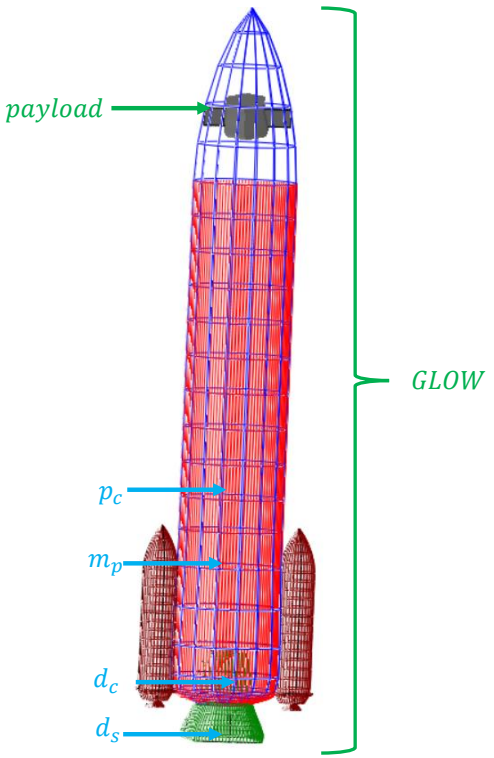
Design variables → **Physical model** → Response of interest

The design of complex systems :

Analysis and **optimization** approaches on a **physical model** in order to :

- Specify the design variables
  - propellant mass m_p,
  - combustion pressure p_c,
  - nozzle diameters d_s, d_c, *etc.*
- most adequate to the purpose of the design system
  - maximize payload mass,
  - minimize Gross Lift Off-Weight (GLOW), *etc.*
- while respecting the system feasibility
  - structural stresses,
  - injection constraint, *etc.*

Design of complex systems



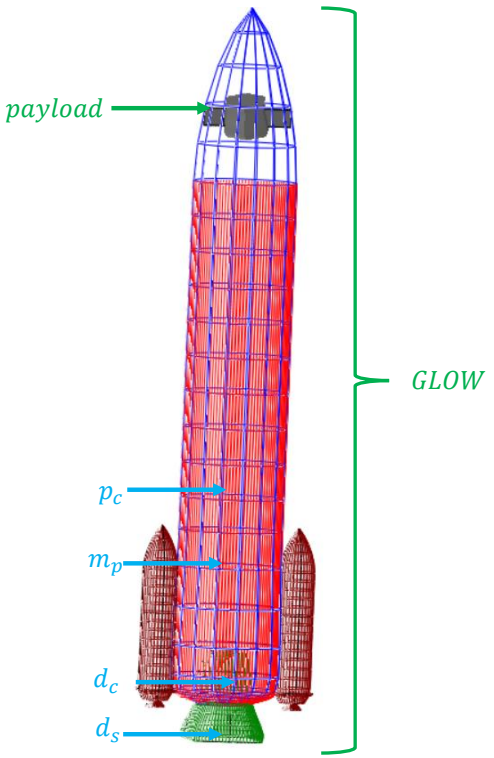
Design variables → **Physical model** → Response of interest

The design of complex systems :

Analysis and **optimization** approaches on a **physical model** in order to :

- Specify the design variables
  - propellant mass m_p,
  - combustion pressure p_c,
  - nozzle diameters d_s, d_c, *etc.*
- most adequate to the purpose of the design system
  - maximize payload mass,
  - minimize Gross Lift Off-Weight (GLOW), *etc.*
- while respecting the system feasibility
  - structural stresses,
  - injection constraint, *etc.*

Design of complex systems



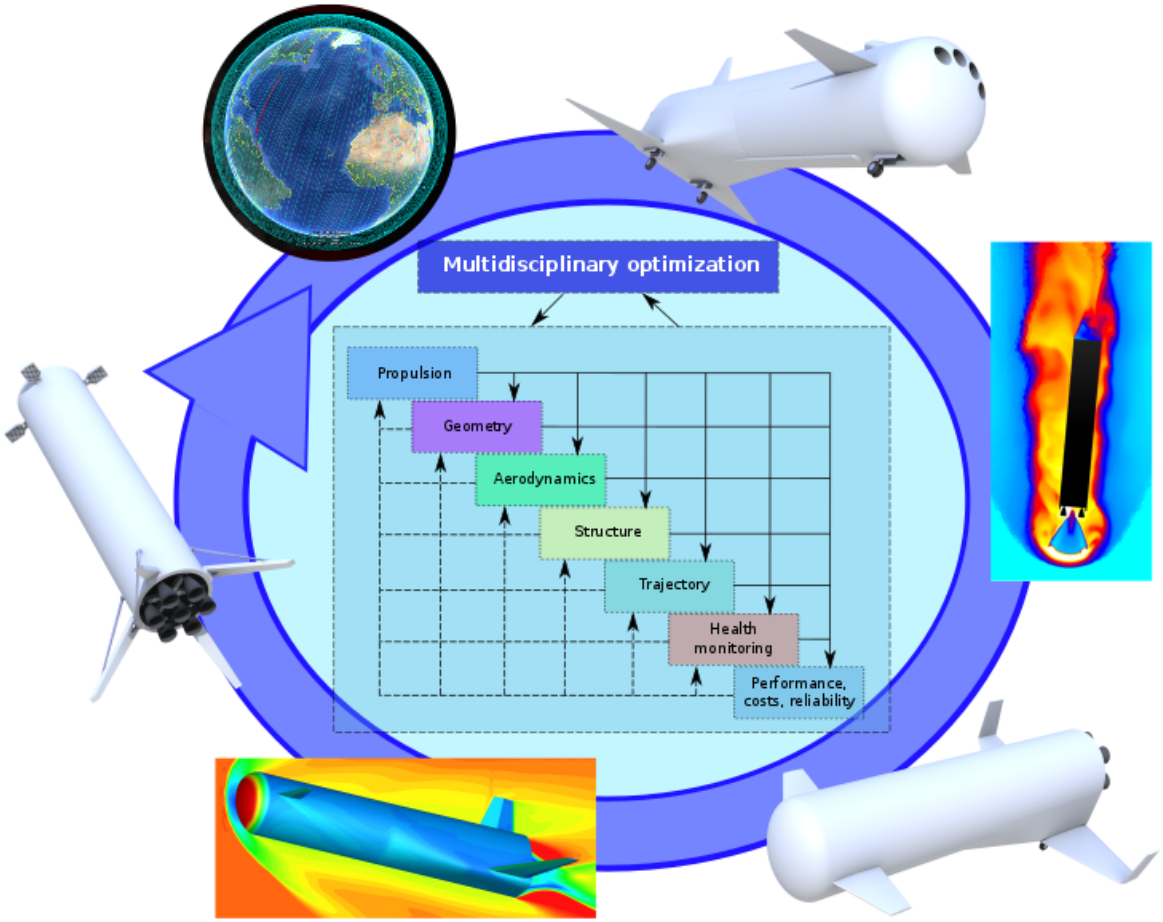
Design variables → **Physical model** → Response of interest

The design of complex systems :

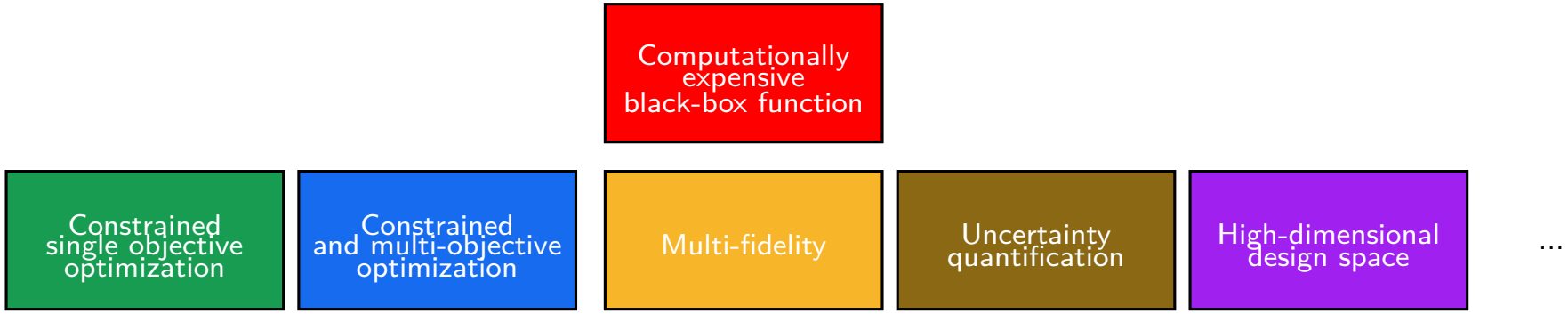
Analysis and **optimization** approaches on a **physical model** in order to :

- Specify the design variables
  - propellant mass m_p,
  - combustion pressure p_c,
  - nozzle diameters d_s, d_c, *etc.*
- most adequate to the purpose of the design system
  - maximize payload mass,
  - minimize Gross Lift Off-Weight (GLOW), *etc.*
- while respecting the system feasibility
  - structural stresses,
  - injection constraint, *etc.*

Design of complex systems



Challenges in the design of complex systems



Challenges in the design of complex systems

Computationally expensive black-box function

Constrained single objective optimization

Constrained and multi-objective optimization

Multi-fidelity

Uncertainty quantification

High-dimensional design space

...

✗ Exact optimization approaches

✗ Classic evolutionary algorithms

✗ Monte Carlo methods

Challenges in the design of complex systems

Computationally expensive black-box function

Constrained single objective optimization

Constrained and multi-objective optimization

Multi-fidelity

Uncertainty quantification

High-dimensional design space

...

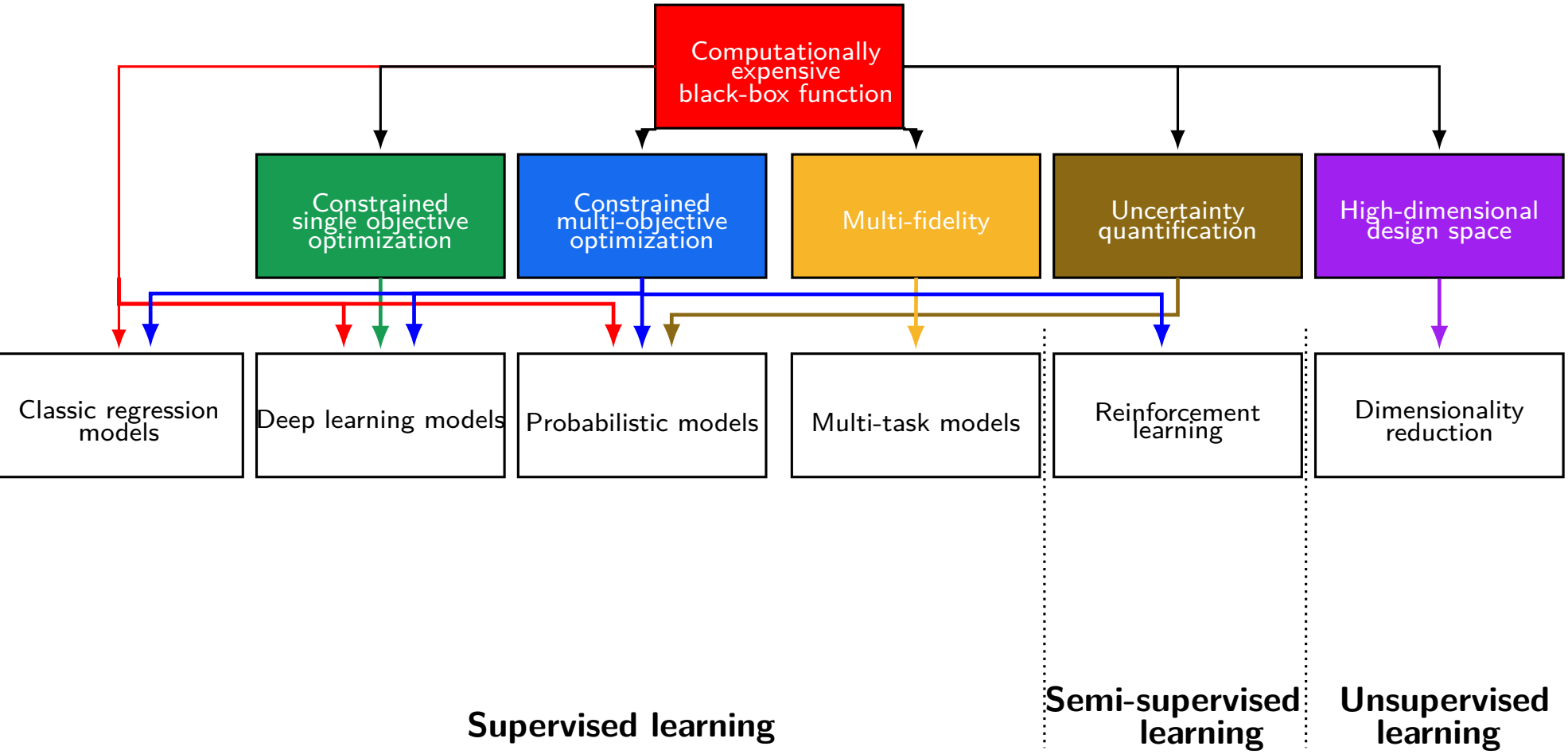
✗ Exact optimization approaches

✗ Classic evolutionary algorithms

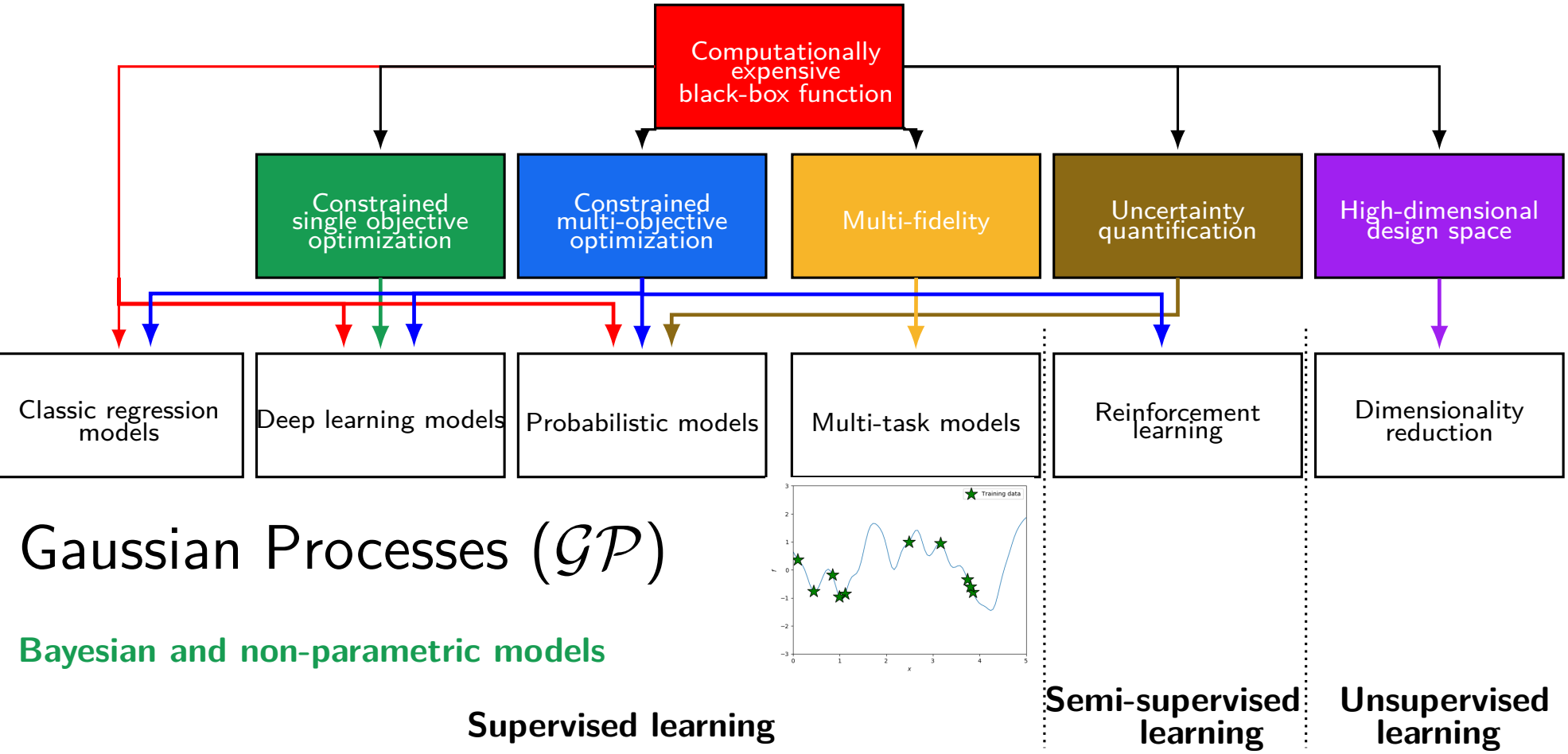
✗ Monte Carlo methods

✓ Machine learning methods

Design of complex systems with data-driven approaches



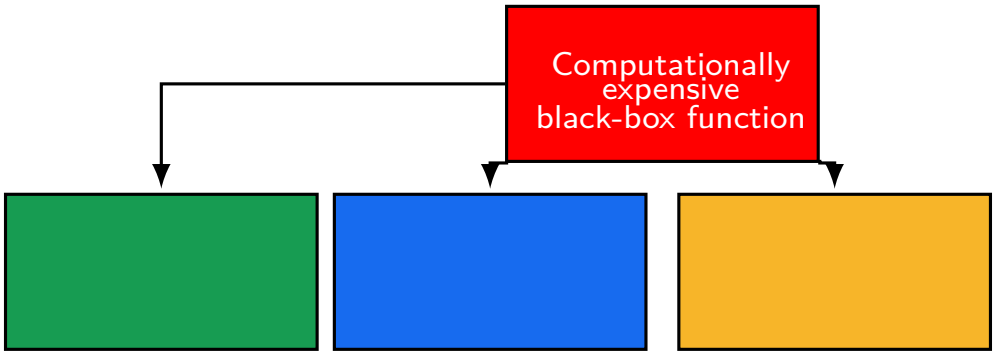
Design of complex systems with data-driven approaches



Gaussian Processes ($\mathcal{GP}$)

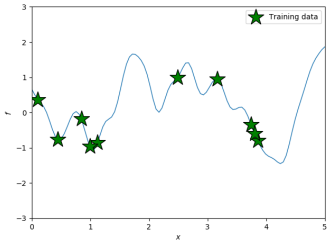
Bayesian and non-parametric models

Design of complex systems with data-driven approaches

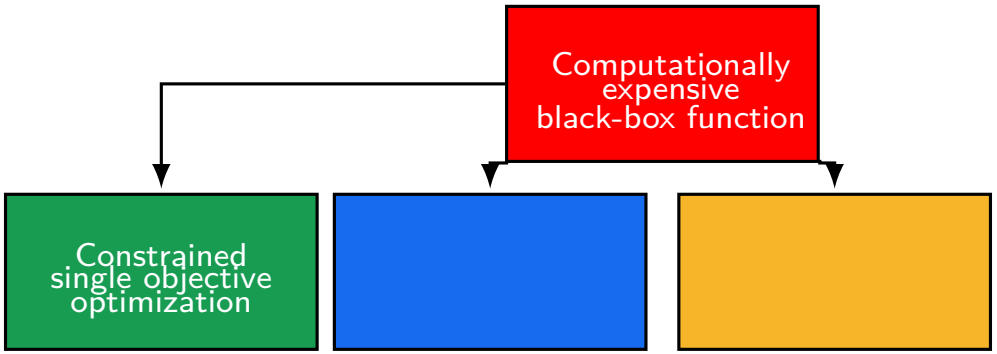


Gaussian Processes ($\mathcal{GP}$)

Bayesian and non-parametric models

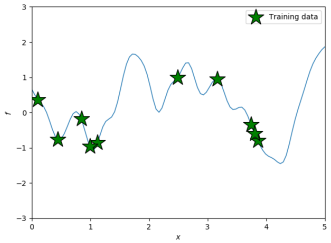


Design of complex systems with data-driven approaches

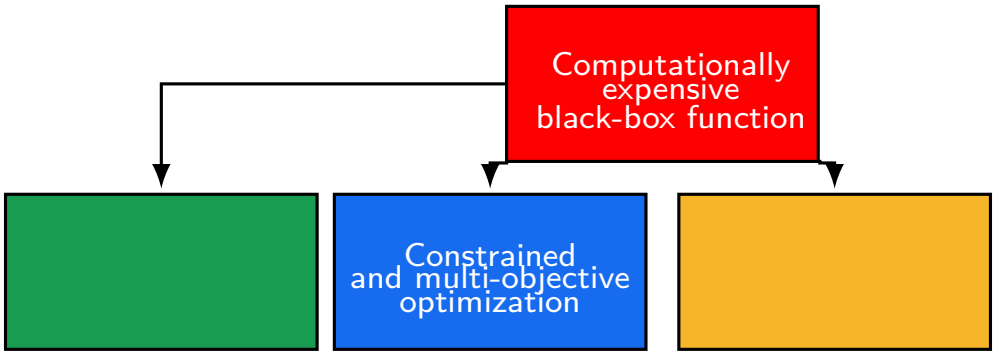


Gaussian Processes ($\mathcal{GP}$)

Bayesian and non-parametric models

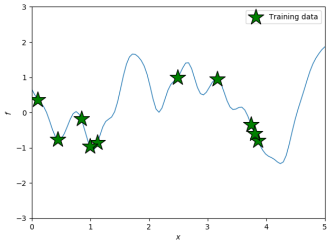


Design of complex systems with data-driven approaches

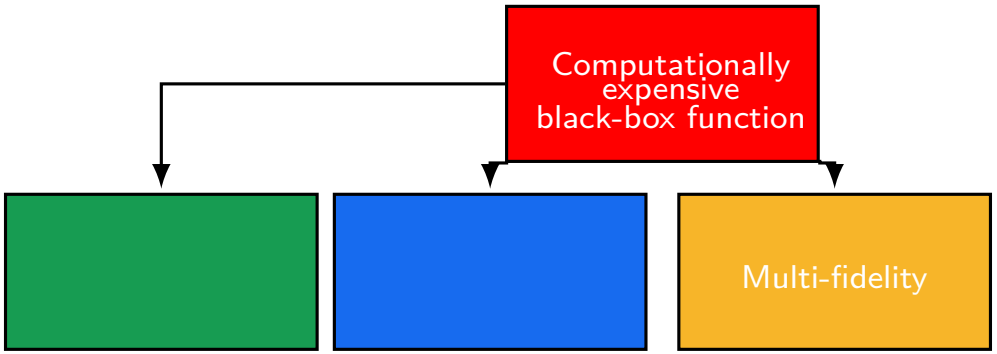


Gaussian Processes ($\mathcal{GP}$)

Bayesian and non-parametric models

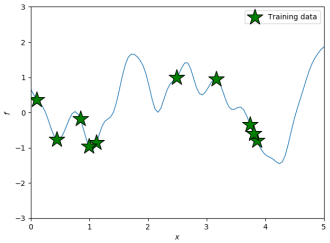


Design of complex systems with data-driven approaches



Gaussian Processes ($\mathcal{GP}$)

Bayesian and non-parametric models



Limits of Gaussian processes

Limits of Gaussian processes for the analysis and optimization of design problems :

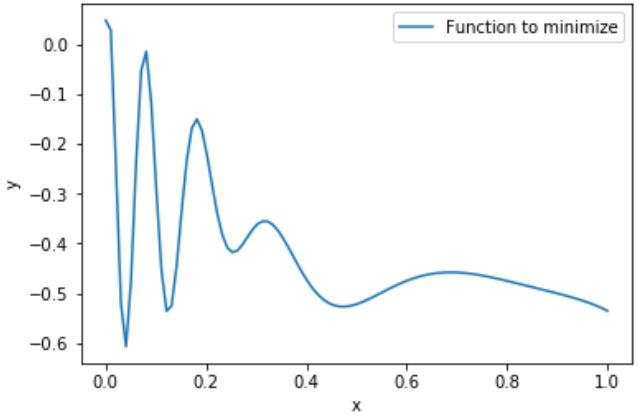
- Optimization of non-stationary functions,
- Multi-objective problems with correlated objectives,
- Multi-fidelity problems with varying input space dimensions.

Limits of Gaussian processes

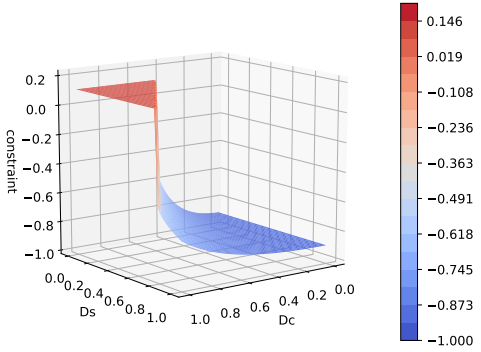
Limits of Gaussian processes for the analysis and optimization of design problems :

- Optimization of non-stationary functions,
- Multi-objective problems with correlated objectives,
- Multi-fidelity problems with varying input space dimensions.

Variation of the smoothness along the input space



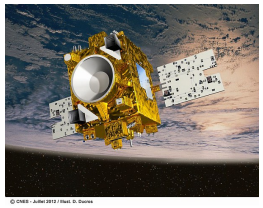
Abrupt change in the function



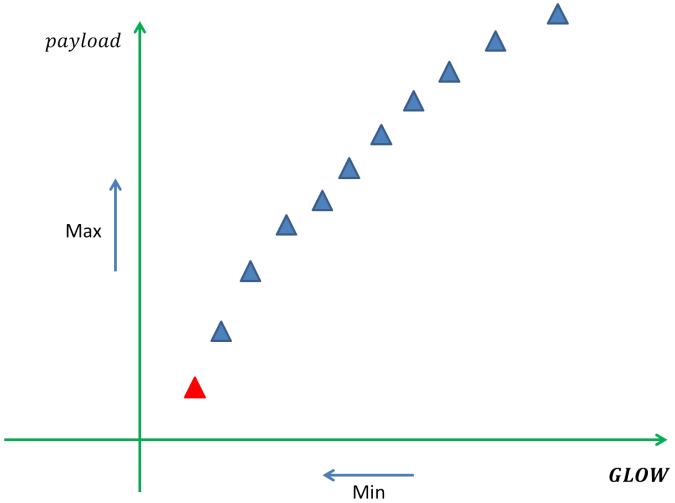
Limits of Gaussian processes

Limits of Gaussian processes for the analysis and optimization of design problems :

- Optimization of non-stationary functions,
- Multi-objective problems with correlated objectives,
- Multi-fidelity problems with varying input space dimensions.



Maximize f_2 : The payload mass.



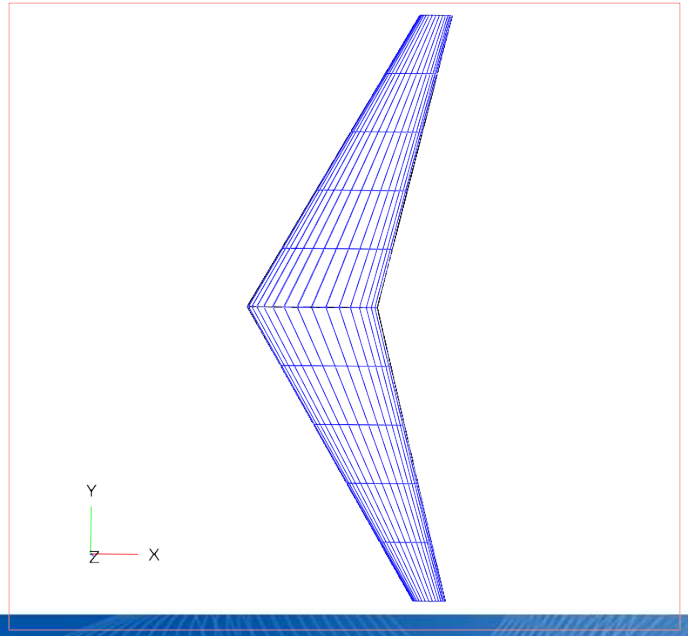
Minimize f_1 : The gross lift off-weight.

Limits of Gaussian processes

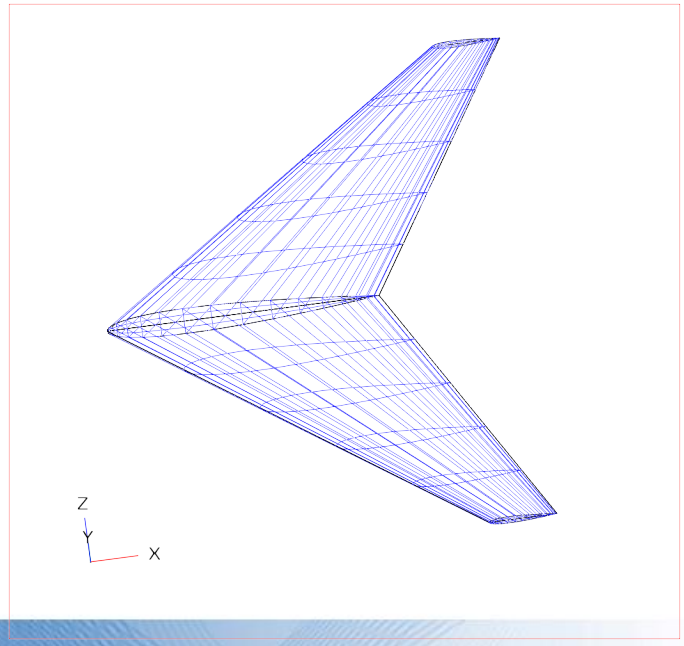
Limits of Gaussian processes for the analysis and optimization of design problems :

- Optimization of non-stationary functions,
- Multi-objective problems with correlated objectives,
- Multi-fidelity problems with varying input space dimensions.

Wing represented as a 2D object in low-fidelity



Wing represented as a 3D object in high-fidelity



Thesis objective

Thesis objective

Develop new algorithms and models based on the hierarchical generalization of Gaussian processes called **deep Gaussian processes** [Damianou and Lawrence, 2013] to overcome the limitations of regular Gaussian processes in the analysis and optimization of complex systems

Thesis objective

Thesis objective

Develop new algorithms and models based on the hierarchical generalization of Gaussian processes called **deep Gaussian processes** [Damianou and Lawrence, 2013] to overcome the limitations of regular Gaussian processes in the analysis and optimization of complex systems

Contributions on Bayesian Optimization

Thesis objective

Thesis objective

Develop new algorithms and models based on the hierarchical generalization of Gaussian processes called **deep Gaussian processes** [Damianou and Lawrence, 2013] to overcome the limitations of regular Gaussian processes in the analysis and optimization of complex systems

Contributions on Bayesian Optimization

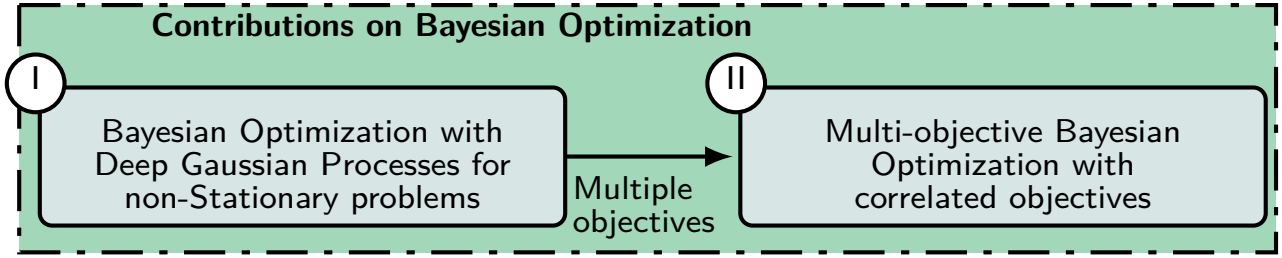
I

Bayesian Optimization with Deep Gaussian Processes for non-Stationary problems

Thesis objective

Thesis objective

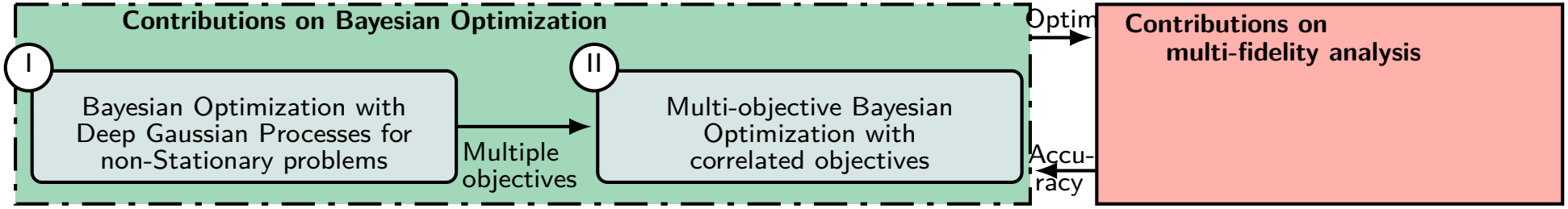
Develop new algorithms and models based on the hierarchical generalization of Gaussian processes called **deep Gaussian processes** [Damianou and Lawrence, 2013] to overcome the limitations of regular Gaussian processes in the analysis and optimization of complex systems



Thesis objective

Thesis objective

Develop new algorithms and models based on the hierarchical generalization of Gaussian processes called **deep Gaussian processes** [Damianou and Lawrence, 2013] to overcome the limitations of regular Gaussian processes in the analysis and optimization of complex systems



Thesis objective

Thesis objective

Develop new algorithms and models based on the hierarchical generalization of Gaussian processes called **deep Gaussian processes** [Damianou and Lawrence, 2013] to overcome the limitations of regular Gaussian processes in the analysis and optimization of complex systems

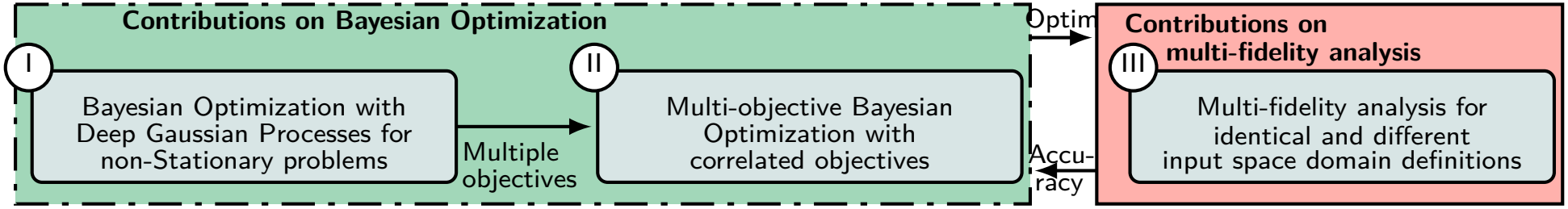


Table of Contents

Introduction

Background on Gaussian processes

Bayesian optimization for non-stationary functions

Multi-objective Bayesian optimization

Multi-fidelity with varying input spaces

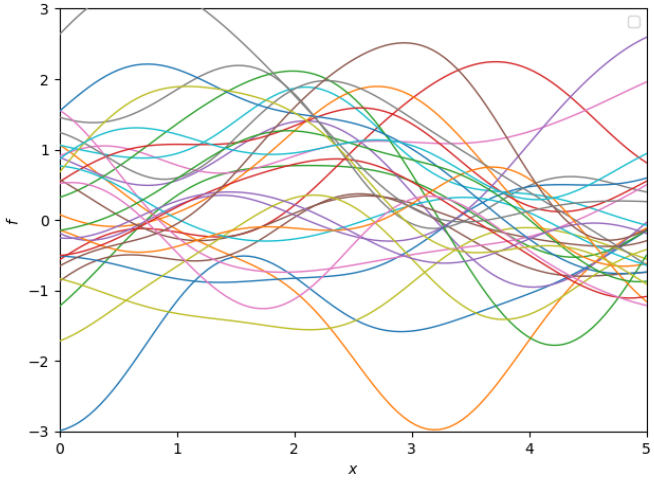
Conclusions and perspectives

Gaussian process regression

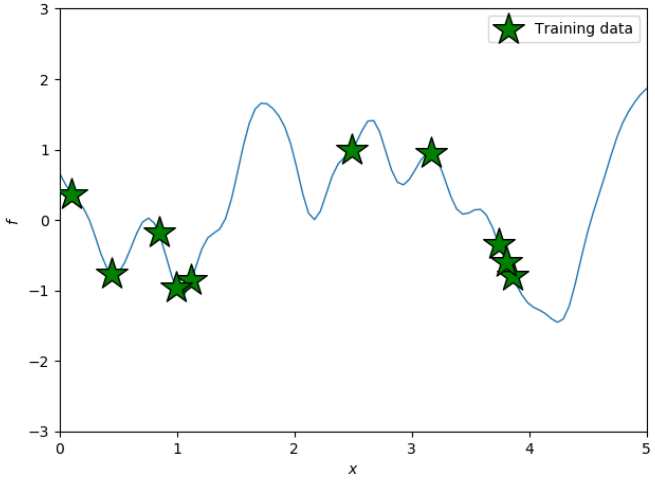
Gaussian process [Rasmussen, 2004]

A Gaussian process is used to describe a distribution over function. It is a collection of infinite random variables, **any finite number of which have a joint Gaussian distribution**.
It is defined by its mean function $\mu(.)$ and covariance function $k^{\Theta}(.)$ (Kernel) :
$f(.) \sim \mathcal{GP}(\mu(.), k^{\Theta}(.))$

GP with a squared exponential kernel **prior** samples



GP with a squared exponential kernel **posterior** samples

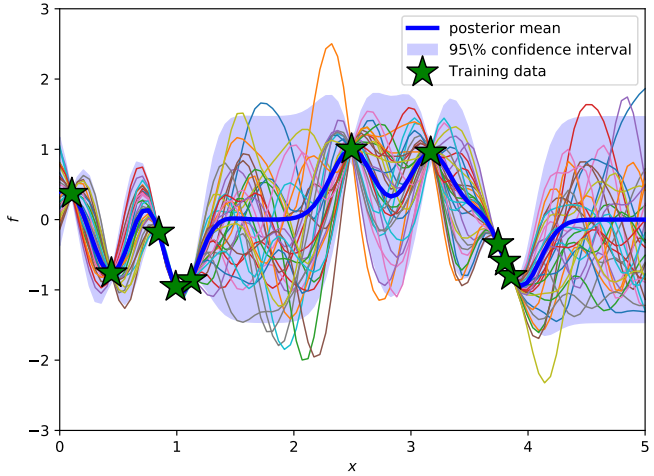


Gaussian process regression

Gaussian process [Rasmussen, 2004]

A Gaussian process is used to describe a distribution over function. It is a collection of infinite random variables, **any finite number of which have a joint Gaussian distribution**.
It is defined by its mean function $\mu(.)$ and covariance function $k^{\Theta}(.)$ (Kernel) :
$f(.) \sim \mathcal{GP}(\mu(.), k^{\Theta}(.))$

Posterior prediction using a GP with a RBF kernel



$$f(\mathbf{x})|\mathbf{y}; \mathbf{X}, \mathbf{x} \sim \mathcal{N}(\hat{f}(\mathbf{x}), \sigma_f(\mathbf{x}))$$

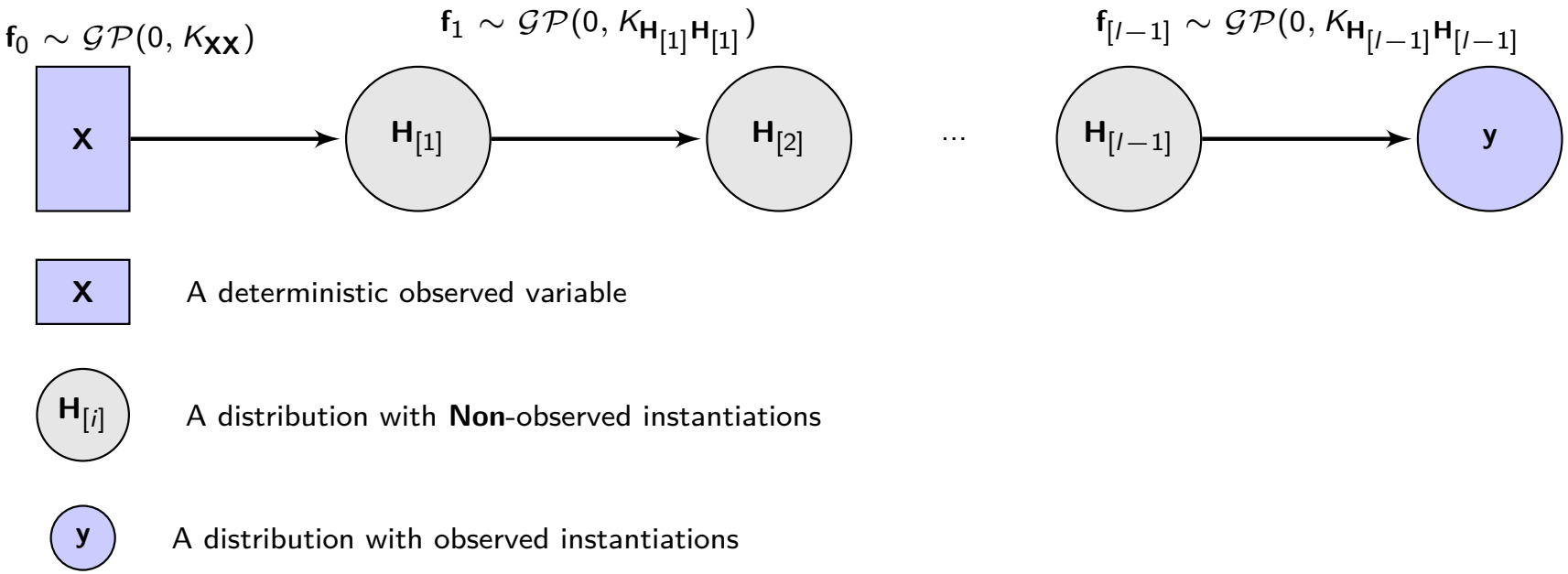
- $\hat{f}(\mathbf{x}) = \mu(\mathbf{x}) + k^{\hat{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{X})\left(k^{\hat{\boldsymbol{\theta}}}(\mathbf{X}, \mathbf{X})\right)^{-1}(\mathbf{y} - \mathbf{1}\mu(\mathbf{X}))$
- $\sigma_f(\mathbf{x}) = k^{\hat{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{x}) - k^{\hat{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{X})\left(k^{\hat{\boldsymbol{\theta}}}(\mathbf{X}, \mathbf{X})\right)^{-1} k^{\hat{\boldsymbol{\theta}}}(\mathbf{X}, \mathbf{x})$

Deep Gaussian processes

Deep Gaussian processes [Damianou and Lawrence, 2013]

Deep Gaussian processes are a class of surrogate models based on the structure of deep neural networks, where each layer is a GP. It considers that the statistical relationship between the inputs and the response is expressed by a functional composition of GPs :

$$y = f_{[l-1]}(\mathbf{f}_{[l-2]}(\ldots(\mathbf{f}_{[1]}(\mathbf{f}_{[0]}(\mathbf{x}))))$$

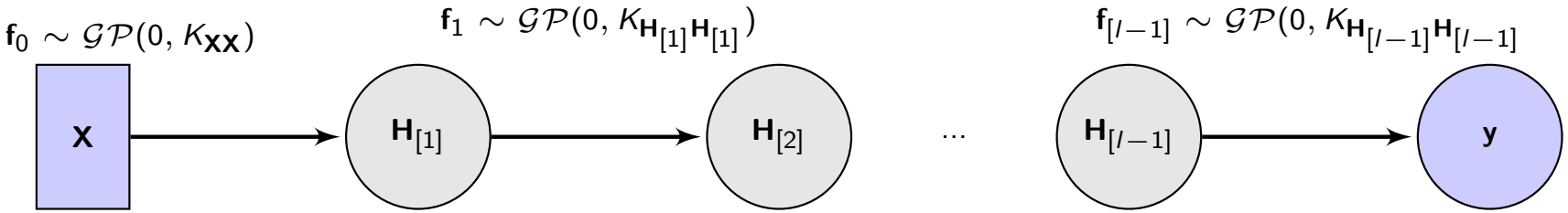


Deep Gaussian processes

Deep Gaussian processes [Damianou and Lawrence, 2013]

Deep Gaussian processes are a class of surrogate models based on the structure of deep neural networks, where each layer is a GP. It considers that the statistical relationship between the inputs and the response is expressed by a functional composition of GPs :

$$y = f_{[l-1]}(\mathbf{f}_{[l-2]}(\ldots(\mathbf{f}_{[1]}(\mathbf{f}_{[0]}(\mathbf{x}))))$$



$$p(\mathbf{y}|\mathbf{X}) = \int_{\{H_{[i]}\}_1^{l-1}} p(\mathbf{y}|\mathbf{H}_{[l-1]}) \ldots p(\mathbf{H}_{[1]}|\mathbf{X}) \mathrm{d}\{\mathbf{H}_{[i]}\}_1^{l-1}$$

Analytically Intractable

Deep Gaussian process inference

Approach	Inference approach	Approximation approach
[Damianou and Lawrence, 2013]	Variational inference	Variational sparse GPs
[Dai et al., 2015]	Variational inference	Variational sparse GPs
[Salimbeni and Deisenroth, 2017]	Variational inference	Variational sparse GPs
[Haibin et al., 2019]	Variational inference	Variational sparse GPs
[Cutajar et al., 2017]	Variational inference	Random feature-based GP
[Bui et al., 2016]	Expectation propagation	Fully independent training conditional GPs
[Havasi et al., 2018]	Markov-Chain Monte-Carlo	Variational sparse GPs
[Rossi et al., 2020]	Markov-Chain Monte-Carlo	Variational sparse GPs

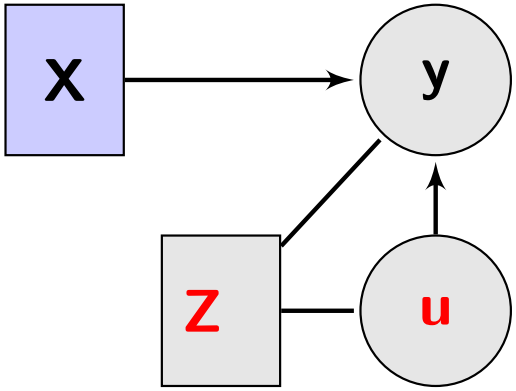
Deep Gaussian process inference

Approach	Inference approach	Approximation approach
[Salimbeni and Deisenroth, 2017]	Variational inference	Variational sparse GPs

Deep Gaussian process inference

Approach	Inference approach	Approximation approach
[Salimbeni and Deisenroth, 2017]	Variational inference	Variational sparse GPs

$f \sim \mathcal{GP}(0, K_{\mathbf{XX}})$

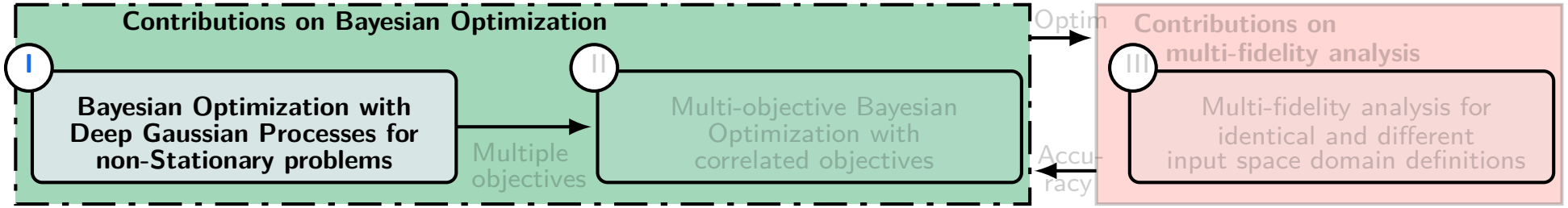


Variational approximation

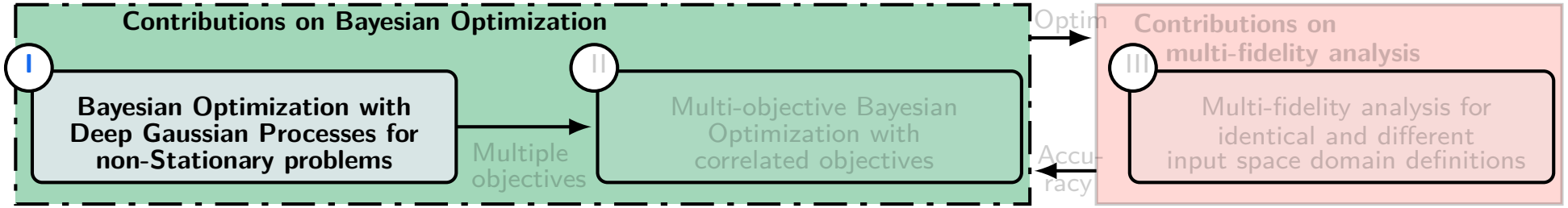
$$q(f, u|y) = p(f|u)q(u)$$

[Titsias, 2009]

Contribution I



Contribution I

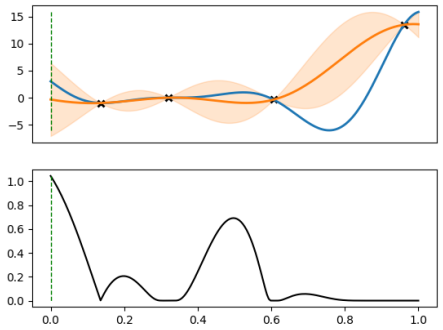


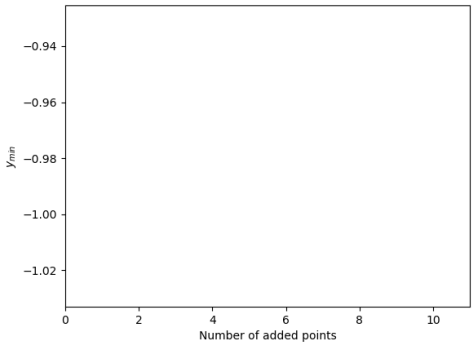
$$
(\mathcal{P}_{\min}) \left| \begin{array}{lll} \underset{\mathbf{x}}{\text{Minimize}} & y & = f(\mathbf{x}) \\ \text{with} & \mathbf{x} & \in \mathcal{X} \subset \mathbb{R}^d \\ \text{subject to} & g_j(\mathbf{x}) & \leq 0, \ \forall j \in \{1, \ldots, n_c\} \end{array} \right.
$$

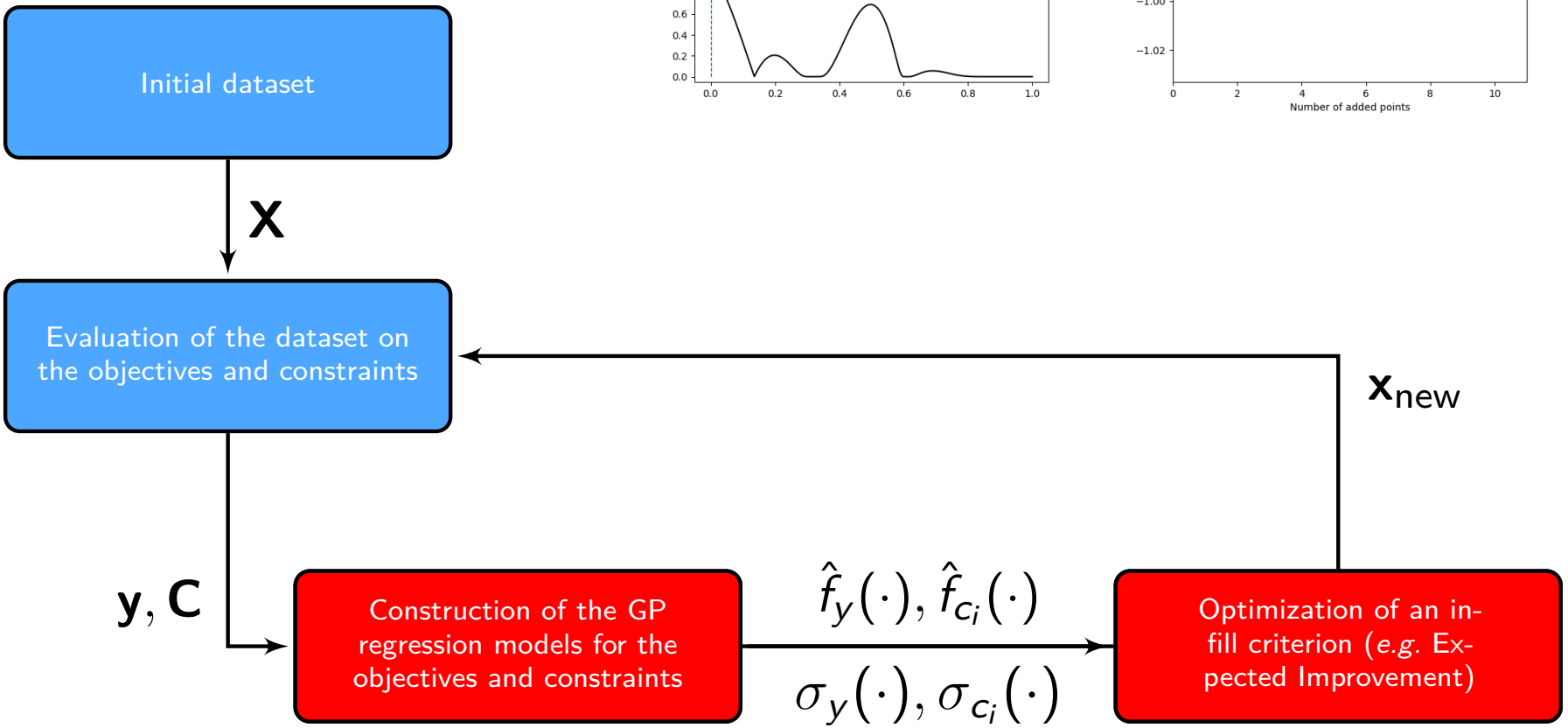
Bayesian optimization framework

Bayesian Optimization (BO)

[Jones et al., 1998]



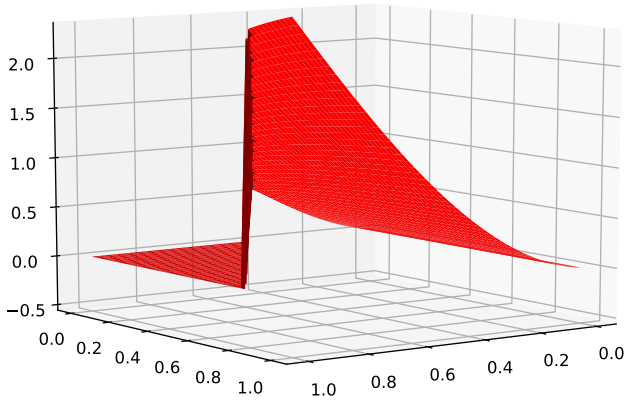




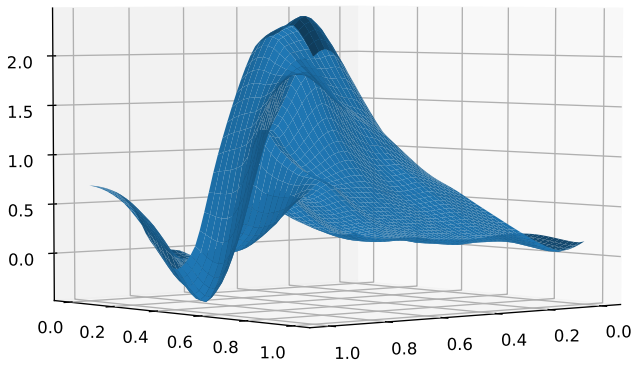
Limits of Bayesian optimization with Gaussian processes

Handling non stationarity

- Abrupt change in the response
- Different behavior according to the input



2-D Function with an abrupt change

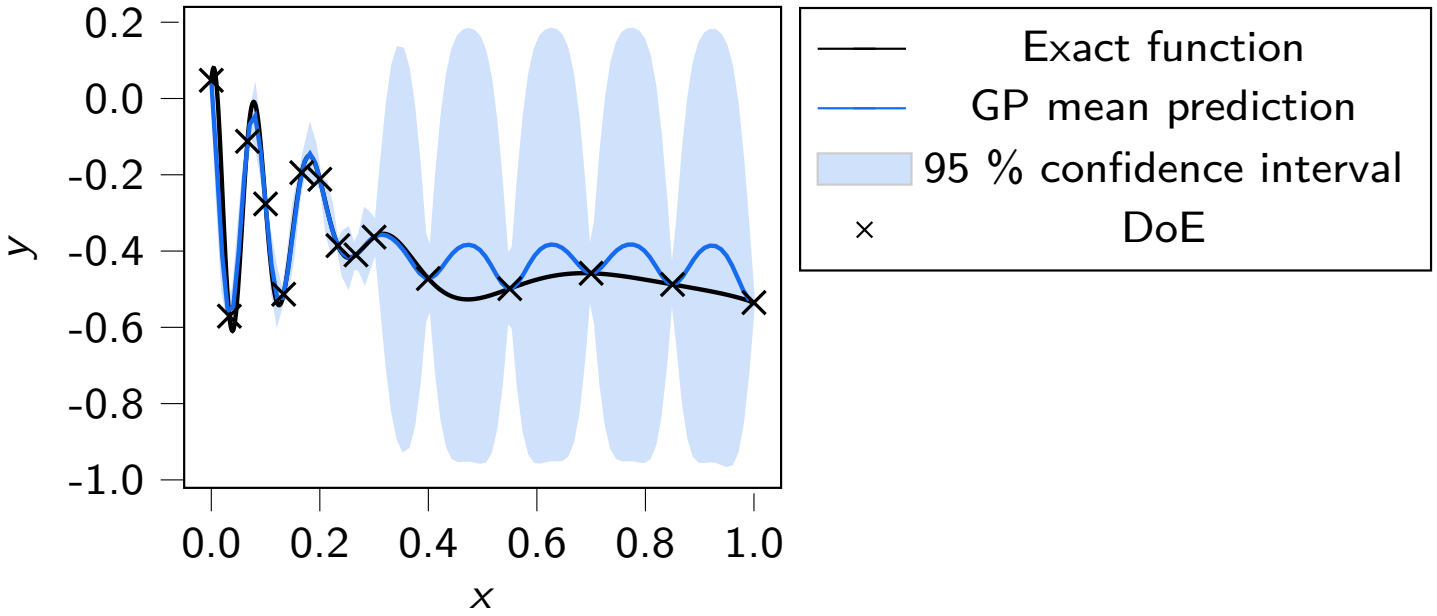


GP Approximation

Limits of Bayesian optimization with Gaussian processes

Handling non stationarity

- Abrupt change in the response
- Different behavior according to the input



1-D

Function with a different behavior according to the input space

Review of methods of handling non stationarity

- Direct formulation of non-stationary covariance function :
  - Non stationary version of the squared exponential covariance function [Higdon et al., 1999],
  - Non stationary version of the Matérn covariance function [Paciorek and Schervish, 2006].
- Local stationary covariance function :
  - Moving window [Haas, 1990], tree-based GPs [Gramacy and Lee, 2008] ...
  - Different stationary GPs in different subspaces of the input space (mixture of GPs) [Rasmussen and Ghahramani, 2002, Bartoli et al., 2017].
- Non-linear mapping :
  - A parametric mapping from the input space into a deformed space [Xiong et al., 2007, Snoek et al., 2014].

Review of methods of handling non stationarity

- Direct formulation of non-stationary covariance function :
  - Non stationary version of the squared exponential covariance function [Higdon et al., 1999],
  - Non stationary version of the Matérn covariance function [Paciorek and Schervish, 2006].
- Local stationary covariance function :
  - Moving window [Haas, 1990], tree-based GPs [Gramacy and Lee, 2008] ...
  - Different stationary GPs in different subspaces of the input space (mixture of GPs) [Rasmussen and Ghahramani, 2002, Bartoli et al., 2017].
- Non-linear mapping :
  - A parametric mapping from the input space into a deformed space [Xiong et al., 2007, Snoek et al., 2014].

Review of methods of handling non stationarity

- Direct formulation of non-stationary covariance function :
  - Non stationary version of the squared exponential covariance function [Higdon et al., 1999],
  - Non stationary version of the Matérn covariance function [Paciorek and Schervish, 2006].
  - ✗ Adapted to only few dimensions
- Local stationary covariance function :
  - Moving window [Haas, 1990], tree-based GPs [Gramacy and Lee, 2008] ...
  - Different stationary GPs in different subspaces of the input space (mixture of GPs) [Rasmussen and Ghahramani, 2002, Bartoli et al., 2017].
- Non-linear mapping :
  - A parametric mapping from the input space into a deformed space [Xiong et al., 2007, Snoek et al., 2014].

Review of methods of handling non stationarity

- Direct formulation of non-stationary covariance function :
  - Non stationary version of the squared exponential covariance function [Higdon et al., 1999],
  - Non stationary version of the Matérn covariance function [Paciorek and Schervish, 2006].
- Local stationary covariance function :
  - Moving window [Haas, 1990], tree-based GPs [Gramacy and Lee, 2008] ...
  - Different stationary GPs in different subspaces of the input space (mixture of GPs) [Rasmussen and Ghahramani, 2002, Bartoli et al., 2017].
- Non-linear mapping :
  - A parametric mapping from the input space into a deformed space [Xiong et al., 2007, Snoek et al., 2014].

Review of methods of handling non stationarity

- Direct formulation of non-stationary covariance function :
  - Non stationary version of the squared exponential covariance function [Higdon et al., 1999],
  - Non stationary version of the Matérn covariance function [Paciorek and Schervish, 2006].
- Local stationary covariance function :
  - Moving window [Haas, 1990], tree-based GPs [Gramacy and Lee, 2008] ...
  - Different stationary GPs in different subspaces of the input space (mixture of GPs) [Rasmussen and Ghahramani, 2002, Bartoli et al., 2017].
  - ✗ Subdivision of scarce data-sets
- Non-linear mapping :
  - A parametric mapping from the input space into a deformed space [Xiong et al., 2007, Snoek et al., 2014].

Review of methods of handling non stationarity

- Direct formulation of non-stationary covariance function :
  - Non stationary version of the squared exponential covariance function [Higdon et al., 1999],
  - Non stationary version of the Matérn covariance function [Paciorek and Schervish, 2006].
- Local stationary covariance function :
  - Moving window [Haas, 1990], tree-based GPs [Gramacy and Lee, 2008] ...
  - Different stationary GPs in different subspaces of the input space (mixture of GPs) [Rasmussen and Ghahramani, 2002, Bartoli et al., 2017].
- Non-linear mapping :
  - A parametric mapping from the input space into a deformed space [Xiong et al., 2007, Snoek et al., 2014].

Review of methods of handling non stationarity

- Direct formulation of non-stationary covariance function :
  - Non stationary version of the squared exponential covariance function [Higdon et al., 1999],
  - Non stationary version of the Matérn covariance function [Paciorek and Schervish, 2006].
- Local stationary covariance function :
  - Moving window [Haas, 1990], tree-based GPs [Gramacy and Lee, 2008] ...
  - Different stationary GPs in different subspaces of the input space (mixture of GPs) [Rasmussen and Ghahramani, 2002, Bartoli et al., 2017].
- Non-linear mapping :
  - A parametric mapping from the input space into a deformed space [Xiong et al., 2007, Snoek et al., 2014].
  - ✗ Parametric mappings are problem dependent

Deep Gaussian processes for non-stationarity

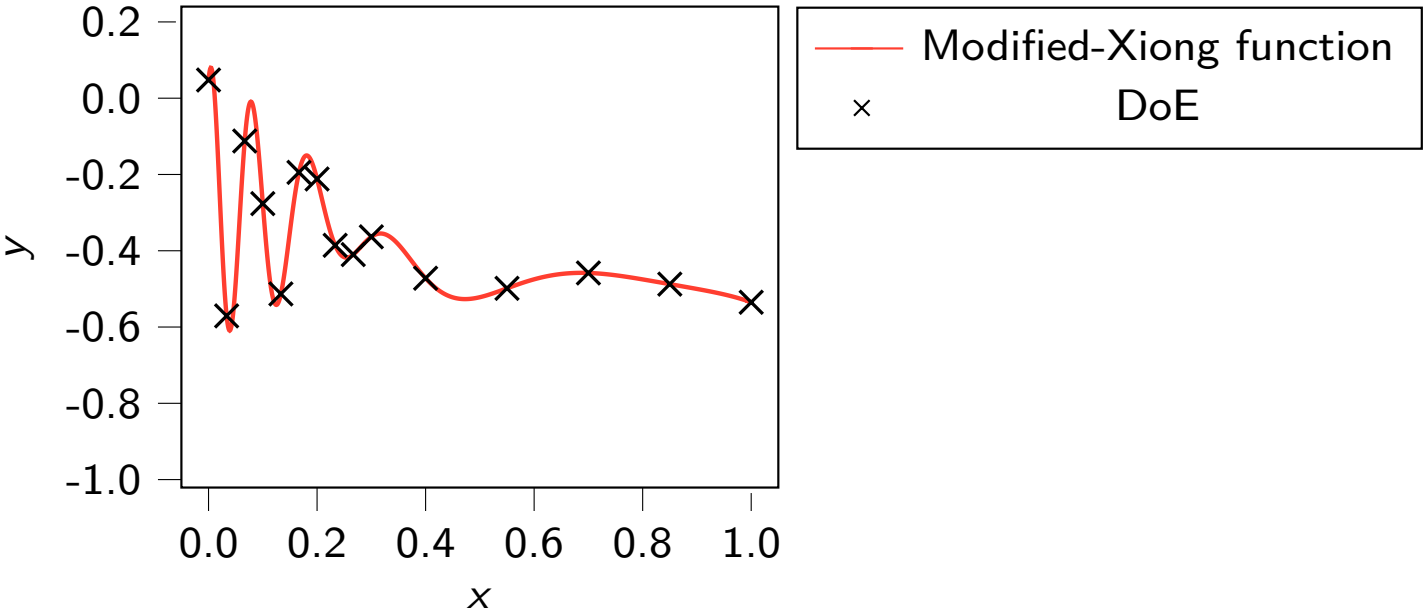
Deep Gaussian processes for non-stationarity

Deep Gaussian processes allow a flexible way of Bayesian kernel construction through input warping and dimensionality expansion to better fit data.

Deep Gaussian processes for non-stationarity

Deep Gaussian processes for non-stationarity

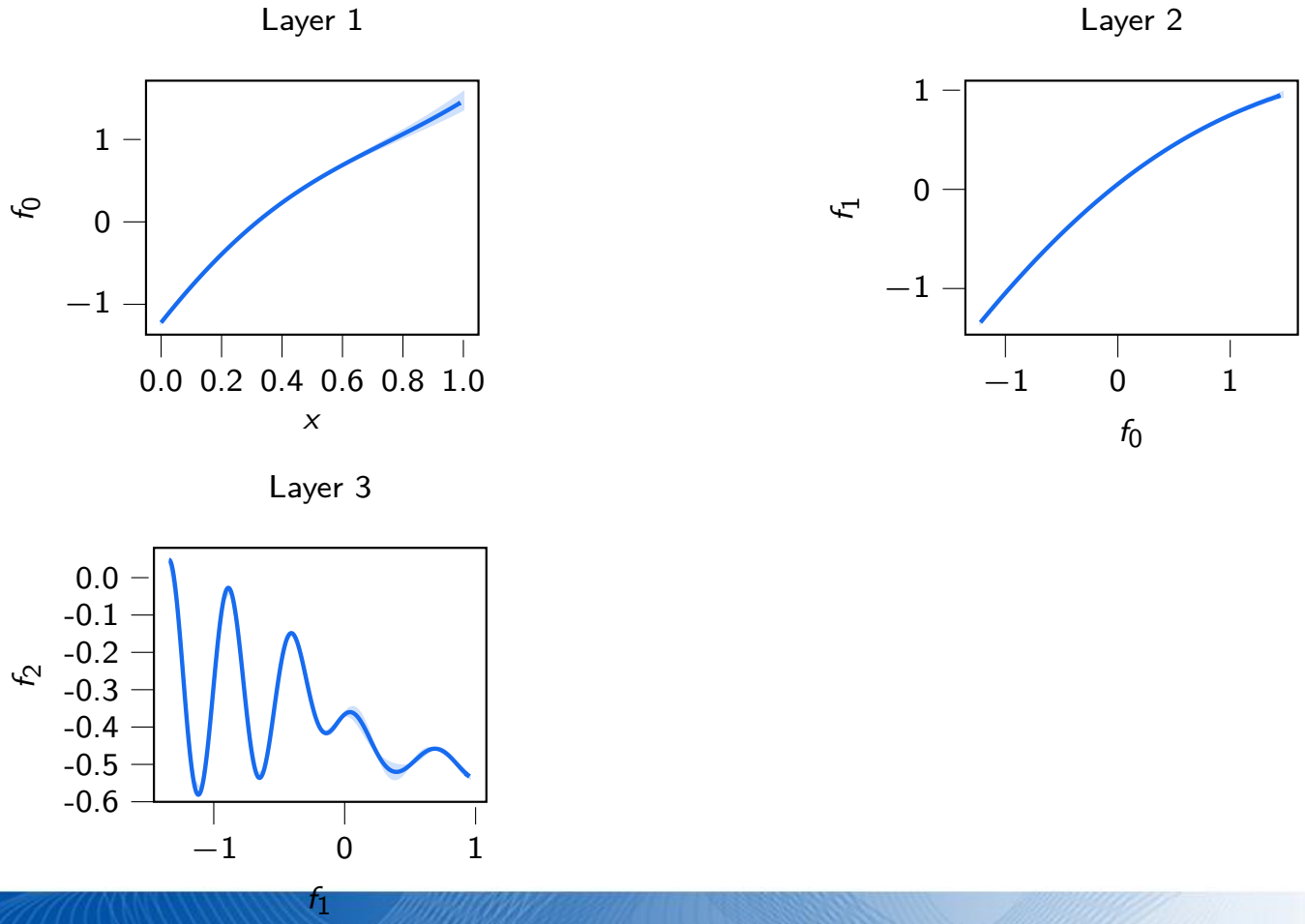
Deep Gaussian processes allow a flexible way of Bayesian kernel construction through input warping and dimensionality expansion to better fit data.



Deep Gaussian processes for non-stationarity

Deep Gaussian processes for non-stationarity

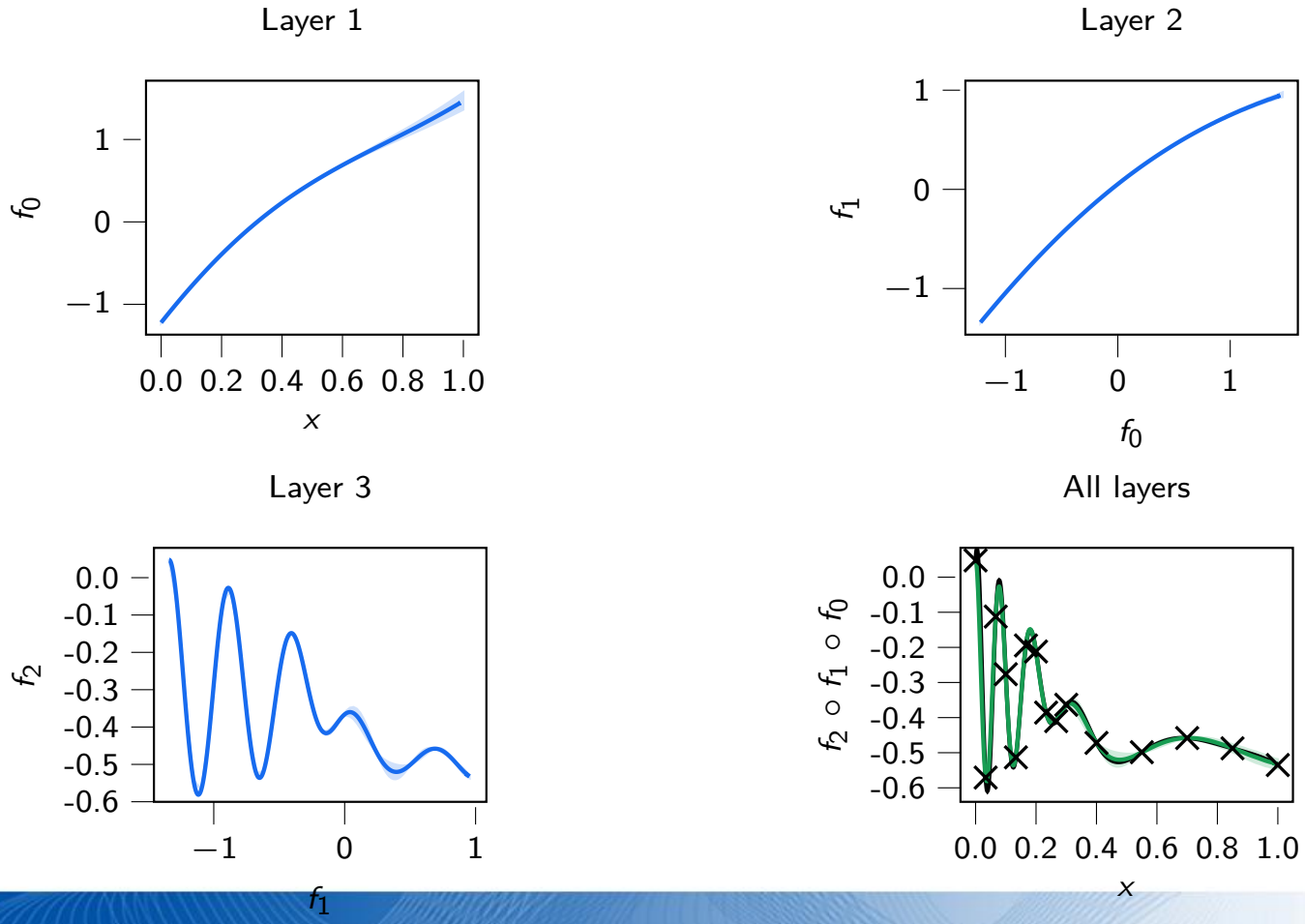
Deep Gaussian processes allow a flexible way of Bayesian kernel construction through input warping and dimensionality expansion to better fit data.



Deep Gaussian processes for non-stationarity

Deep Gaussian processes for non-stationarity

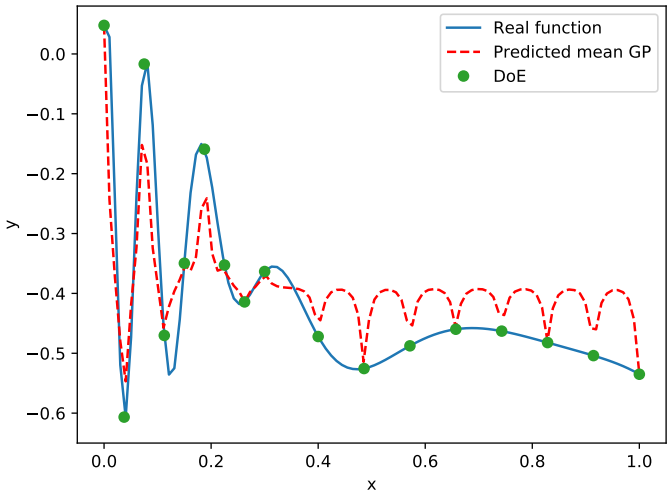
Deep Gaussian processes allow a flexible way of Bayesian kernel construction through input warping and dimensionality expansion to better fit data.



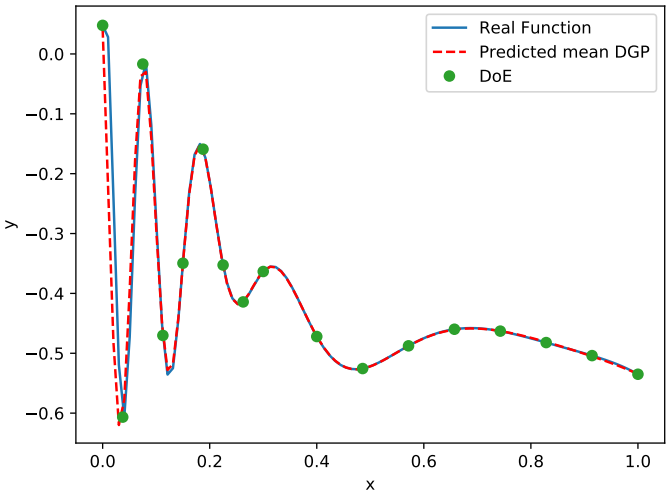
Deep Gaussian processes for non-stationarity

Deep Gaussian processes for non-stationarity

Deep Gaussian processes allow a flexible way of Bayesian kernel construction through input warping and dimensionality expansion to better fit data.



GP prediction of a non-stationary 1-D function.



DGP prediction of a non-stationary 1-D function.

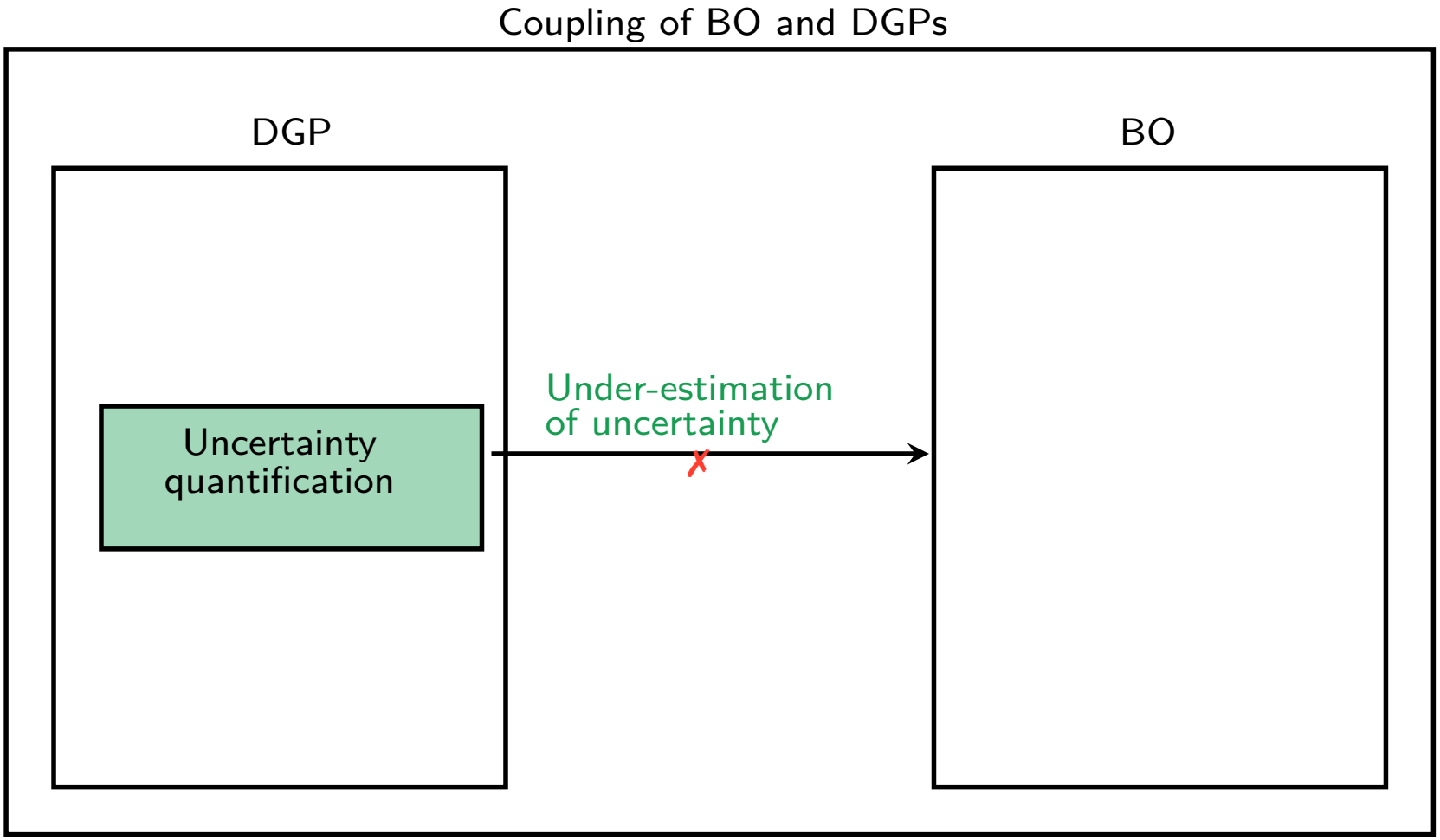
Coupling of Bayesian optimization and deep Gaussian processes

Coupling of BO and DGPs

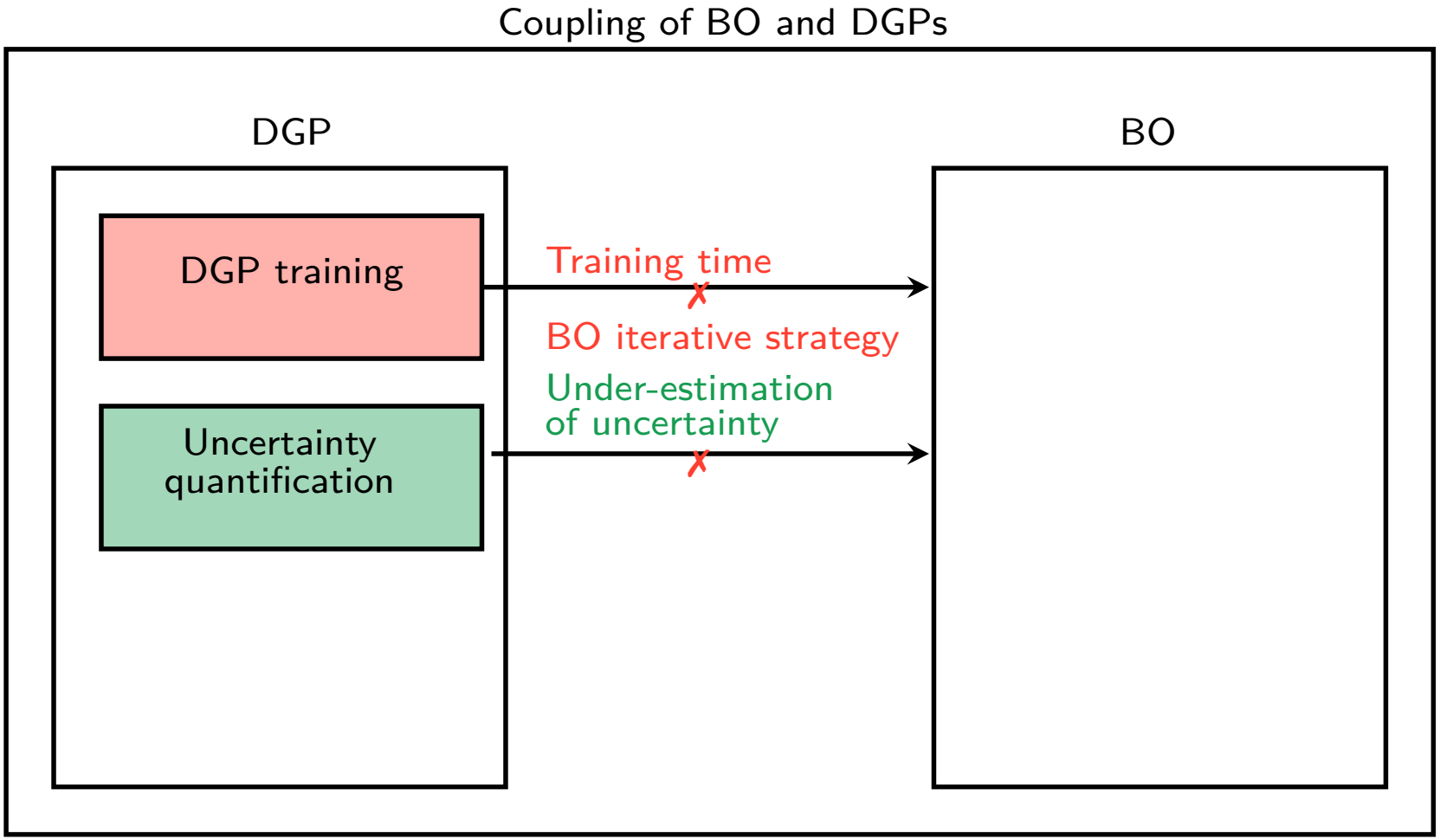
DGP

BO

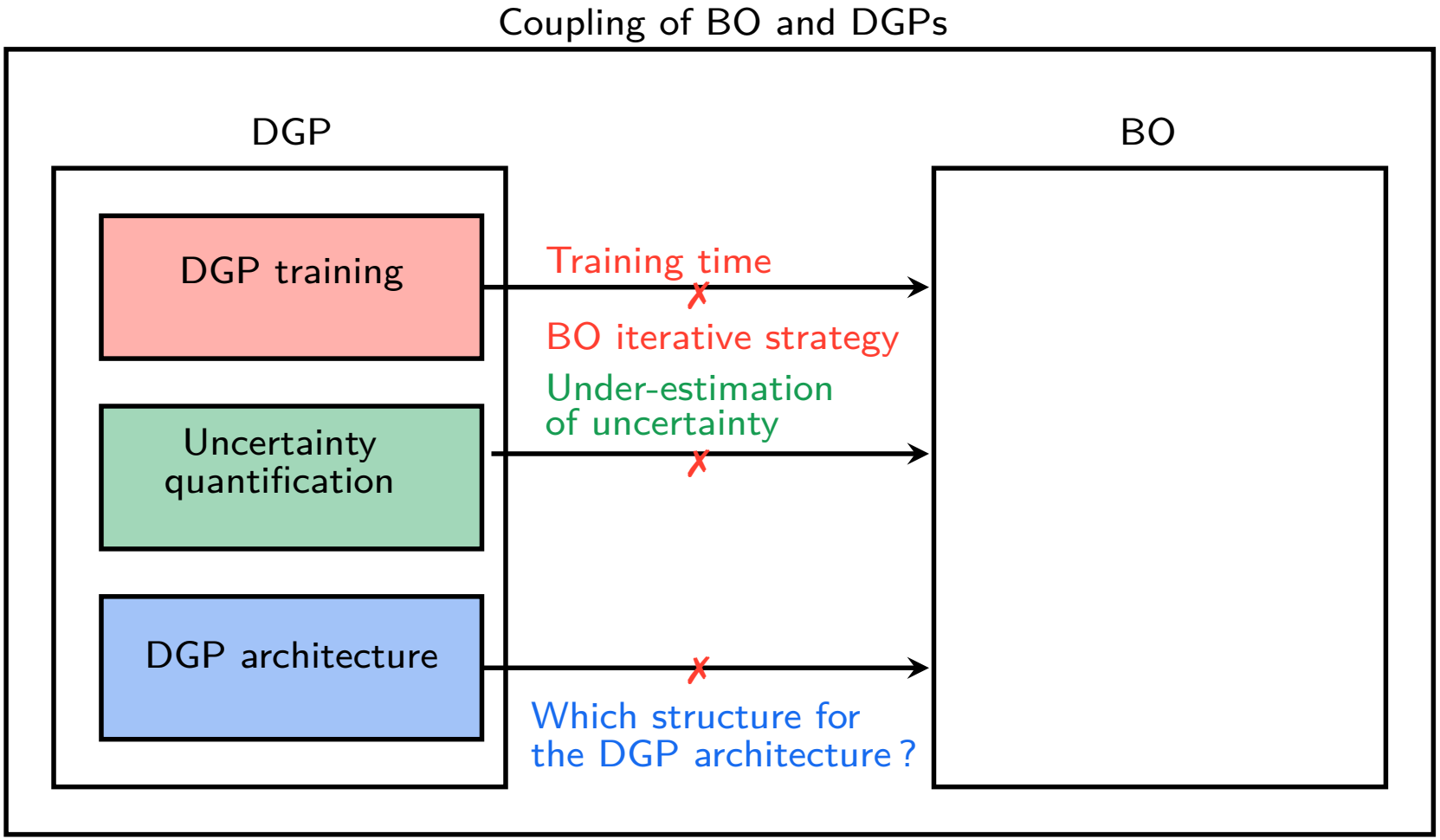
Coupling of Bayesian optimization and deep Gaussian processes



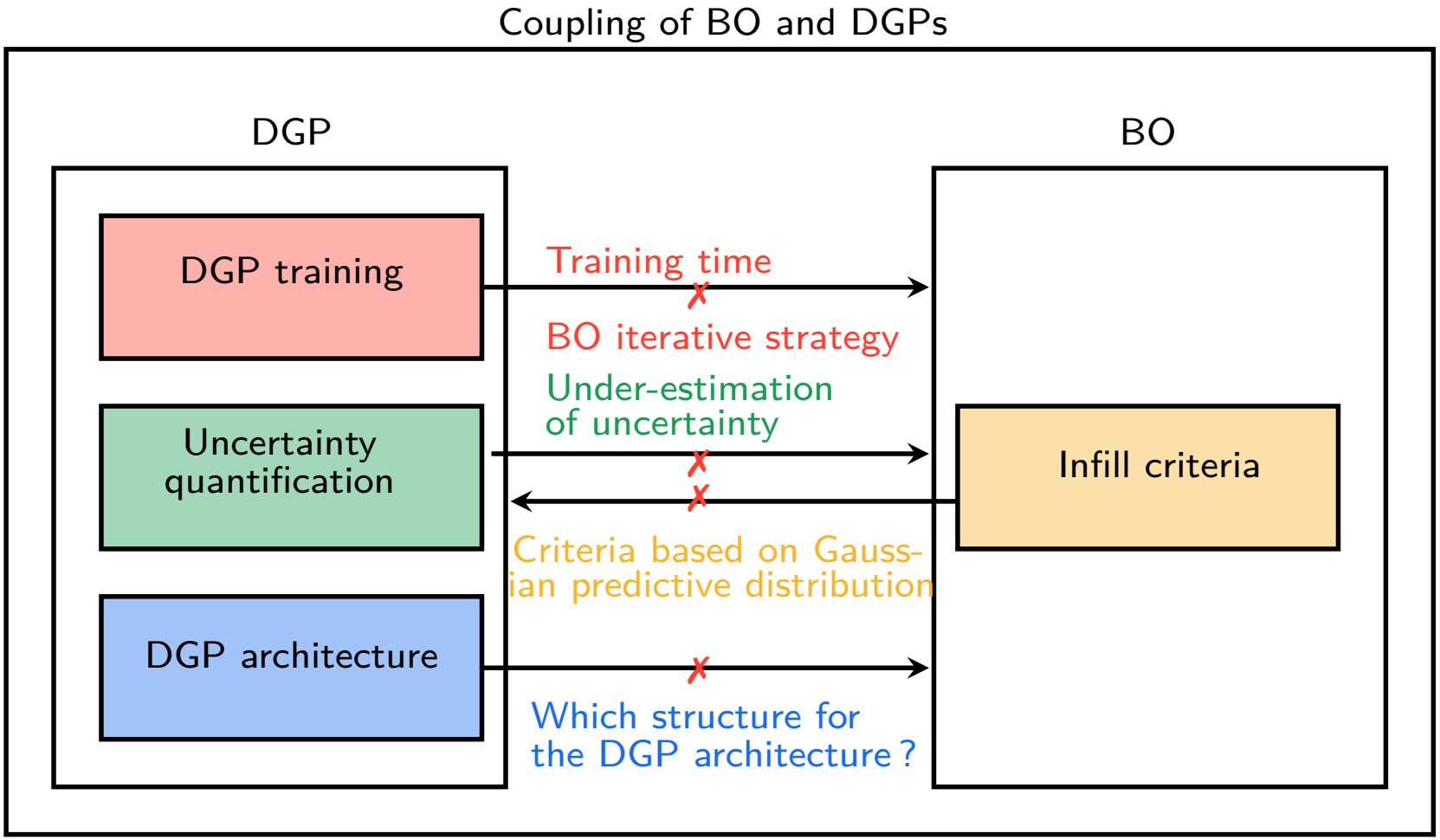
Coupling of Bayesian optimization and deep Gaussian processes



Coupling of Bayesian optimization and deep Gaussian processes

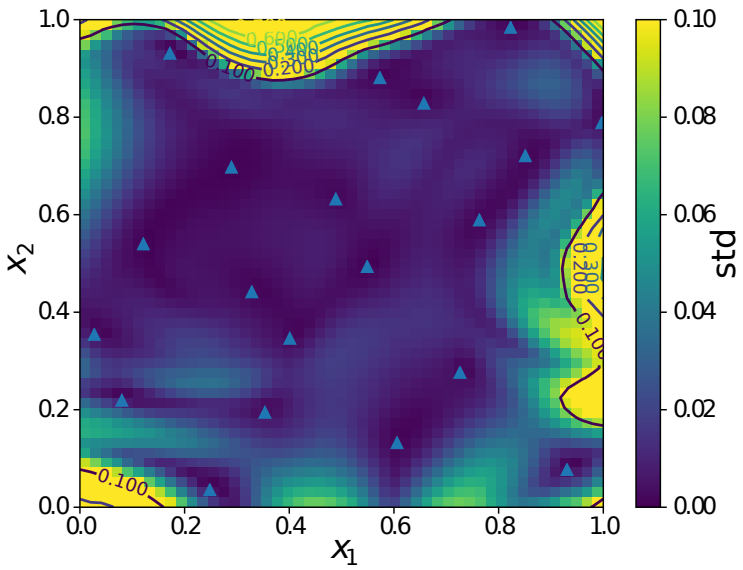


Coupling of Bayesian optimization and deep Gaussian processes



Deep Gaussian processes uncertainty under-estimation

The limitation of DGPs : **uncertainty quantification**

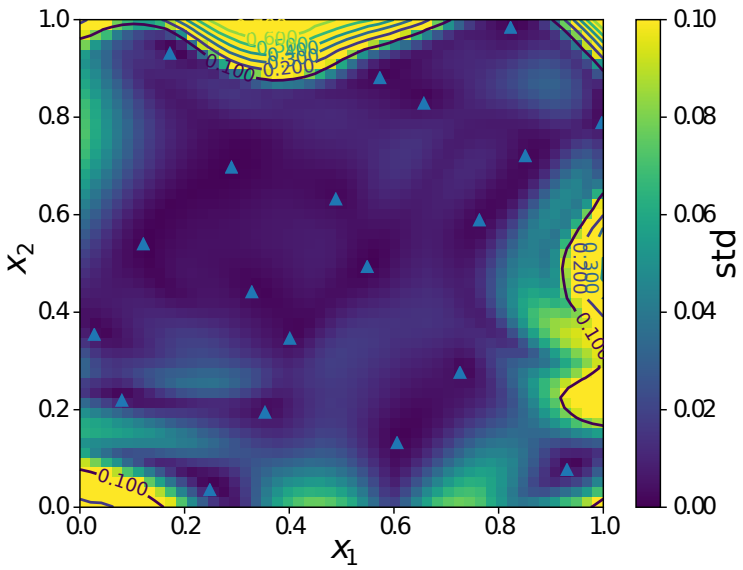


Standard deviation on the prediction given by a 2 hidden layers DGP model trained using Adam Optimizer.

- ✗Variational inference under-estimates the posterior uncertainty.
- ✗ The optimization of the Evidence Lower Bound (ELBO) with an ordinary gradient.

Deep Gaussian processes uncertainty under-estimation

The limitation of DGPs : **uncertainty quantification**

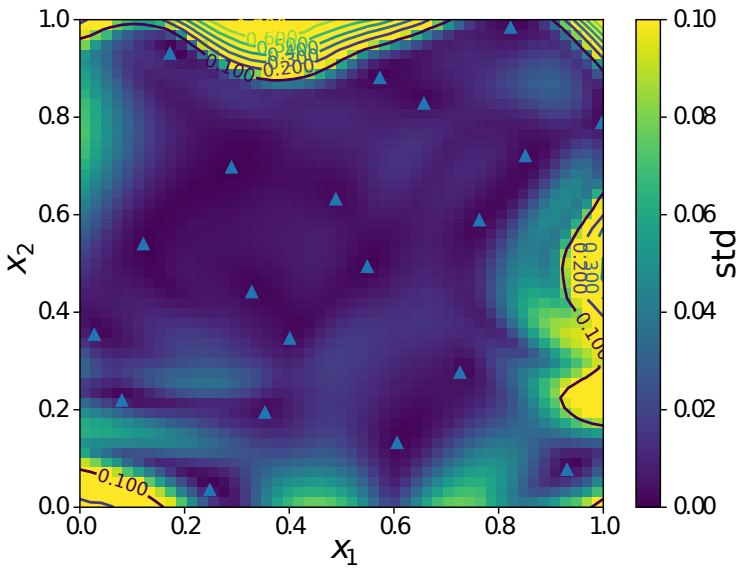


Standard deviation on the prediction given by a 2 hidden layers DGP model trained using Adam Optimizer.

- ✗Variational inference under-estimates the posterior uncertainty.
- ✗ The optimization of the Evidence Lower Bound (ELBO) with an ordinary gradient.

Deep Gaussian processes uncertainty under-estimation

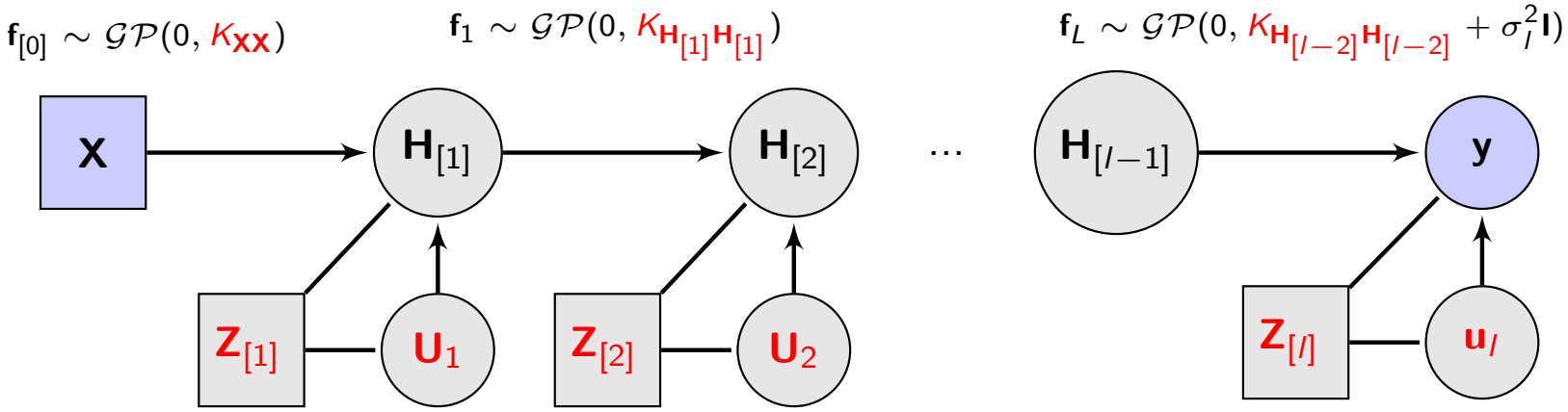
The limitation of DGPs : **uncertainty quantification**



Standard deviation on the prediction given by a 2 hidden layers DGP model trained using Adam Optimizer.

- ✗Variational inference under-estimates the posterior uncertainty.
- ✗ The optimization of the Evidence Lower Bound (ELBO) with an ordinary gradient.

Deep Gaussian processes uncertainty under-estimation

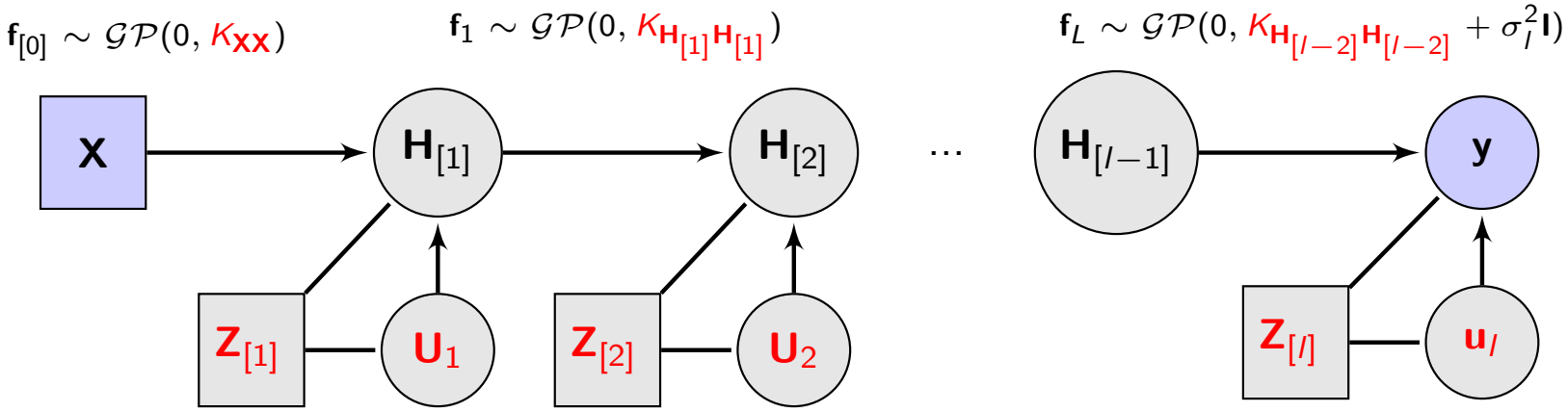


Limits of ordinary gradient

✗ Issue : Parameter space not euclidean ($q(\mathbf{U})$) $\rightarrow$ stochastic ordinary gradient may not be adapted

- Not the steepest descent direction $\rightarrow$ large number of iterations in the training,
- Poor optimization of the Evidence lower bound $\rightarrow$ limited predictive accuracy.

Deep Gaussian processes uncertainty under-estimation



Limits of ordinary gradient

✗ Issue : Parameter space not euclidean ($q(\mathbf{U})$) $\rightarrow$ stochastic ordinary gradient may not be adapted

- Not the steepest descent direction $\rightarrow$ large number of iterations in the training,
- Poor optimization of the Evidence lower bound $\rightarrow$ limited predictive accuracy.

Deep Gaussian processes uncertainty under-estimation

Limits of ordinary gradient

✗ Issue : Parameter space not euclidean ($q(\mathbf{U})$) $\rightarrow$ stochastic ordinary gradient may not be adapted

Combining natural and stochastic gradient for training

Optimization process adapted to the differential geometry of the parameter space $\rightarrow$ generalization of natural gradient [Amari, 1998] to deep Gaussian processes

Deep Gaussian processes uncertainty under-estimation

Limits of ordinary gradient

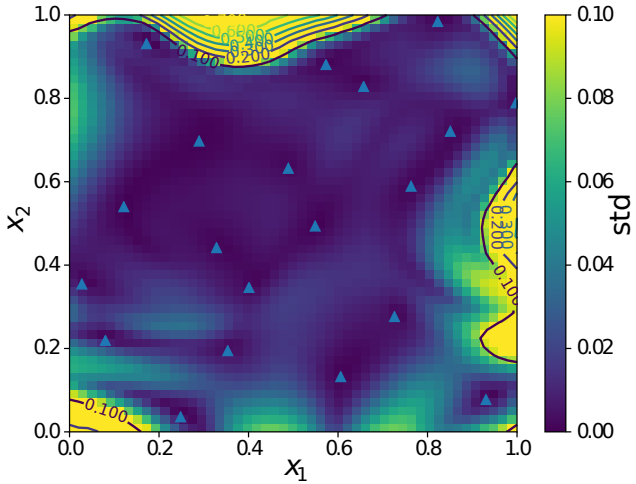
✗ Issue : Parameter space not euclidean ($q(\mathbf{U})$) $\rightarrow$ stochastic ordinary gradient may not be adapted

Combining natural and stochastic gradient for training

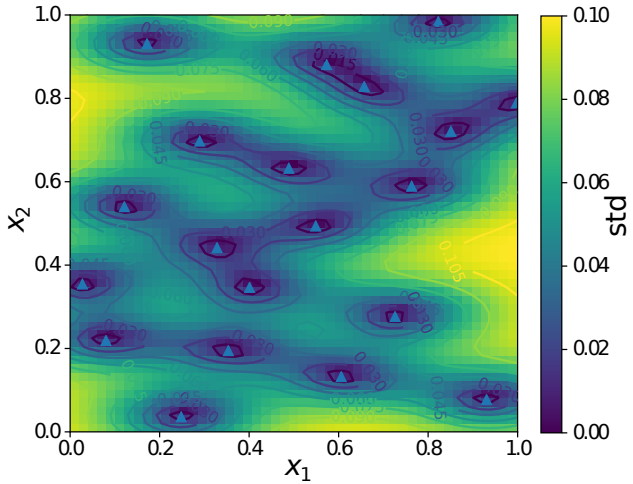
✓ Proposed optimization process adapted to the parameter space :
Loop between :

- step using stochastic ordinary gradient (Adam Optimizer) with respect to $\left(\{\Theta_{[i]}\}_{i=1}^{l}, \{\mathbf{Z}_{[i]}\}_{i=1}^{l}\right)$
- step using natural gradient with respect to all the variational distributions $\left(q\{(\mathbf{u}_{[i]})\}_{i=1}^{s}\right)$.

Deep Gaussian processes uncertainty under-estimation



A DGP predictive uncertainty using classic training.

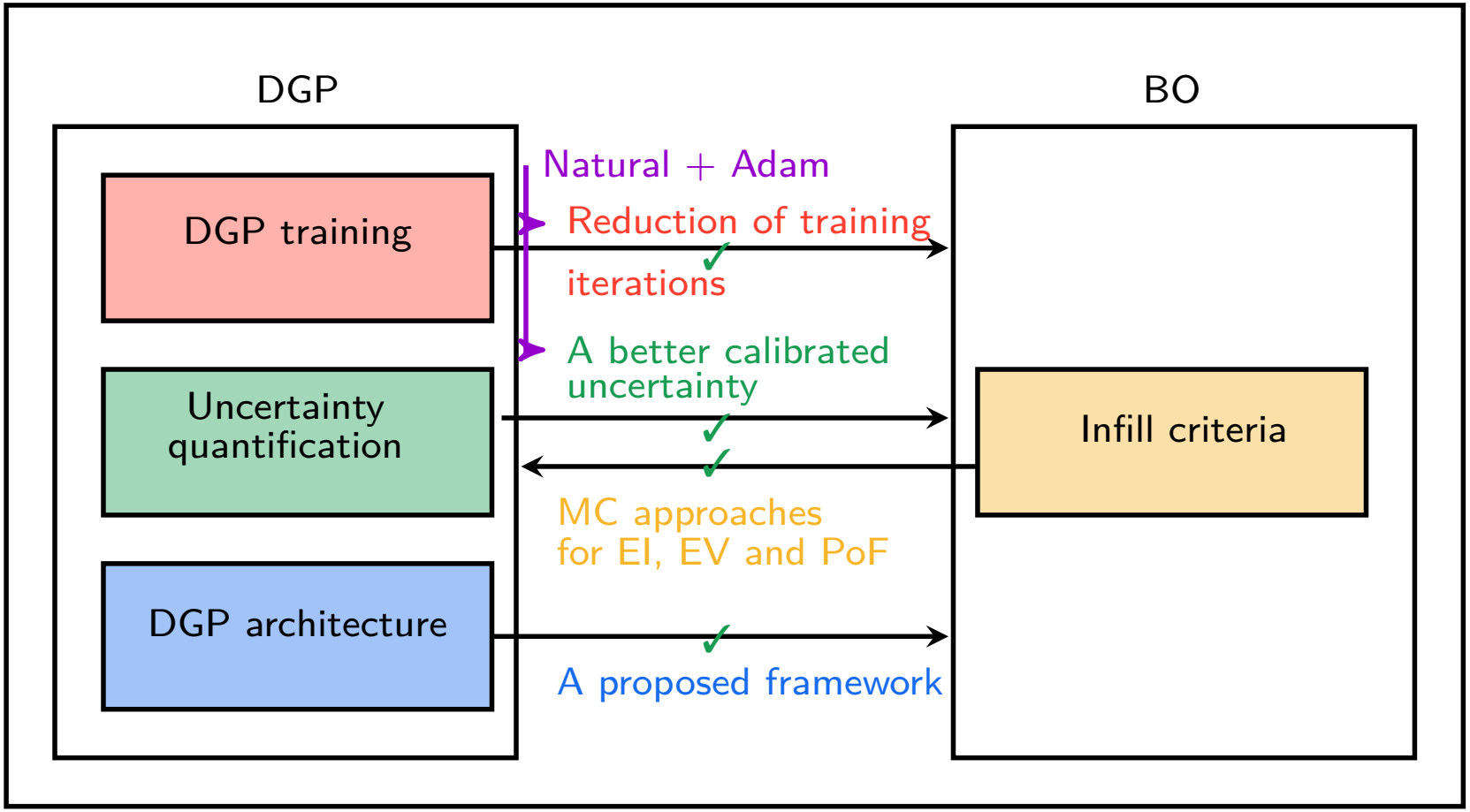


A DGP predictive uncertainty using the proposed training.

Reduction of the mean negative test-log likelihood by a factor of 100.

Bayesian optimization and deep Gaussian processes framework

Coupling of BO and DGPs



Experimentations

Assessment of the performance of the coupling Bayesian optimization & deep Gaussian processes :

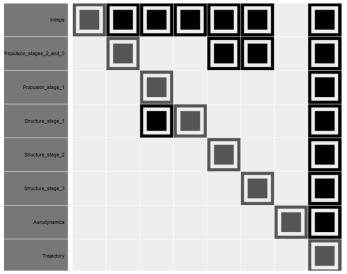
- on a benchmark of analytical test cases,
- on 2 design aerospace problems.

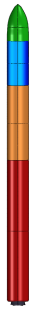
with respect to state of the art approaches to handle non-stationarity in Bayesian optimization.

Application to industrial test case : design of an aerospace vehicle

Problem : Design optimization of a three stage sounding rocket.
Objective : Maximize the final altitude h_{max} to release the payload experiments.

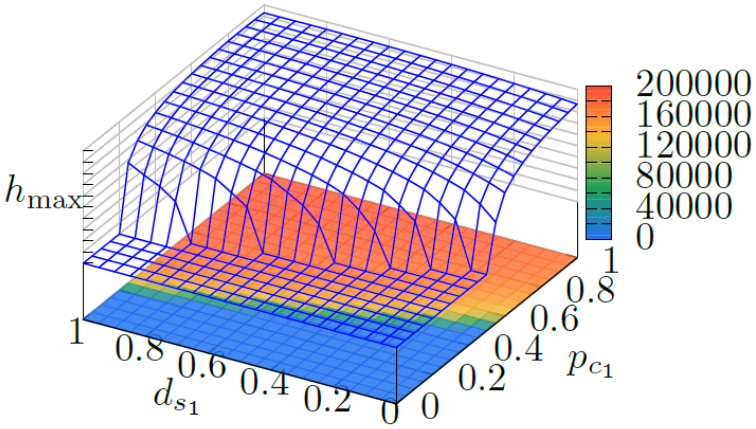
$$
\begin{array}{ll}
\text{Minimize :} & -h_{\max}(\mathbf{x}) \\
\text{w.r.t :} & \mathbf{x} = [d_1, d_2, d_3, m_1, m_2, m_3, p_{c_1}, p_{c_2}, \\
& p_{c_3}, d_{c_1}, d_{s_1}, q_2, q_3, OF_2, OF_3] \\
\text{s.t :} & \left\{ \begin{array}{l} \text{1 structural constraint} \\ \text{8 constraints on the integrity} \\ \text{of the first stage} \\ \text{3 constraints on the maximal} \\ \text{axial load factor} \\ \text{1 constraint on the maximal} \\ \text{GLOW allowed} \end{array} \right.
\end{array}
$$





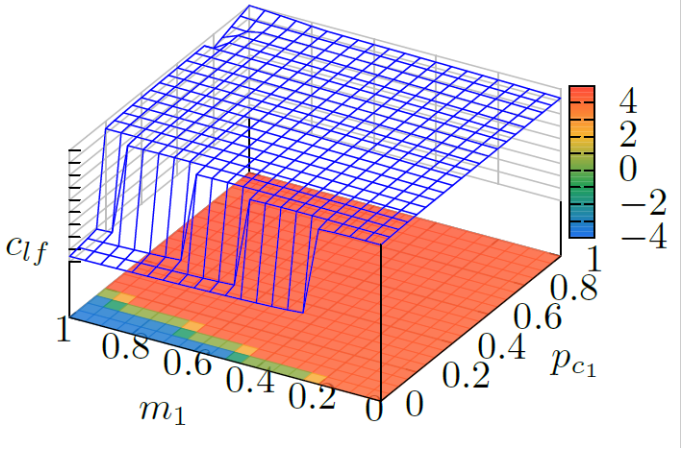
Design process of the three stage sounding rocket

Application to industrial test case : design of aerospace vehicle



The final altitude $h_{\max}$ w.r.t :

- the exit nozzle diameter d_{s_1}
- the chamber pressure p_{c_1}



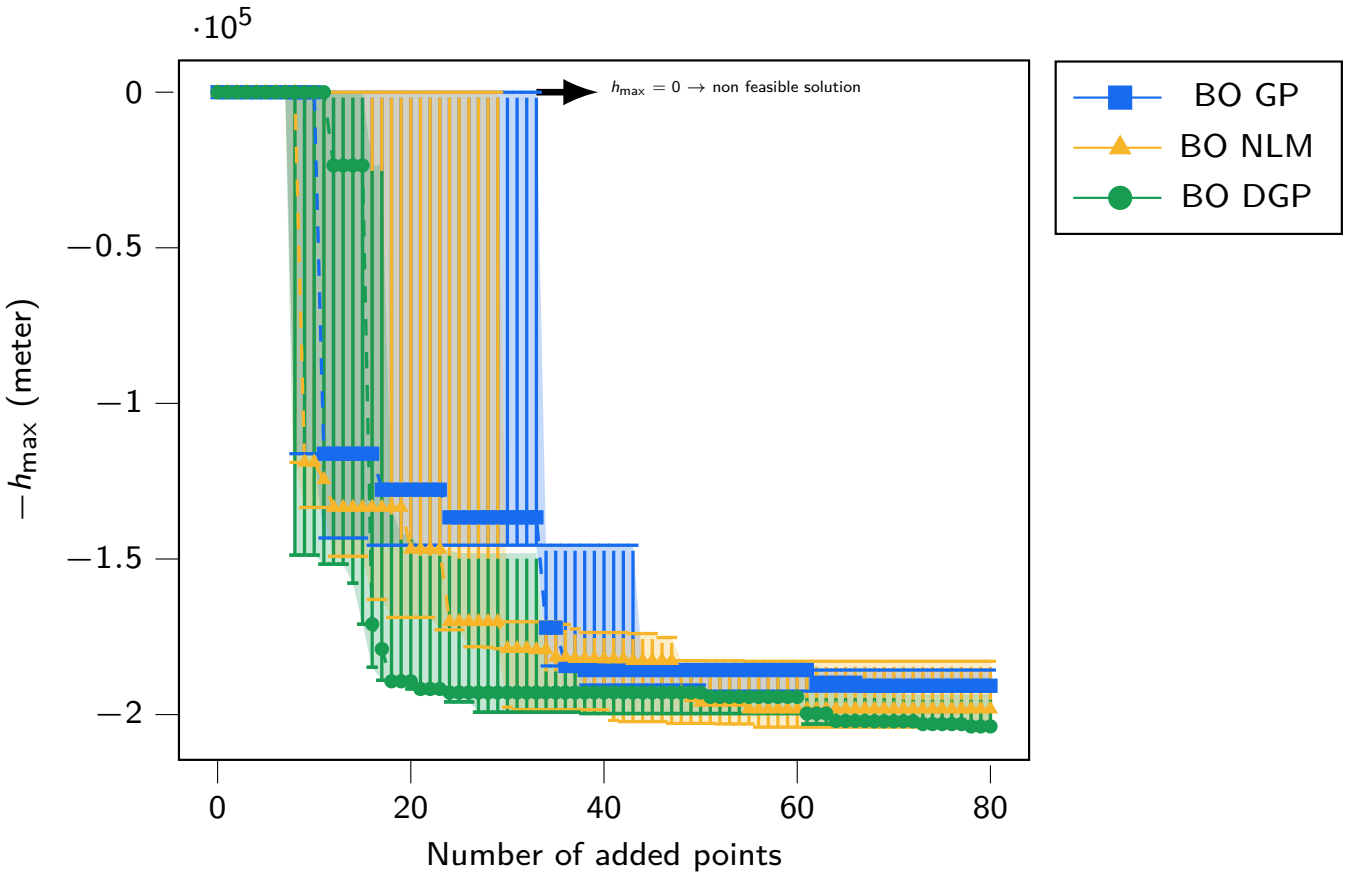
The maximal axial load factor w.r.t :

- the combustion pressure p_{c_1}
- the propellant mass m_1

Application to industrial test case : design of aerospace vehicle

- Initial DoEs size : 75 data points.
- Initialization using random Latin Hyper-Cube Sampling [Loh et al., 1996].
- 5 repetitions are performed.
- Infill criterion : WB2S [Bartoli et al., 2019] and the Expected Violation [Sasena, 2002].
- Stopping criterion : 80 added points.

Application to industrial test case : design of aerospace vehicle



Convergence curve of the negative altitude $-h_{\max}$ of BO with GP, BO with Bayesian Non-linear mapping and BO with a 2-layer DGP.

Summary of contributions I

Bayesian optimization with deep Gaussian processes

- Proposition of a training approach for deep Gaussian processes to improve uncertainty quantification,
- Proposition of a framework coupling of deep Gaussian processes with Bayesian optimization,
- Application of the framework to aerospace design problems.

Summary of contributions I

Bayesian optimization with deep Gaussian processes

- Proposition of a training approach for deep Gaussian processes to improve uncertainty quantification,
- Proposition of a framework coupling of deep Gaussian processes with Bayesian optimization,
- Application of the framework to aerospace design problems.

Summary of contributions I

Bayesian optimization with deep Gaussian processes

- Proposition of a training approach for deep Gaussian processes to improve uncertainty quantification,
- Proposition of a framework coupling of deep Gaussian processes with Bayesian optimization,
- Application of the framework to aerospace design problems.

Summary of contributions I

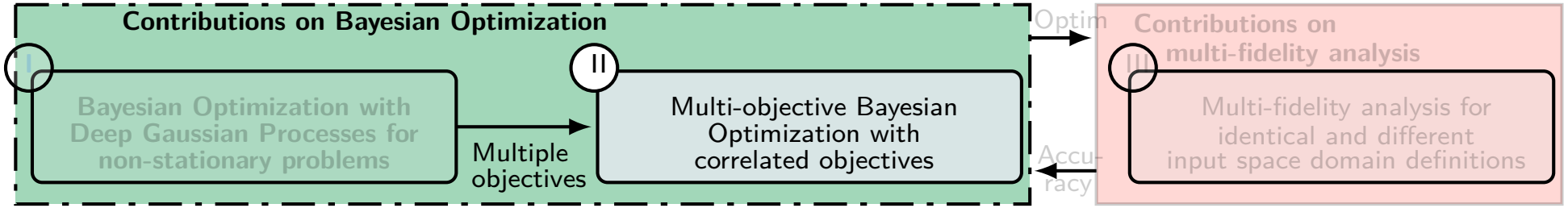
Bayesian optimization with deep Gaussian processes

- Proposition of a training approach for deep Gaussian processes to improve uncertainty quantification,
- Proposition of a framework coupling of deep Gaussian processes with Bayesian optimization,
- Application of the framework to aerospace design problems.

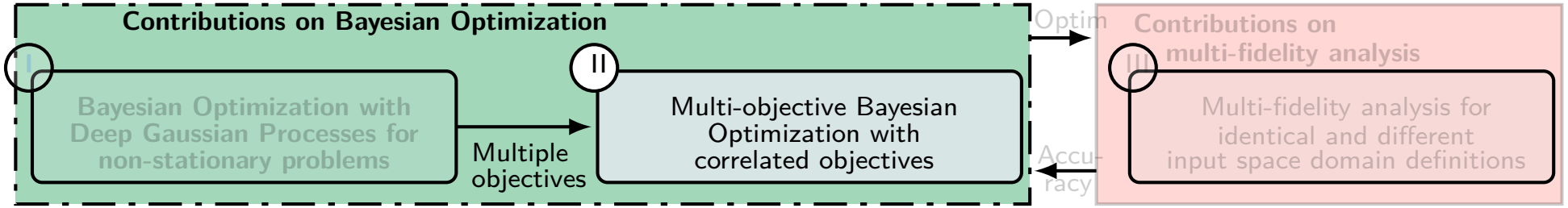
Hebbal, A., Brevault, L., Balesdent, M., Talbi, E. G., & Melab, N. Bayesian optimization using deep Gaussian processes with applications to aerospace system design. In *Optimization and Engineering*, 1-41. Springer, 2020.

Hebbal, A., Brevault, L., Balesdent, M., Taibi, E. G., & Melab, N. Efficient global optimization using deep Gaussian processes. In *IEEE Congress on evolutionary computation (CEC)* 2018 (pp. 1-8). Rio de Janeiro, Brazil.

Contribution II

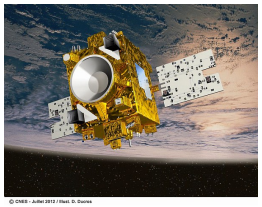


Contribution II

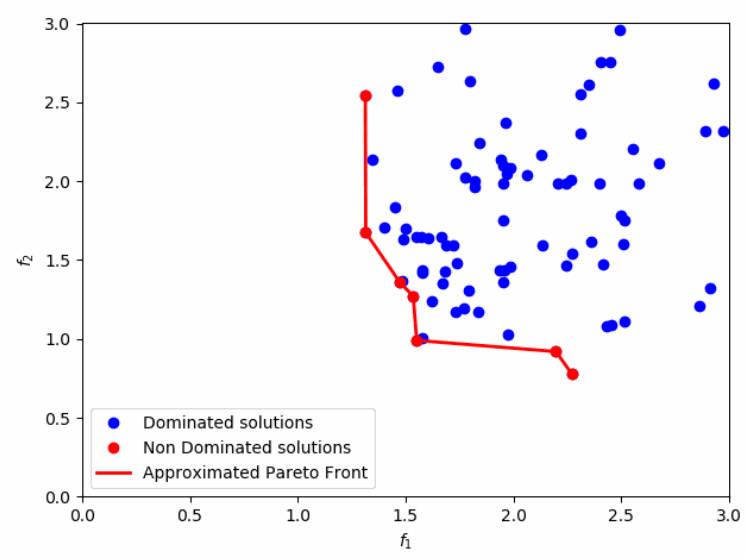


$$
(\mathcal{P}_{CMO}) \left| \begin{array}{lll} \underset{\mathbf{x}}{\text{Minimize}} & \mathbf{y} & = \mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_{n_o}(\mathbf{x})]^\intercal \\ \text{with} & \mathbf{x} & \in \mathcal{X} \subset \mathbb{R}^d \\ \text{subject to} & g_j(\mathbf{x}) & \leq 0, \ \forall j \in \{1, \ldots, n_c\} \end{array} \right.
$$

Multi-objective Bayesian optimization with independent objectives

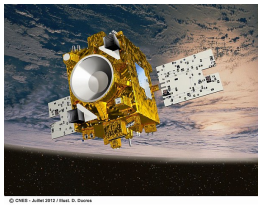


Maximize f_1 : The payload mass.

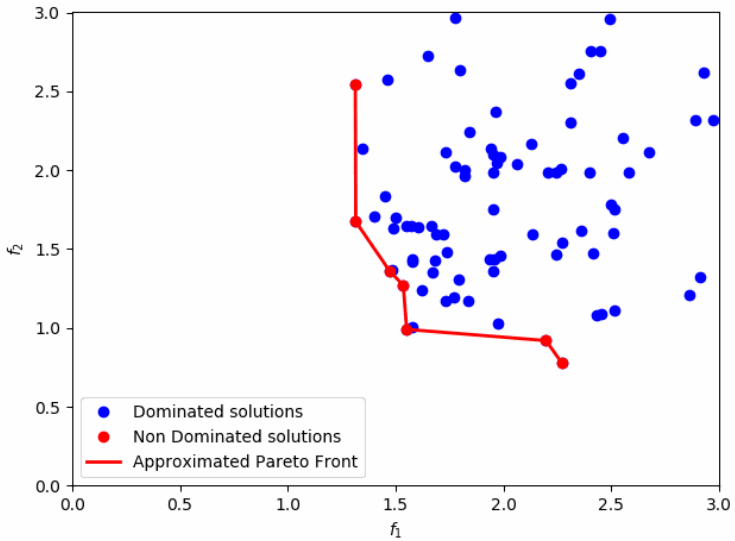


Minimize f_2 : The gross lift off-weight.

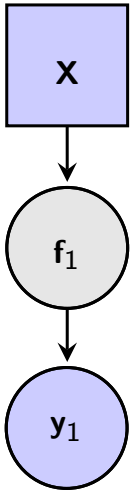
Multi-objective Bayesian optimization with independent objectives

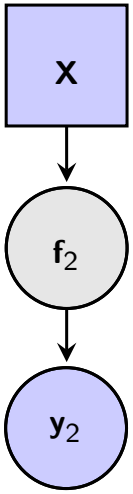


Maximize f_1 : The payload mass.



Minimize f_2 : The gross lift off-weight.

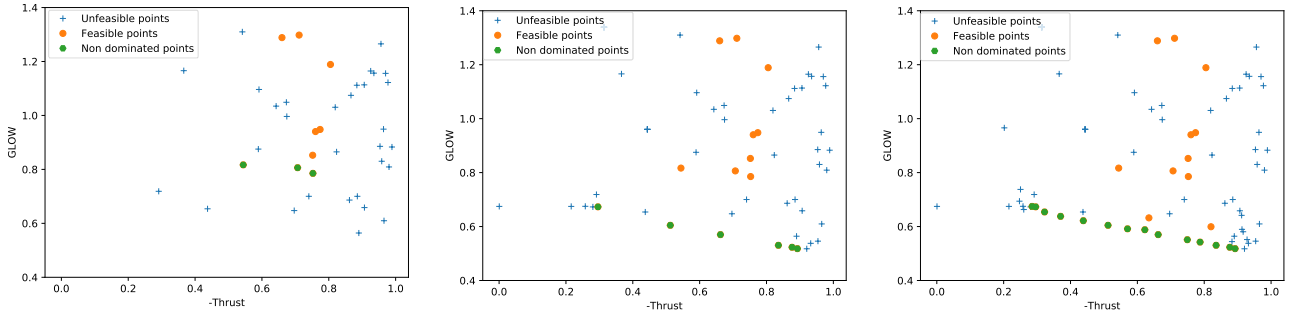


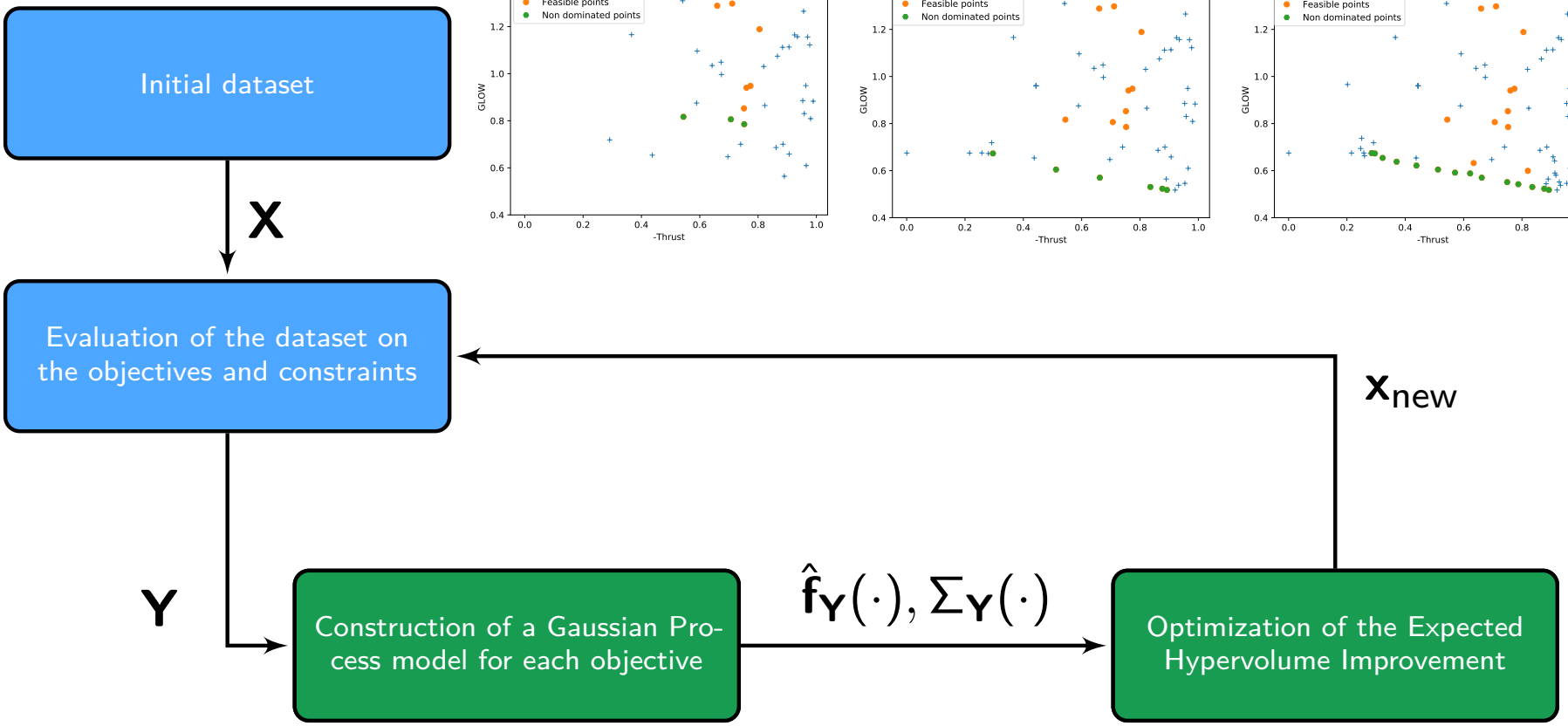


Multi-objective Bayesian optimization with independent objectives

Multi-Objective Bayesian Optimization (MO-BO) [Emmerich et al., 2006]

Initial DoE | 30 added points | 50 added points





Gaussian Processes with linear model of coregionalization

- Classic MO-BO approaches use an independent GP for each objective $\rightarrow$ assumption of independency between the objectives.
- GPs with linear model of coregionalization [Shah and Ghahramani, 2016] : exhibit correlation between functions by introducing a coregionalization matrix K^{coreg} :

$$\mathrm{Cov}\left(f_i(\mathbf{x}), f_j(\mathbf{x}')\right) = K_{ij}^{coreg} k(\mathbf{x}, \mathbf{x}')$$

- Each function is marginally identically distributed up to a scaling factor.

Gaussian Processes with linear model of coregionalization

- Classic MO-BO approaches use an independent GP for each objective $\rightarrow$ assumption of independency between the objectives.
- GPs with linear model of coregionalization [Shah and Ghahramani, 2016] : exhibit correlation between functions by introducing a coregionalization matrix K^{coreg} :

$$\mathrm{Cov}\left(f_i(\mathbf{x}), f_j(\mathbf{x}')\right) = K_{ij}^{coreg} k(\mathbf{x}, \mathbf{x}')$$

- Each function is marginally identically distributed up to a scaling factor.

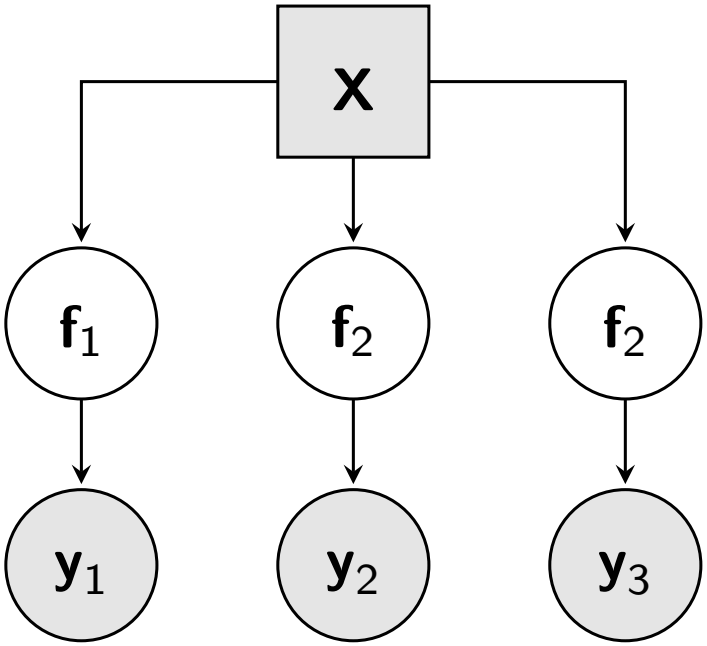
Gaussian Processes with linear model of coregionalization

- Classic MO-BO approaches use an independent GP for each objective $\rightarrow$ assumption of independency between the objectives.
- GPs with linear model of coregionalization [Shah and Ghahramani, 2016] : exhibit correlation between functions by introducing a coregionalization matrix K^{coreg} :

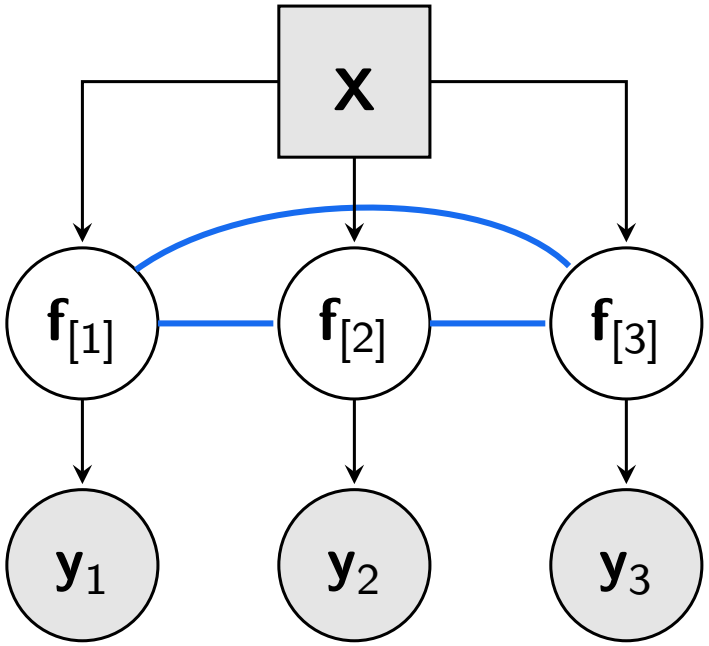
$$\mathrm{Cov}\left(f_i(\mathbf{x}), f_j(\mathbf{x}')\right) = K_{ij}^{coreg} k(\mathbf{x}, \mathbf{x}')$$

- Each function is marginally identically distributed up to a scaling factor.

MO-DGP model



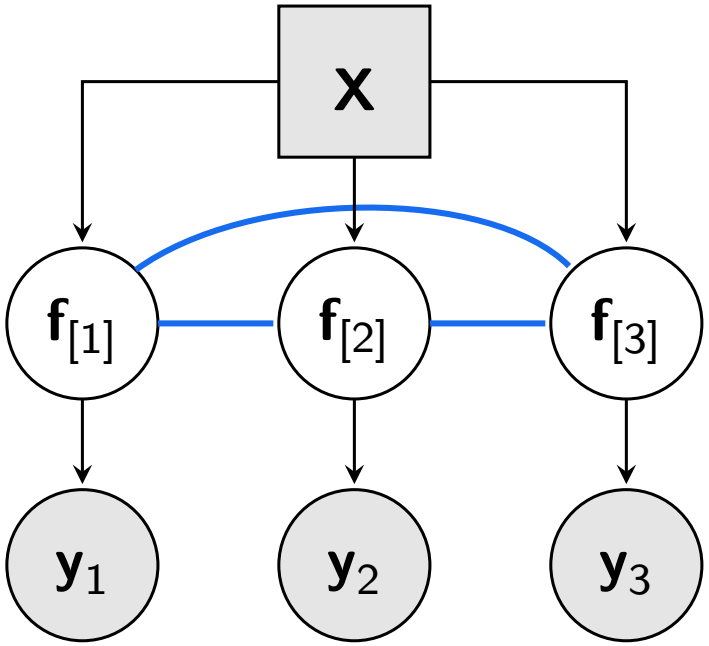
Existing approach : A model for each objective is trained independently.



Proposed approach : The objectives are jointly modeled.

MO-DGP model

- A covariance function to capture the correlations in the objective space as well as in the original input space.



Proposed approach : The objectives are jointly modeled.

MO-DGP inference

$$p\left(\{\mathbf{f}_{[i]}\}_{i=1}^{n_o}\right) \neq \prod_{i=1}^{n_o} p\left(\mathbf{f}_{[i]}|\mathbf{f}_{[i-1]}\right)$$

The following variational approximation is considered :

$$q\left(\{\mathbf{f}_{[i]}\}_{i=1}^{n_o}, \{\mathbf{u}_{[i]}\}_{i=1}^{n_o}\right) = p(\{\mathbf{f}_{[i]}\}_{i=1}^{n_o}|\{\mathbf{u}_{[i]}\}_{i=1}^{n_o}, [\mathbf{X}, \mathbf{f}_{[-i]}], \mathbf{Z}_{[i]}) \times \prod_{i=1}^{n_o} q(\mathbf{u}_{[i]})$$

The obtained Evidence Lower Bound :

$$\mathcal{L}_{\text{MO-DGP}} = \sum_{i=1}^{n_o}\sum_{j=1}^{n} \mathbb{E}_{q(\{f_{[k]}^{(j)}\}_{k=1}^{n_o})} \log\left(p\left(y_i^{(j)}|f_{[i]}^{(j)}\right)\right) - \sum_{i=1}^{n_o} \mathbb{KL}\left[q\left(\mathbf{u}_{[i]}\right)||p(\mathbf{u}_{[i]}|\mathbf{Z}_{[i]})\right]$$

A **Gibbs sampling** procedure to estimate the expectation term.

Experimentations

Assessment of the performance of MO-DGP :

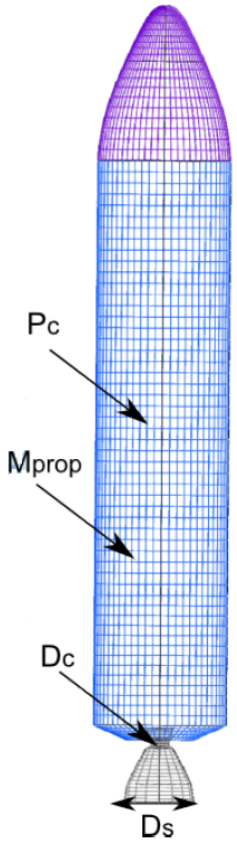
- on a benchmark of analytical test cases,
- on an aerospace design optimization problem.

with respect to state-of-the art approaches.

Experimentations

Problem : Optimization of a set of objectives for a booster engine.
Objectives : Minimization of the Gross Lift-off Weight ($GLOW$) and maximization of the change in velocity (ΔV).

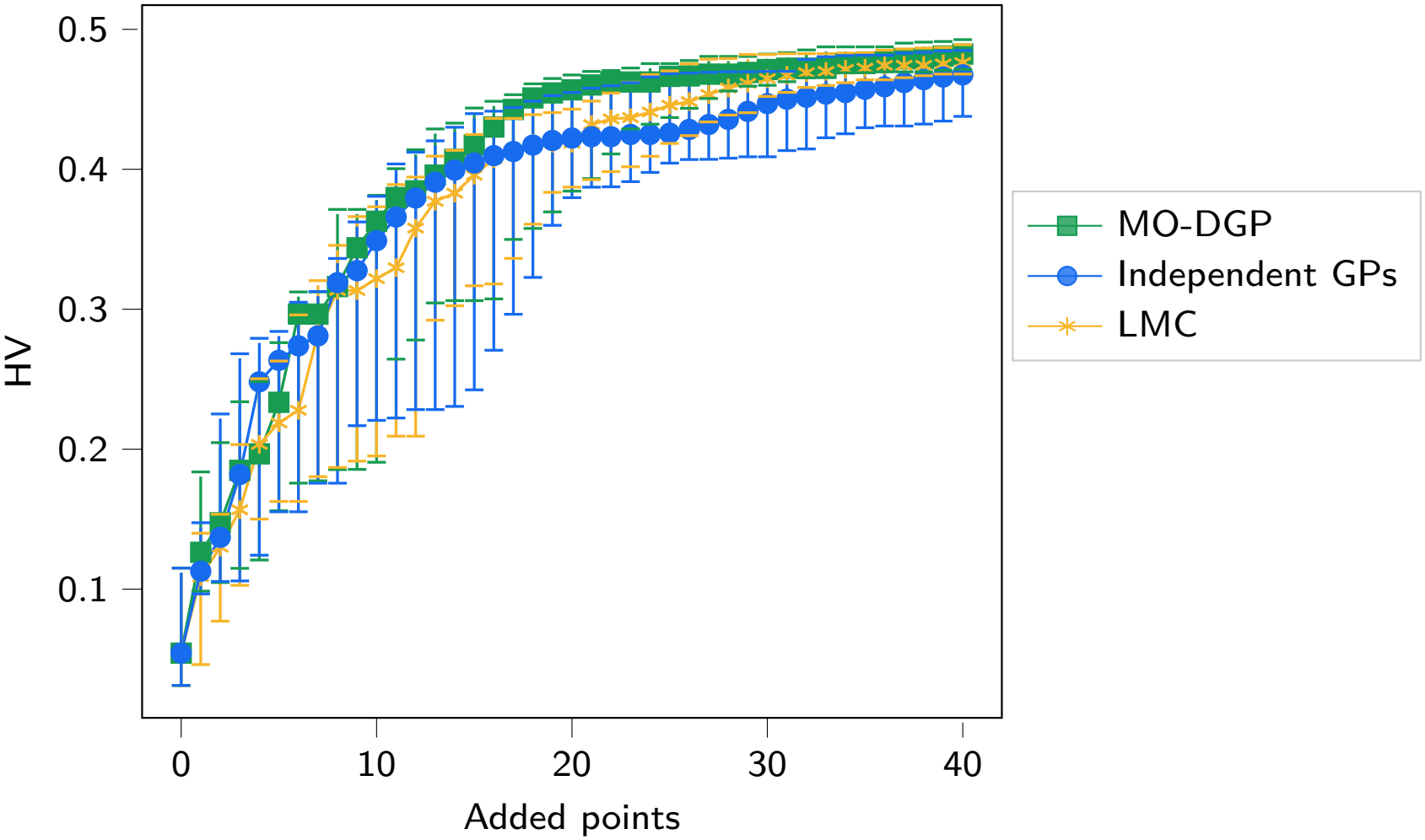
$$
\begin{array}{ll}
\text{Minimize :} & -\Delta V(\mathbf{x}) \\
\text{Minimize :} & GLOW(\mathbf{x}) \\
\text{w.r.t :} & \mathbf{x} = [m_{prop}, p_c, d_c, d_s] \\
\text{s.t :} & \left\{ \begin{array}{l} \text{1 structural constraint} \\ \text{6 geometrical constraints} \\ \text{1 jet breakaway constraint} \end{array} \right.
\end{array}
$$



Application to industrial test case : design of aerospace vehicle

- Initial DoEs size : 20 data points.
- Initialization using random Latin Hyper-Cube Sampling.
- 20 repetitions are performed.
- Infill criterion : EHVI [Emmerich et al., 2006] and the Expected Violation [Sasena, 2002].
- Stopping criterion : 40 added points.

Application to industrial test case : design of aerospace vehicle



Summary of contributions II

Multi-objective Bayesian optimization with correlated objectives

- Development of the multi-objective deep Gaussian process model to encode the potential correlation between objectives,
- Development of a computational approach for the expected hyper-volume improvement with correlated objectives,
- Application of the developed approaches to an aerospace design problem.

Summary of contributions II

Multi-objective Bayesian optimization with correlated objectives

- Development of the multi-objective deep Gaussian process model to encode the potential correlation between objectives,
- Development of a computational approach for the expected hyper-volume improvement with correlated objectives,
- Application of the developed approaches to an aerospace design problem.

Summary of contributions II

Multi-objective Bayesian optimization with correlated objectives

- Development of the multi-objective deep Gaussian process model to encode the potential correlation between objectives,
- Development of a computational approach for the expected hyper-volume improvement with correlated objectives,
- Application of the developed approaches to an aerospace design problem.

Summary of contributions II

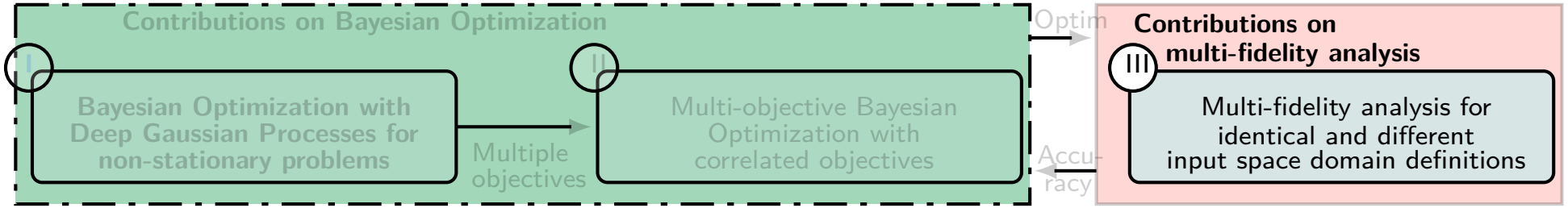
Multi-objective Bayesian optimization with correlated objectives

- Development of the multi-objective deep Gaussian process model to encode the potential correlation between objectives,
- Development of a computational approach for the expected hyper-volume improvement with correlated objectives,
- Application of the developed approaches to an aerospace design problem.

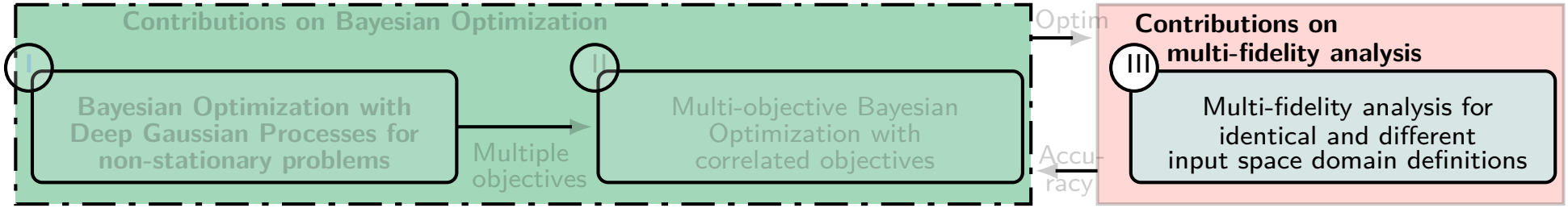
Hebbal, A., Brevault, L., Balesdent, M., Talbi, E., and Melab, N., A Deep Gaussian Process based model for Multi-Objective optimization. The 13th International Conference on Multiple Objective Programming and Goal Programming (MOPGP) 2019.

Hebbal, A., Brevault, L., Balesdent, M., Talbi, E., and Melab, N., Multi-objective optimization using Deep Gaussian Processes : Application to Aerospace Vehicle Design, AIAA Scitech 2019.

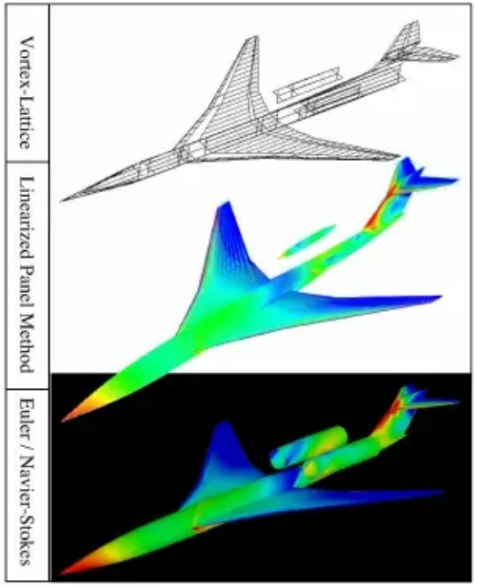
Contribution III



Contribution III



- High-fidelity (HF) models $\rightarrow$ **accurate data** but on a **limited dataset**.
- Low-fidelity (LF) models $\rightarrow$ **large amount of data** of **approximated data**.
- Use of multi-fidelity model to exhibit correlations between datasets of s increasing levels of fidelity $(X^t, \mathbf{y}^t), \forall 1 \leq t \leq s$, to improve prediction accuracy.



[Choi et al., 2004]

Linear Auto-Regressive

Auto-Regressive model (AR1) [Kennedy and O'Hagan, 2001] :

$$f_t(\mathbf{x}) = \rho_{t-1} f_{t-1}(\mathbf{x}) + \gamma_t(\mathbf{x})$$

$f_t(\cdot)$	Fidelity function of level t	$(\mathcal{GP})$
$f_{t-1}(\cdot)$	Fidelity function of level $t-1$	$(\mathcal{GP})$
ρ_{t-1}	Scaling factor of the lower-fidelity	(constant)
$\gamma_t(\cdot)$	Bias between fidelity t and $t-1$	$(\mathcal{GP})$

Linear Auto-Regressive

Auto-Regressive model (AR1) [Kennedy and O'Hagan, 2001] :

$$f_t(\mathbf{x}) = \rho_{t-1} f_{t-1}(\mathbf{x}) + \gamma_t(\mathbf{x})$$

$f_t(\cdot)$	Fidelity function of level t	$(\mathcal{GP})$
$f_{t-1}(\cdot)$	Fidelity function of level $t-1$	$(\mathcal{GP})$
ρ_{t-1}	Scaling factor of the lower-fidelity	(constant)
$\gamma_t(\cdot)$	Bias between fidelity t and $t-1$	$(\mathcal{GP})$

|Multiple variations of the AR1 have been developed [Le Gratiet and Garnier, 2014]

Linear Auto-Regressive

Auto-Regressive model (AR1) [Kennedy and O'Hagan, 2001] :

$$f_t(\mathbf{x}) = \rho_{t-1} f_{t-1}(\mathbf{x}) + \gamma_t(\mathbf{x})$$

$f_t(\cdot)$	Fidelity function of level t	$(\mathcal{GP})$
$f_{t-1}(\cdot)$	Fidelity function of level $t-1$	$(\mathcal{GP})$
ρ_{t-1}	Scaling factor of the lower-fidelity	(constant)
$\gamma_t(\cdot)$	Bias between fidelity t and $t-1$	$(\mathcal{GP})$

| Linear correction with a scaling factor ρ_{t-1} and a bias $\gamma_t(\cdot)$
$\rightarrow$ ✗ captures only linear correlations between fidelities.

Non-Linear Auto-Regressive

Non-Linear Auto-Regressive model (NARGP) [Perdikaris et al., 2017] :

$$f_t(\mathbf{x}) = g_{t-1}(f^*_{t-1}(\mathbf{x}), \mathbf{x})$$

$f_t(\cdot)$	Fidelity function of level t	$(\mathcal{GP})$
$f^*_{t-1}(\cdot)$	mean of the posterior of the $\mathcal{GP}$ $f_{t-1}(\cdot)$	(Real-valued function)
$g_{t-1}(\cdot)$	Non-linear transformation of fidelity $t-1$	$(\mathcal{GP})$

Non-Linear Auto-Regressive

Non-Linear Auto-Regressive model (NARGP) [Perdikaris et al., 2017] :

$$f_t(\mathbf{x}) = g_{t-1}(f^*_{t-1}(\mathbf{x}), \mathbf{x})$$

$f_t(\cdot)$	Fidelity function of level t	$(\mathcal{GP})$
$f^*_{t-1}(\cdot)$	mean of the posterior of the $\mathcal{GP}$ $f_{t-1}(\cdot)$	(Real-valued function)
$g_{t-1}(\cdot)$	Non-linear transformation of fidelity $t-1$	$(\mathcal{GP})$

- ✓ Captures non-linearity using $g_{t-1}(\cdot)$

Non-Linear Auto-Regressive

Non-Linear Auto-Regressive model (NARGP) [Perdikaris et al., 2017] :

$$f_t(\mathbf{x}) = g_{t-1}(f^*_{t-1}(\mathbf{x}), \mathbf{x})$$

$f_t(\cdot)$	Fidelity function of level t	$(\mathcal{GP})$
$f^*_{t-1}(\cdot)$	mean of the posterior of the $\mathcal{GP}$ $f_{t-1}(\cdot)$	(Real-valued function)
$g_{t-1}(\cdot)$	Non-linear transformation of fidelity $t-1$	$(\mathcal{GP})$

- ✗ Disjoint optimization of the fidelity models

Multi-fidelity Deep Gaussian Process

Deep Gaussian process for multi-fidelity
[Marmin and Filippone, 2018, Cutajar et al., 2019] :

$$f_t(\mathbf{x}) = g_{t-1}(f_{t-1}(\mathbf{x}), \mathbf{x})$$

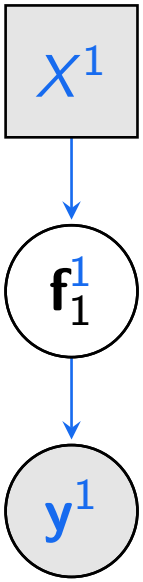
$f_t(\cdot)$	Fidelity function of level t	$(\mathcal{GP})$
$f_{t-1}(\cdot)$	Fidelity function of level $t-1$	$(\mathcal{GP})$
$g_{t-1}(\cdot)$	Non-linear transformation of fidelity $t-1$	$(\mathcal{GP})$

- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes

Multi-fidelity Deep Gaussian Process

$$f_t(\mathbf{x}) = g_{t-1}\big(f_{t-1}(\mathbf{x}), \mathbf{x}\big)$$

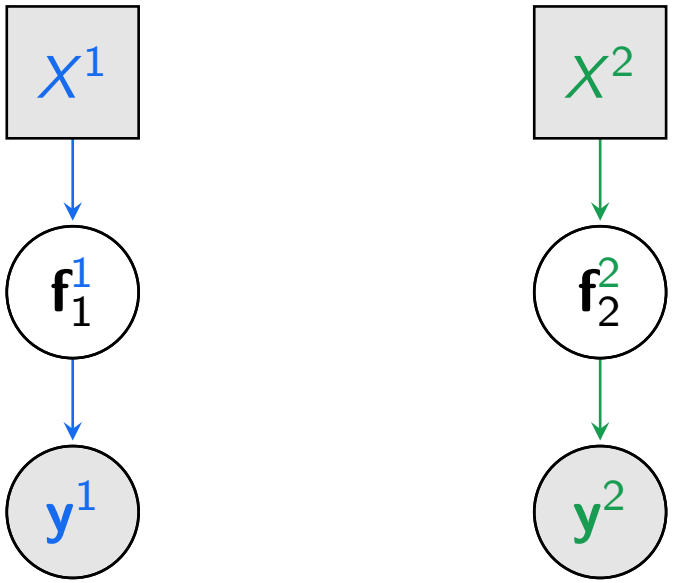
- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes



Multi-fidelity Deep Gaussian Process

$$f_t(\mathbf{x}) = g_{t-1}\big(f_{t-1}(\mathbf{x}), \mathbf{x}\big)$$

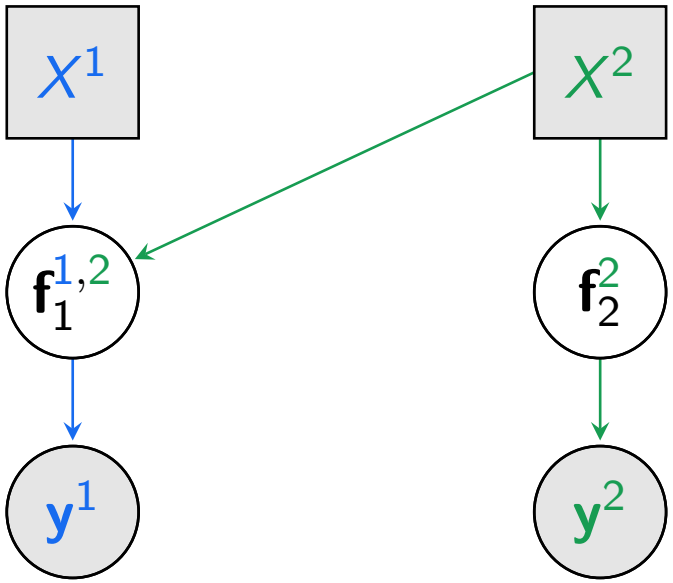
- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes



Multi-fidelity Deep Gaussian Process

$$f_t(\mathbf{x}) = g_{t-1}(f_{t-1}(\mathbf{x}), \mathbf{x})$$

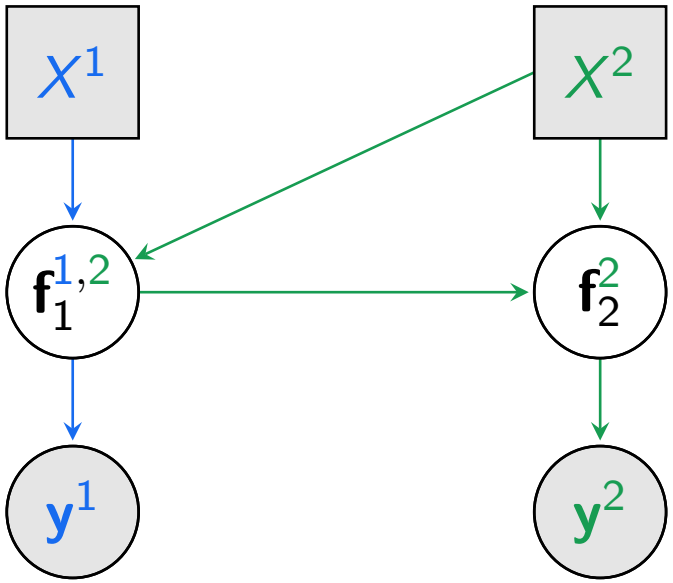
- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes



Multi-fidelity Deep Gaussian Process

$$f_t(\mathbf{x}) = g_{t-1}(f_{t-1}(\mathbf{x}), \mathbf{x})$$

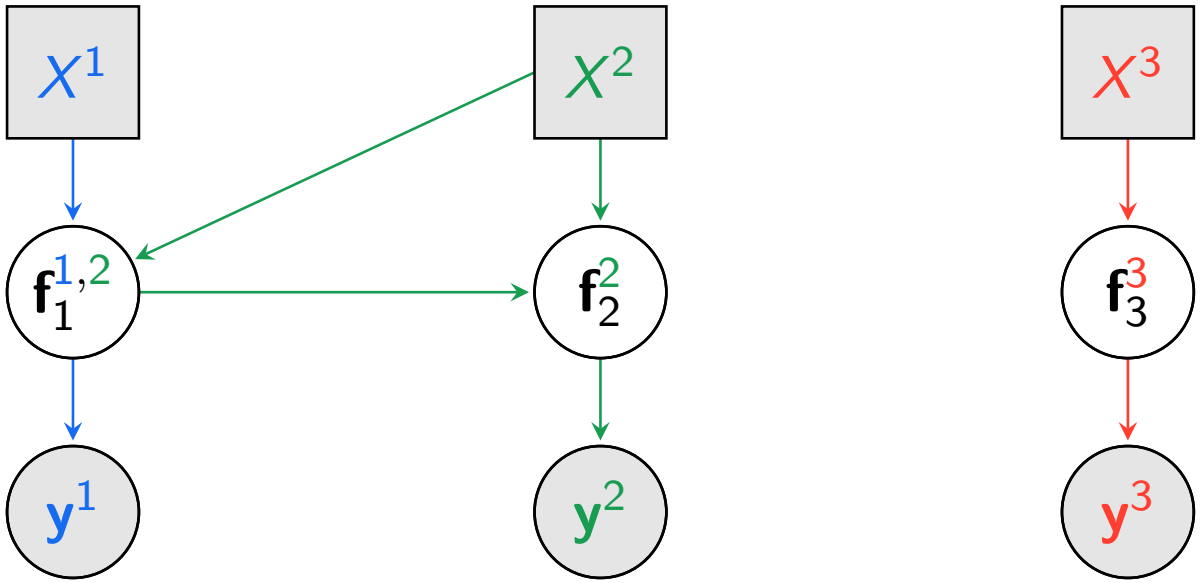
- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes



Multi-fidelity Deep Gaussian Process

$$f_t(\mathbf{x}) = g_{t-1}\big(f_{t-1}(\mathbf{x}), \mathbf{x}\big)$$

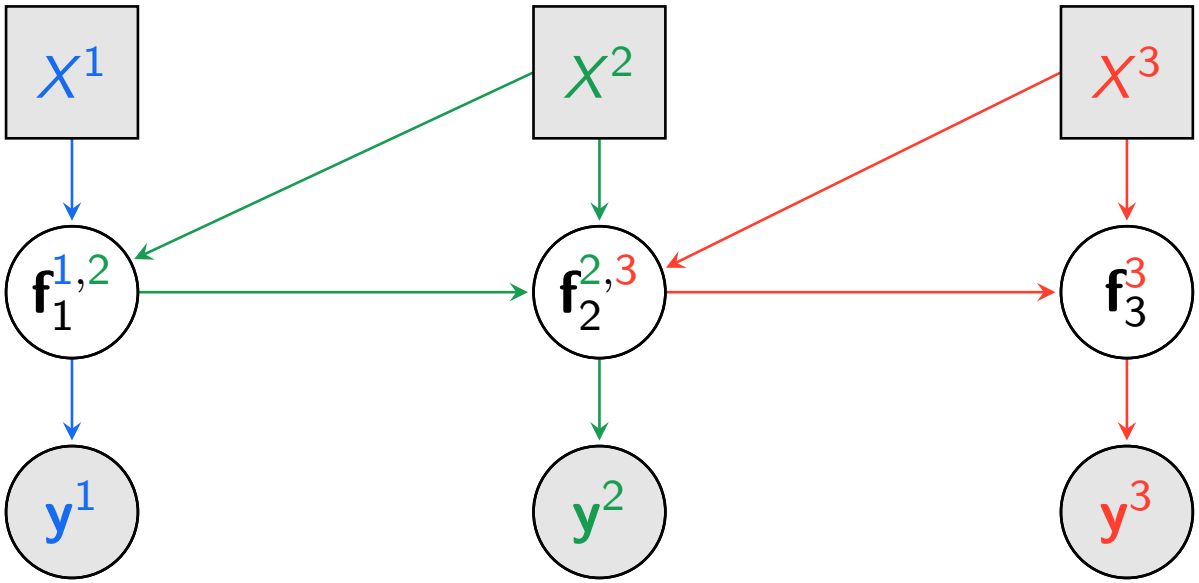
- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes



Multi-fidelity Deep Gaussian Process

$$f_t(\mathbf{x}) = g_{t-1}(f_{t-1}(\mathbf{x}), \mathbf{x})$$

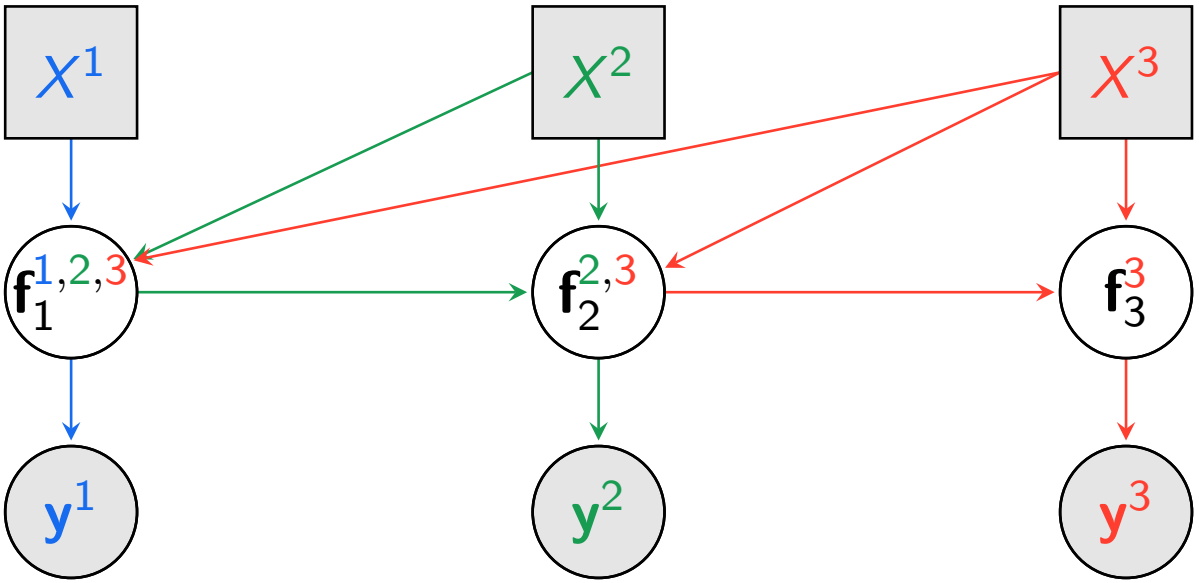
- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes



Multi-fidelity Deep Gaussian Process

$$f_t(\mathbf{x}) = g_{t-1}(f_{t-1}(\mathbf{x}), \mathbf{x})$$

- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes

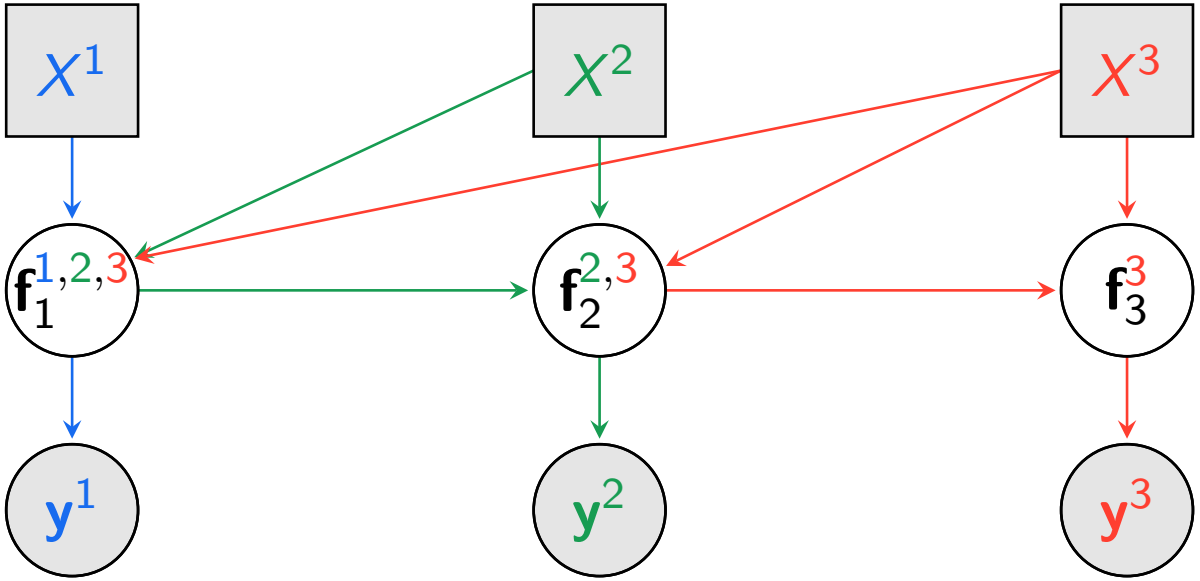


Multi-fidelity Deep Gaussian Process

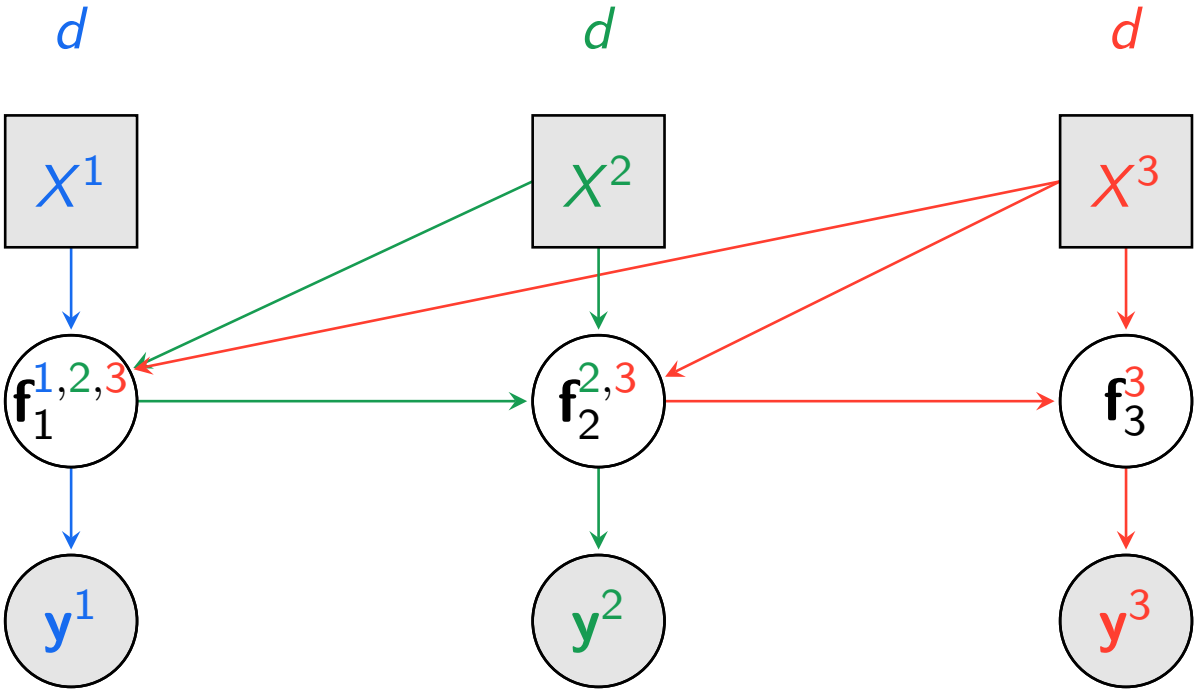
- Keeping the GP prior $f_{t-1}(\cdot)$ implies a composition of GPs $\rightarrow$ deep Gaussian processes

✓ Joint optimization of the different fidelities. $p\left(\mathbf{y}^1, \ldots, \mathbf{y}^s | X^1, \ldots, X^s\right)$

✓ Well-calibrated uncertainty quantification.



Multi-fidelity Deep Gaussian Process

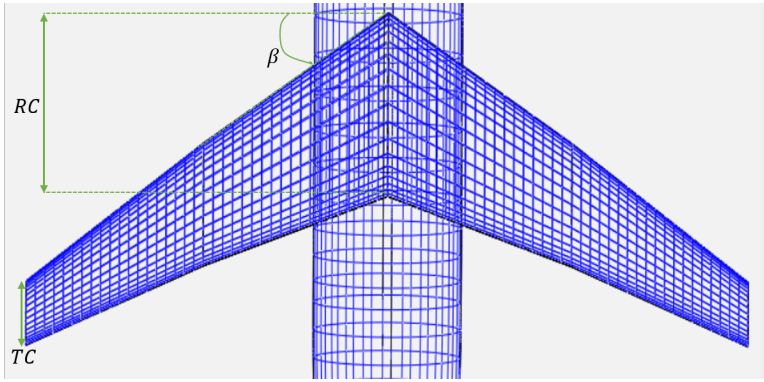


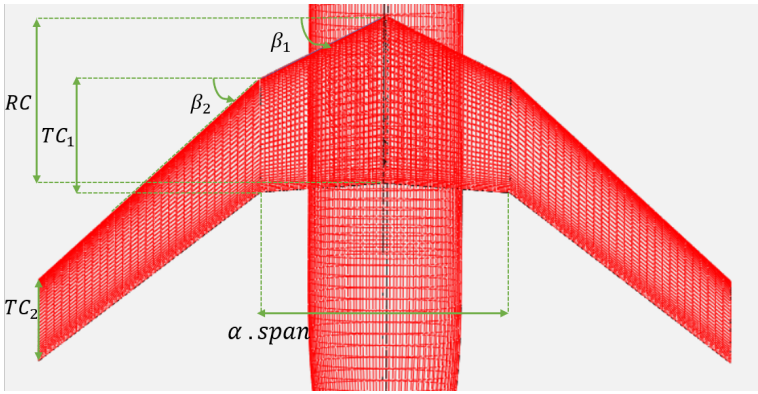
Multi-fidelity with varying input spaces

Multi-fidelity with varying input spaces

In some industrial problems each fidelity can be characterized by its own input space due to :

- Different modeling approaches in each fidelity,
- Omission of some variables in the lower fidelities.





A one-section wing characterized by 3 design variables (left) can be used as a low-fidelity model of a two-sections wing characterized by 6 design variables (right).

Nominal mapping

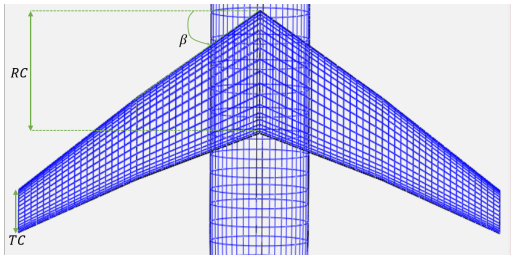
A nominal mapping $g_0(\cdot)$ is an input mapping based on theoretical insight of the multi-fidelity problem at hand [Tao et al., 2019].

Ex : The mapping maps the design variables of the HF into the LF space to obtain an identical defined quantity of interest (*e.g.* the volume defined by the HF variables equals to the volume defined by the mapped HF variables)

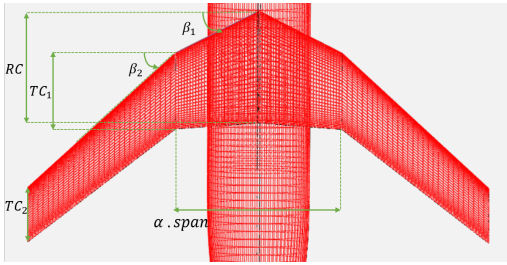
Nominal mapping

A nominal mapping $g_0(\cdot)$ is an input mapping based on theoretical insight of the multi-fidelity problem at hand [Tao et al., 2019].

Ex : The mapping maps the design variables of the HF into the LF space to obtain an identical defined quantity of interest (*e.g.* the volume defined by the HF variables equals to the volume defined by the mapped HF variables)



$$S\left(g_0(\mathbf{x}^{hf})\right) = \qquad S\left(\mathbf{x}^{hf}\right)$$

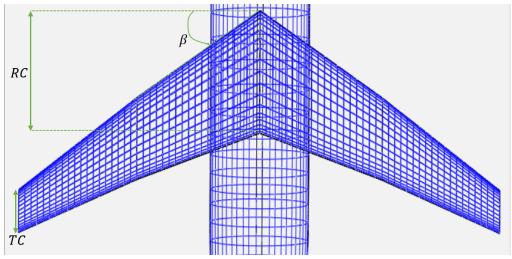


$$\mathbf{x}^{lf} = \begin{bmatrix} R_C \\ T_{C1} + (1-\alpha)TC_2 + (\alpha)RC \\ \alpha\beta_1 + (1-\alpha)\beta_2 \end{bmatrix} \qquad \overset{g_0(\cdot)}{\leftarrow} \qquad \mathbf{x}^{hf} = \begin{bmatrix} R_C \\ T_{C1} \\ T_{C2} \\ \beta_1 \\ \beta_2 \\ \alpha \end{bmatrix}$$

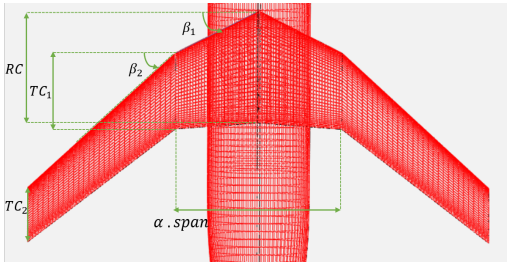
Nominal mapping

A nominal mapping $g_0(\cdot)$ is an input mapping based on theoretical insight of the multi-fidelity problem at hand [Tao et al., 2019].

Ex : The mapping maps the design variables of the HF into the LF space to obtain an identical defined quantity of interest (*e.g.* the volume defined by the HF variables equals to the volume defined by the mapped HF variables)



$$S\left(g_0(\mathbf{x}^{hf})\right) = \qquad S\left(\mathbf{x}^{hf}\right)$$



$$\mathbf{x}^{lf} = \begin{bmatrix} R_C \\ T_{C1} + (1-\alpha)TC_2 + (\alpha)RC \\ \alpha\beta_1 + (1-\alpha)\beta_2 \end{bmatrix} \qquad \xleftarrow{g_0(\cdot)} \qquad \mathbf{x}^{hf} = \begin{bmatrix} R_C \\ T_{C1} \\ T_{C2} \\ \beta_1 \\ \beta_2 \\ \alpha \end{bmatrix}$$

The input mapping may be computationally expensive.

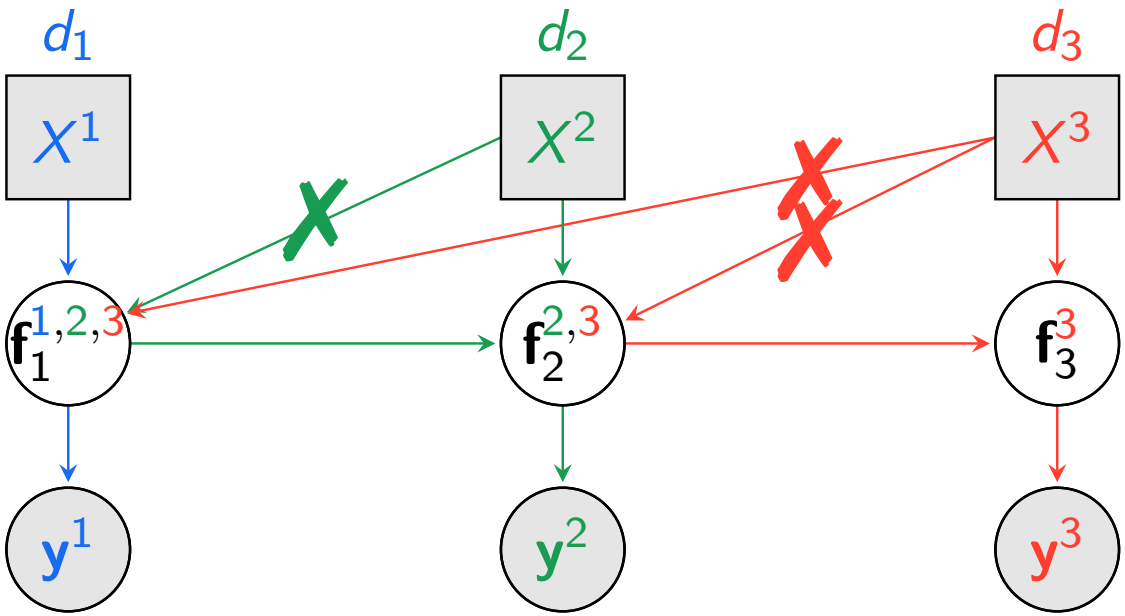
Nominal mapping

A nominal mapping $g_0(\cdot)$ is an input mapping based on theoretical insight of the multi-fidelity problem at hand [Tao et al., 2019].

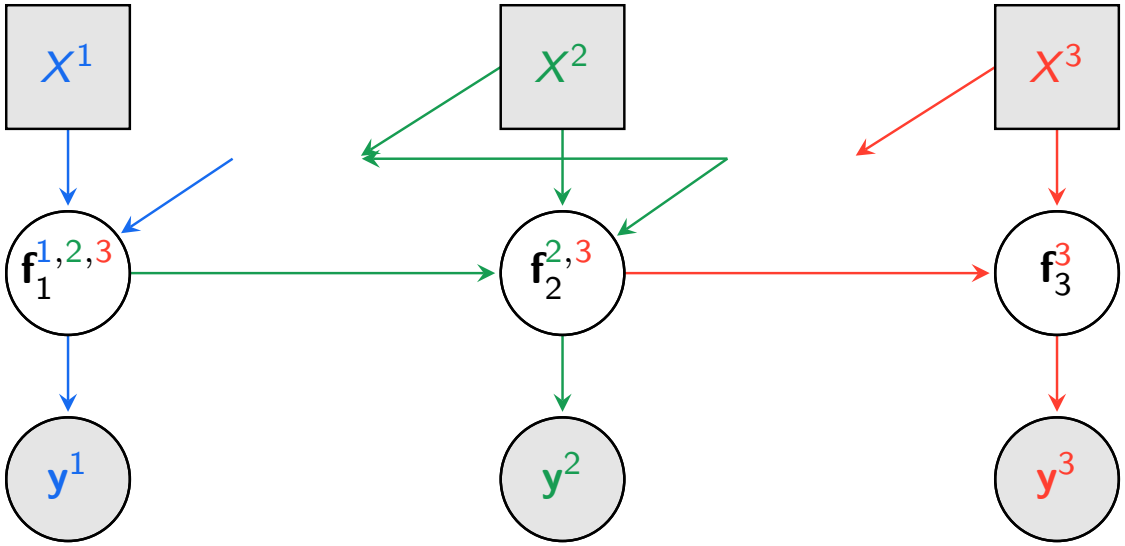
Classic approach : Bias correction with nominal mapping [Li et al., 2016] :

$$f_{hf}(\mathbf{x}^{hf}) = f_{lf}(g_0(\mathbf{x}^{hf})) + \gamma(\mathbf{x}^{hf})$$

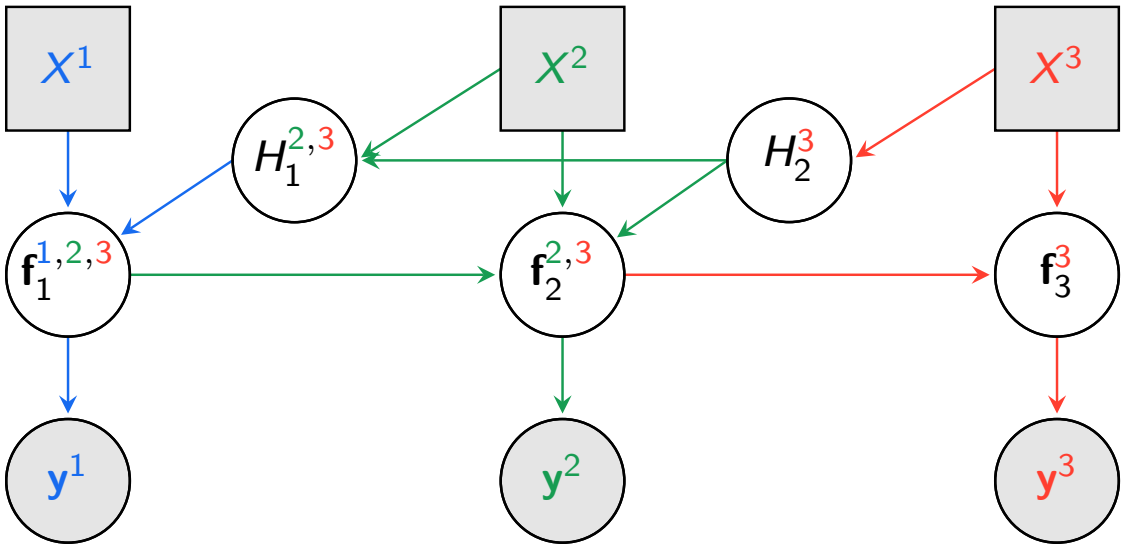
MF-DGP for varying input space parametrizations



Proposed model : A two-level DGP

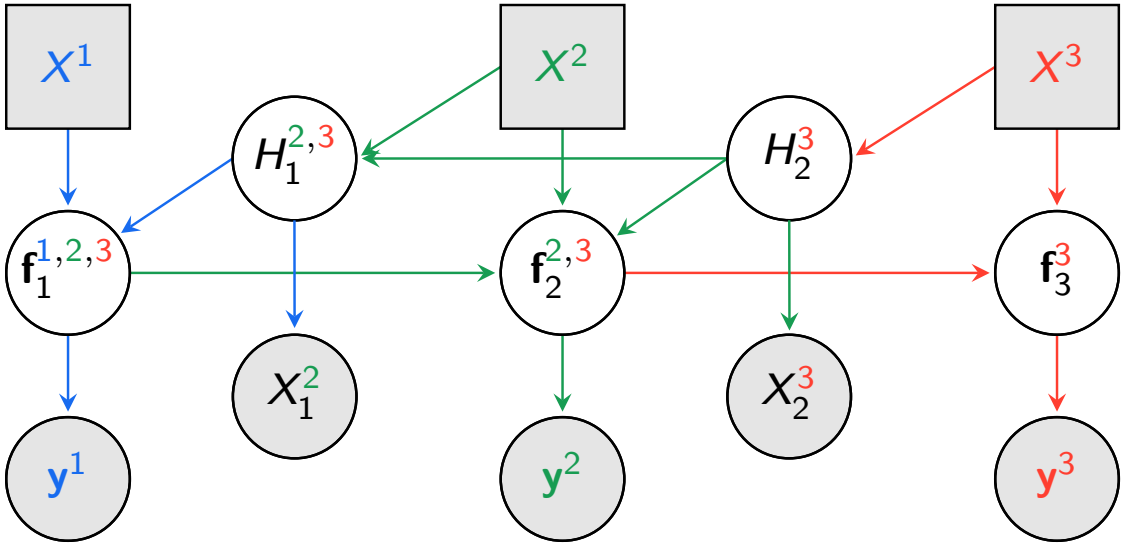


Proposed model : A two-level DGP



- Multi-output GPs H_t^{t+1} :$d_{t+1} \rightarrow d_t$,
- Conditioning of the GPs input mapping on the nominal mapped values of the training data X_t^{t+1}.

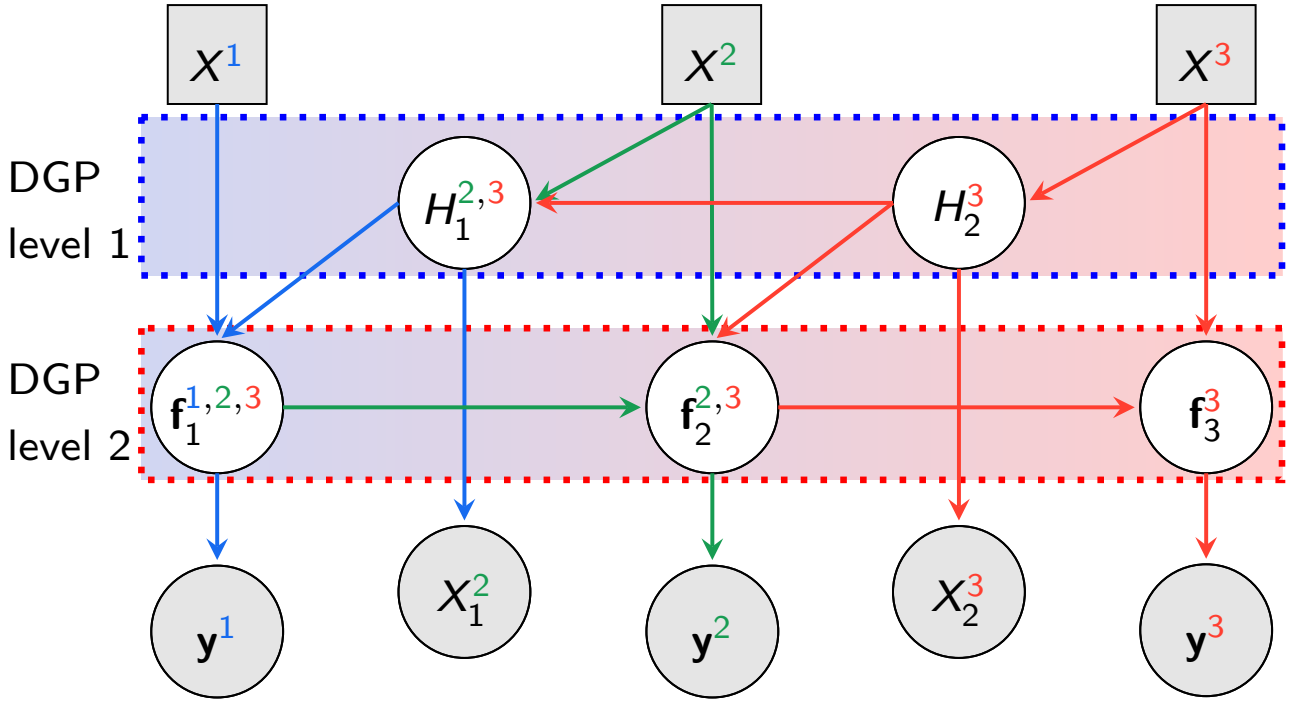
Proposed model : A two-level DGP



MF-DGP Embedded mapping (MF-DGP-EM) graphical representation

- Multi-output GPs H_t^{t+1} :$d_{t+1} \rightarrow d_t$,
- Conditioning of the GPs input mapping on the nominal mapped values of the training data X_t^{t+1}.

Proposed model : A two-level DGP



MF-DGP Embedded mapping (MF-DGP-EM) graphical representation

- ✓ Joint optimization of the mapping and the multi-fidelity model $p\left(\mathbf{y}^1, \ldots, \mathbf{y}^s, X_1^2 \ldots, X_{s-1}^s | X^1, \ldots, X^s\right)$,
- ✓ Correlation computed in the original input space of each fidelity,
- ✓ Adapted to computationally expensive nominal input mapping,
- ✓ Uncertainty quantification on the input space mapping.

Experimentations

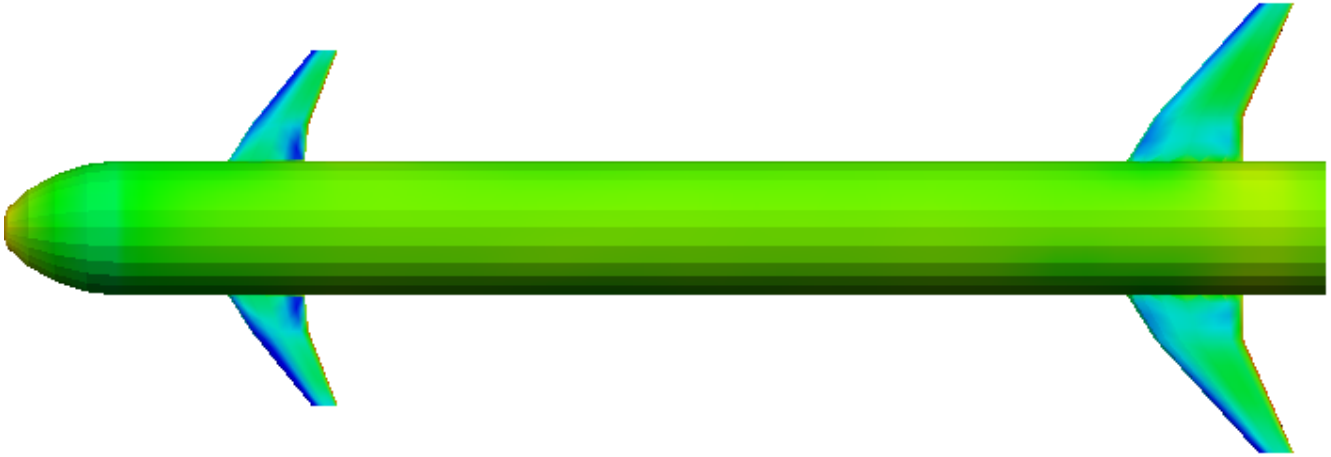
Assessment of the performance of MO-DGP :

- on a benchmark of analytical test cases,
- on a structural problem,
- on an aerospace design problem.

with respect to state-of-the art approaches.

Aerodynamic problem

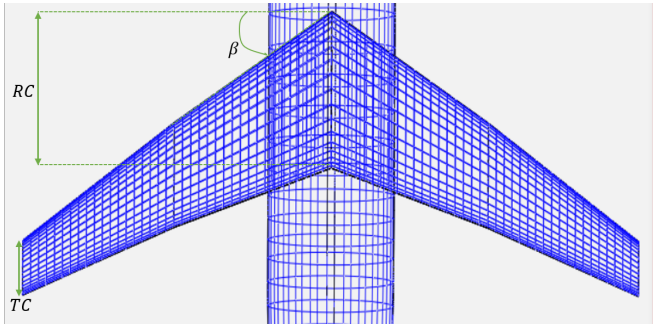
Modeling of the lift coefficient (CL) of a winged reusable launch vehicle composed of a core, two wings, and two canards.



The Vortex lattice method (VLM), is used for the computation of CL using openVSP.

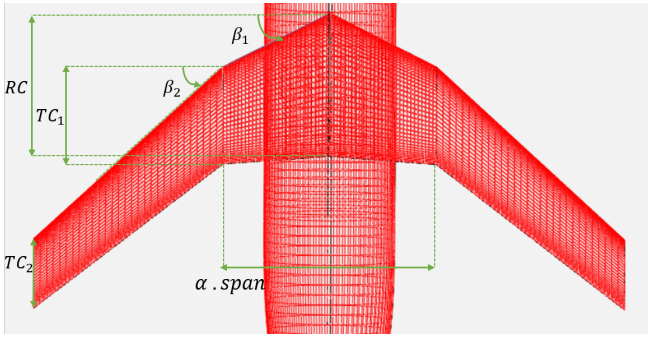
Aerodynamic problem

Low-fidelity :



- The main wing and canards are considered as one-section wings.
- Main wings and canards characterized by 3 variables each, root chord, tip chord, and sweep angle. (**6 variables**)

High-fidelity :

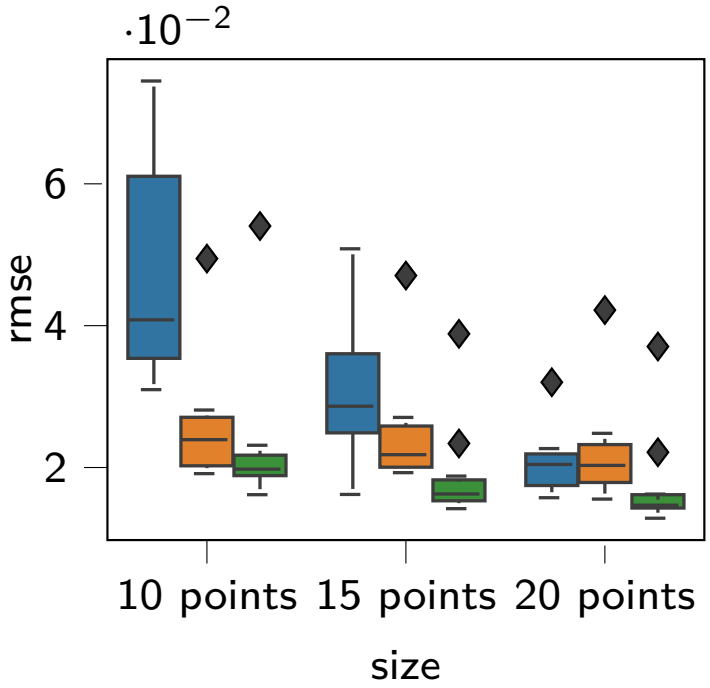


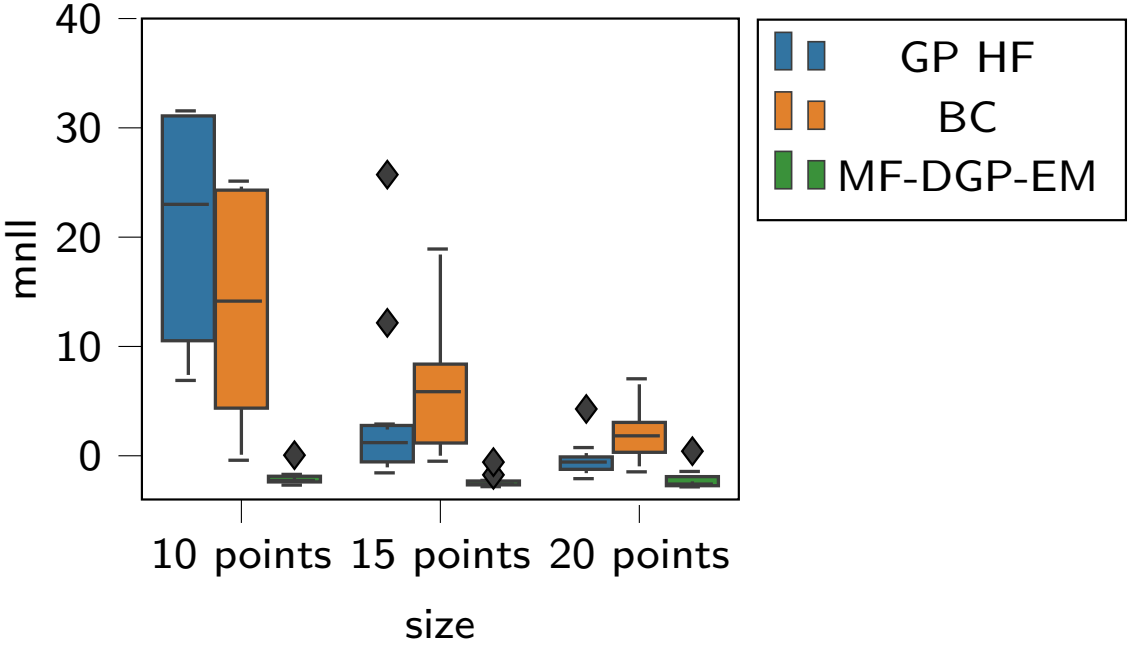
- The main wing and canards are two-section wings.
- Main wings and canards characterized by 6 variables each, root chord, tip chord 1, tip chord 2, sweep angle 1, sweep angle 2, relative span of the first section. (**12 variables**)
- **Meshes densified**

Aerodynamic problem

- Low-fidelity data-size : 120 data points.
- Different HF DoE sizes are considered : 10, 15, and 20 data points.
- Initialization using random Latin Hyper-Cube Sampling.
- 10 repetitions are performed.

Aerodynamic Problem





- With a DoE size of only 10 points for HF, the MF-DGP-EM already obtains a robust and efficient result in prediction accuracy and uncertainty quantification

Summary of contributions III

Multi-fidelity analysis

- Identically defined design spaces
  - Development of a training approach for the inducing points in MF-DGP,
  - Benchmark of Gaussian process-based multi-fidelity approaches on aerospace test cases.
- Varying design spaces
  - Development of MF-DGP with embedded mapping for multi-fidelity with different fidelity input spaces,
  - Application of the developed approaches to design problems.

Summary of contributions III

Multi-fidelity analysis

- Identically defined design spaces
  - Development of a training approach for the inducing points in MF-DGP,
  - Benchmark of Gaussian process-based multi-fidelity approaches on aerospace test cases.
- Varying design spaces
  - Development of MF-DGP with embedded mapping for multi-fidelity with different fidelity input spaces,
  - Application of the developed approaches to design problems.

Summary of contributions III

Multi-fidelity analysis

- Identically defined design spaces
  - Development of a training approach for the inducing points in MF-DGP,
  - Benchmark of Gaussian process-based multi-fidelity approaches on aerospace test cases.
- Varying design spaces
  - Development of MF-DGP with embedded mapping for multi-fidelity with different fidelity input spaces,
  - Application of the developed approaches to design problems.

Summary of contributions III

Multi-fidelity analysis

- Identically defined design spaces
  - Development of a training approach for the inducing points in MF-DGP,
  - Benchmark of Gaussian process-based multi-fidelity approaches on aerospace test cases.
- Varying design spaces
  - Development of MF-DGP with embedded mapping for multi-fidelity with different fidelity input spaces,
  - Application of the developed approaches to design problems.

Summary of contributions III

Multi-fidelity analysis

- Identically defined design spaces
  - Development of a training approach for the inducing points in MF-DGP,
  - Benchmark of Gaussian process-based multi-fidelity approaches on aerospace test cases.
- Varying design spaces
  - Development of MF-DGP with embedded mapping for multi-fidelity with different fidelity input spaces,
  - Application of the developed approaches to design problems.

Summary of contributions III

Multi-fidelity analysis

- Identically defined design spaces
  - Development of a training approach for the inducing points in MF-DGP,
  - Benchmark of Gaussian process-based multi-fidelity approaches on aerospace test cases.
- Varying design spaces
  - Development of MF-DGP with embedded mapping for multi-fidelity with different fidelity input spaces,
  - Application of the developed approaches to design problems.

Summary of contributions III

Multi-fidelity analysis

- Identically defined design spaces
  - Development of a training approach for the inducing points in MF-DGP,
  - Benchmark of Gaussian process-based multi-fidelity approaches on aerospace test cases.
- Varying design spaces
  - Development of MF-DGP with embedded mapping for multi-fidelity with different fidelity input spaces,
  - Application of the developed approaches to design problems.

Hebbal, A., Brevault, L., Balesdent, M., Talbi, E. G., & Melab, N. Multi-fidelity modeling with different input domain definitions using deep Gaussian processes. In *Structural and Multidisciplinary Optimization Journal*.

Hebbal, A., Brevault, L., Balesdent, M., Talbi, E., & Melab, N. Multi-fidelity modeling using DGPs : Improvements and a generalization to varying input space dimensions. *4th workshop on Bayesian Deep Learning (NeurIPS 2019)*, Vancouver, Canada.

Table of Contents

Introduction

Background on Gaussian processes

Bayesian optimization for non-stationary functions

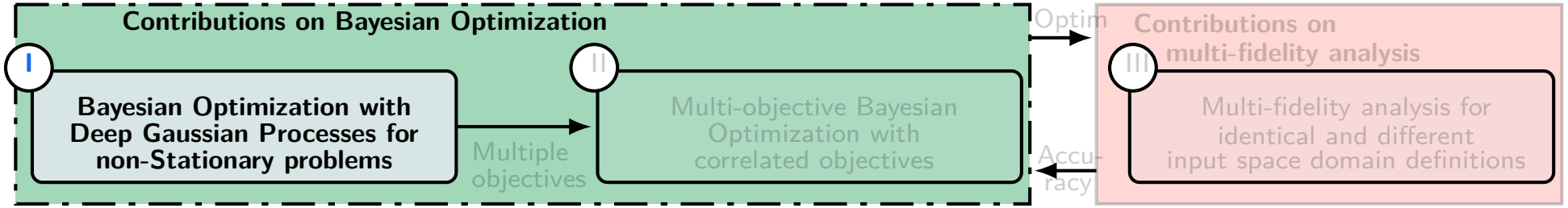
Multi-objective Bayesian optimization

Multi-fidelity with varying input spaces

Conclusions and perspectives

Conclusions

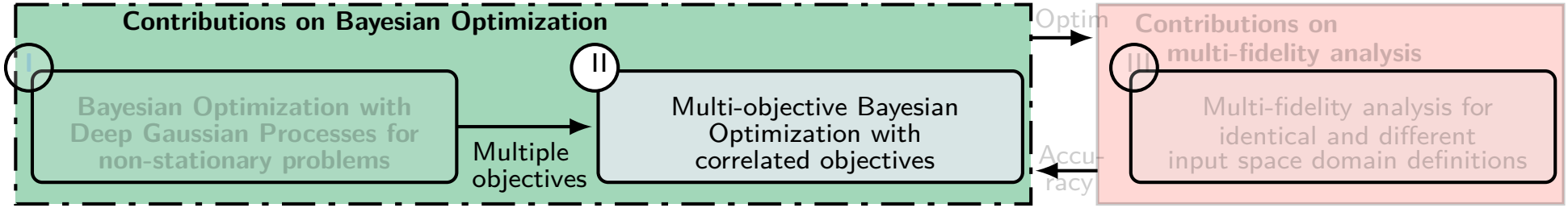
Contributions :



- Coupling of deep Gaussian processes with single and multi-objective Bayesian optimization.
- Proposition of MO-DGP model taking into account the correlations in the objective space.
- Proposition of a MF-DGP model for varying input space multi-fidelity problems.
- Application of the different approaches on aerospace design problems.
- An open source Python implementation of the developed approaches (*release in progress*).

Conclusions

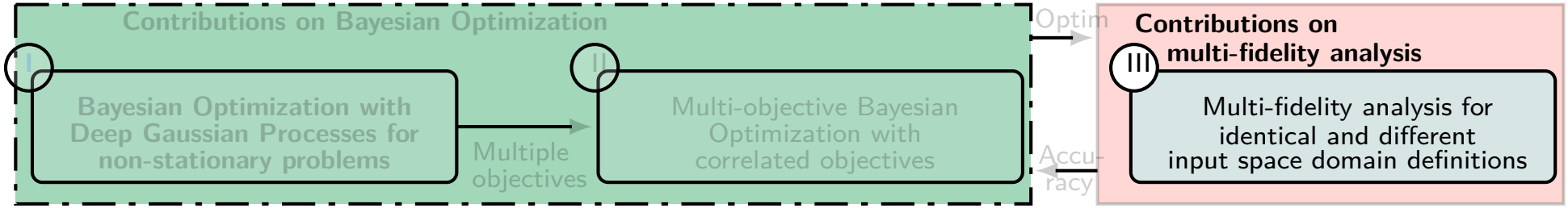
Contributions :



- Coupling of deep Gaussian processes with single and multi-objective Bayesian optimization.
- Proposition of MO-DGP model taking into account the correlations in the objective space.
- Proposition of a MF-DGP model for varying input space multi-fidelity problems.
- Application of the different approaches on aerospace design problems.
- An open source Python implementation of the developed approaches (*release in progress*).

Conclusions

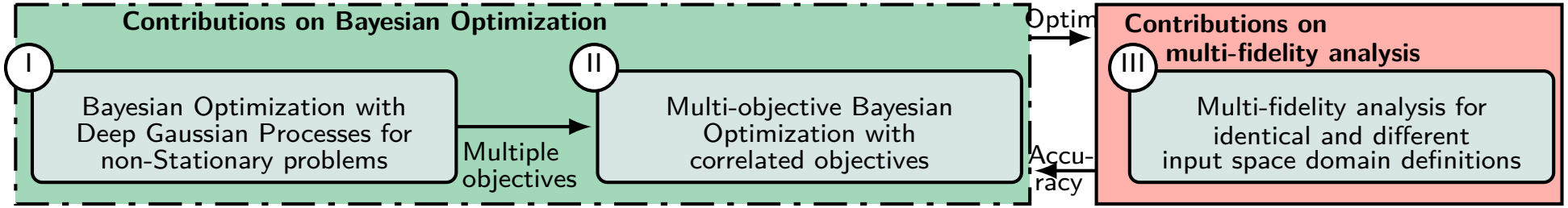
Contributions :



- Coupling of deep Gaussian processes with single and multi-objective Bayesian optimization.
- Proposition of MO-DGP model taking into account the correlations in the objective space.
- Proposition of a MF-DGP model for varying input space multi-fidelity problems.
- Application of the different approaches on aerospace design problems.
- An open source Python implementation of the developed approaches (*release in progress*).

Conclusions

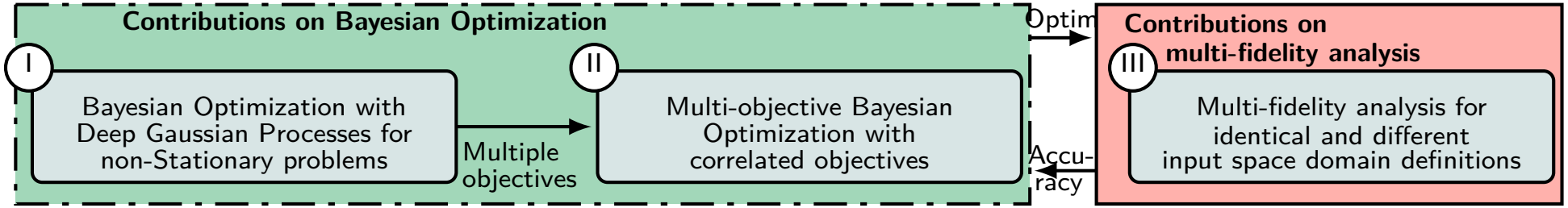
Contributions :



- Coupling of deep Gaussian processes with single and multi-objective Bayesian optimization.
- Proposition of MO-DGP model taking into account the correlations in the objective space.
- Proposition of a MF-DGP model for varying input space multi-fidelity problems.
- Application of the different approaches on aerospace design problems.
- An open source Python implementation of the developed approaches (*release in progress*).

Conclusions

Contributions :



- Coupling of deep Gaussian processes with single and multi-objective Bayesian optimization.
- Proposition of MO-DGP model taking into account the correlations in the objective space.
- Proposition of a MF-DGP model for varying input space multi-fidelity problems.
- Application of the different approaches on aerospace design problems.
- An open source Python implementation of the developed approaches (*release in progress*).

Limits

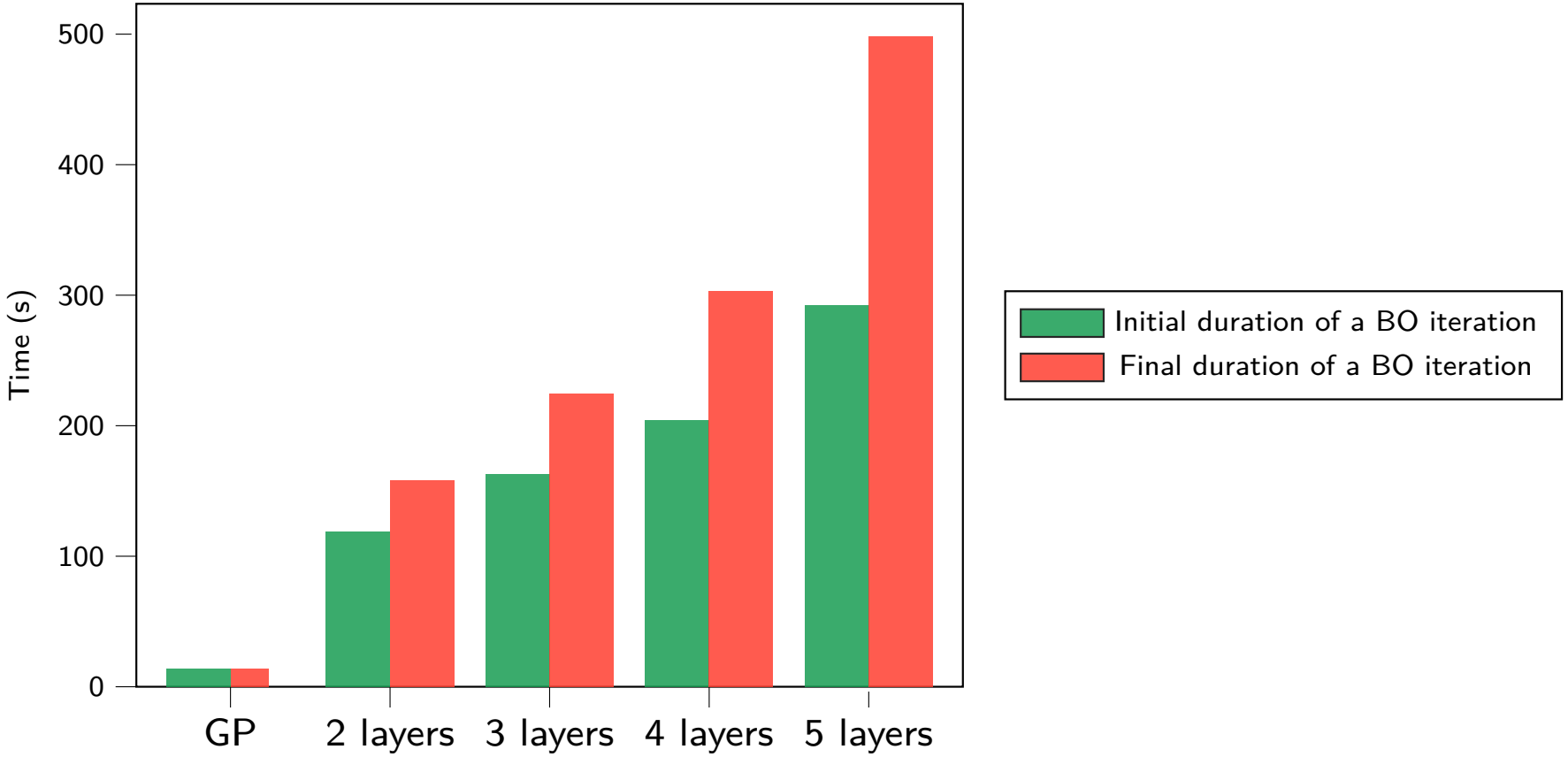
Limits of deep Gaussian process based approaches

- The computational time for training the models.
- The curse of dimension

Limits

Limits of deep Gaussian process based approaches

- The computational time for training the models.
- The curse of dimension



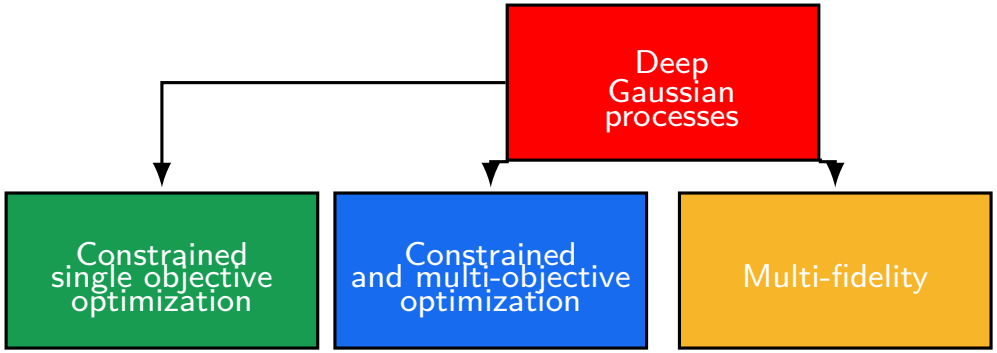
Limits

Limits of deep Gaussian process based approaches

- The computational time for training the models.
- The curse of dimension

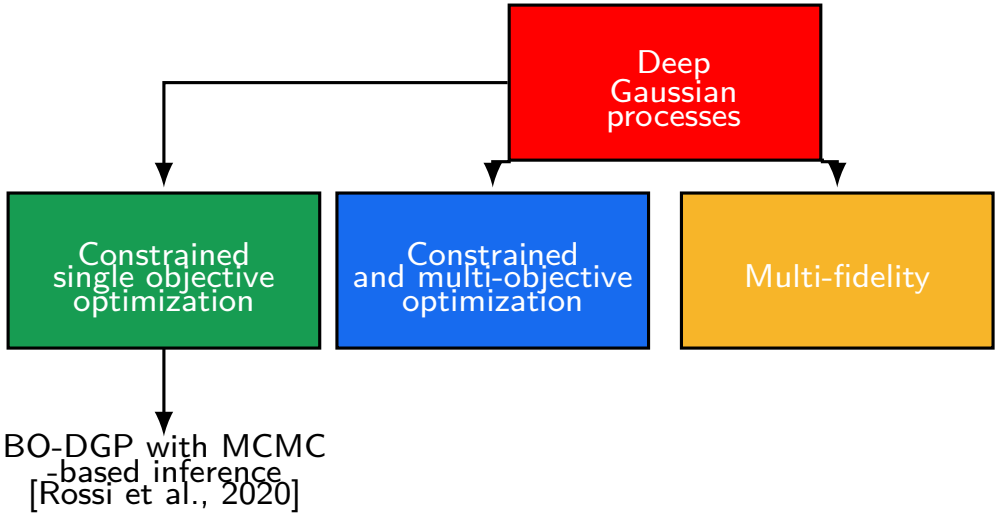
Perspectives

Vertical extensions



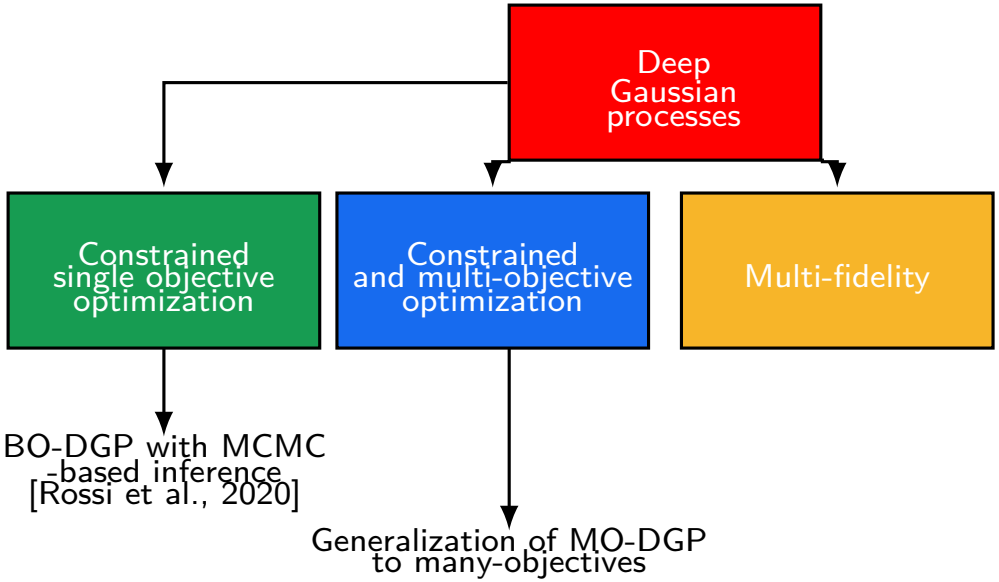
Perspectives

Vertical extensions



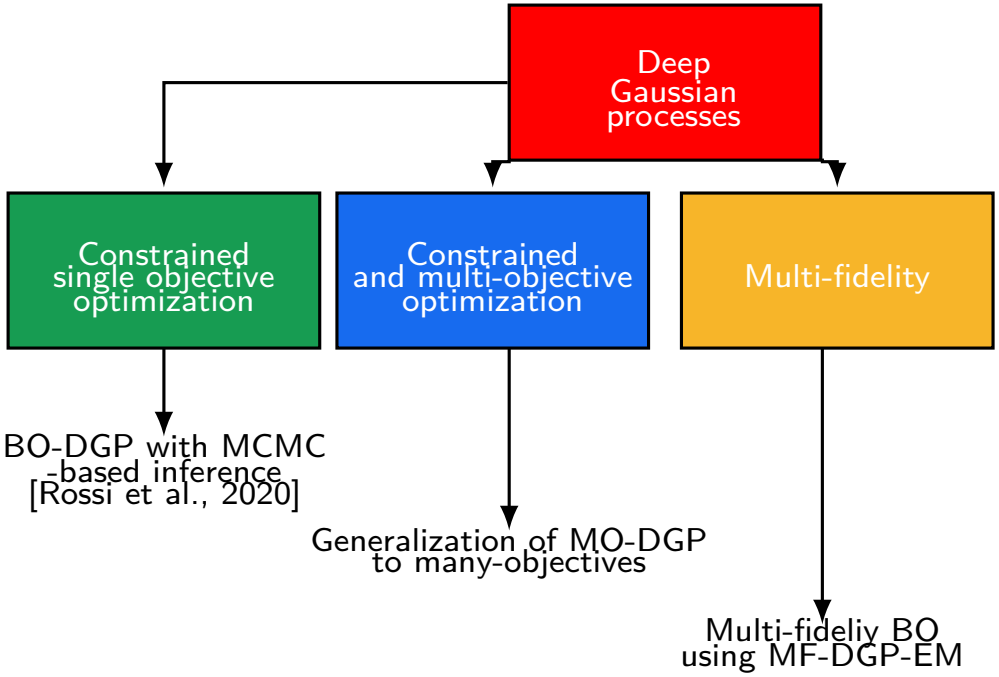
Perspectives

Vertical extensions



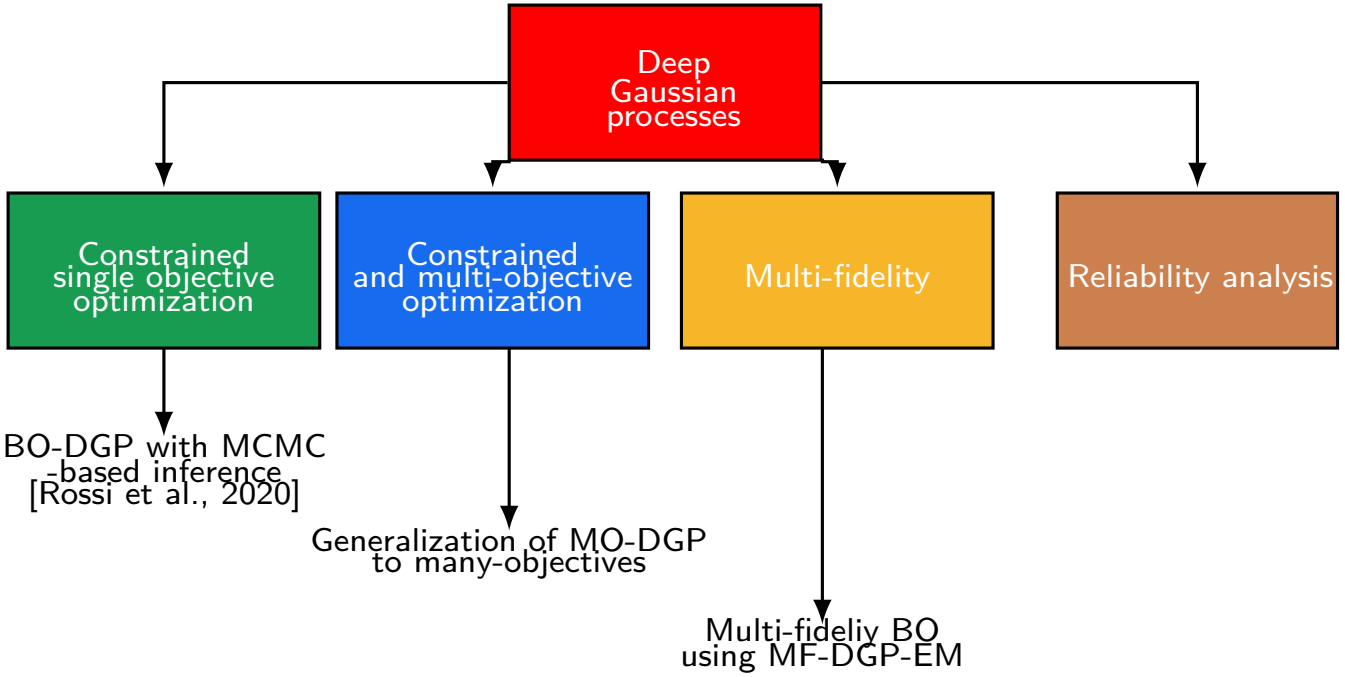
Perspectives

Vertical extensions



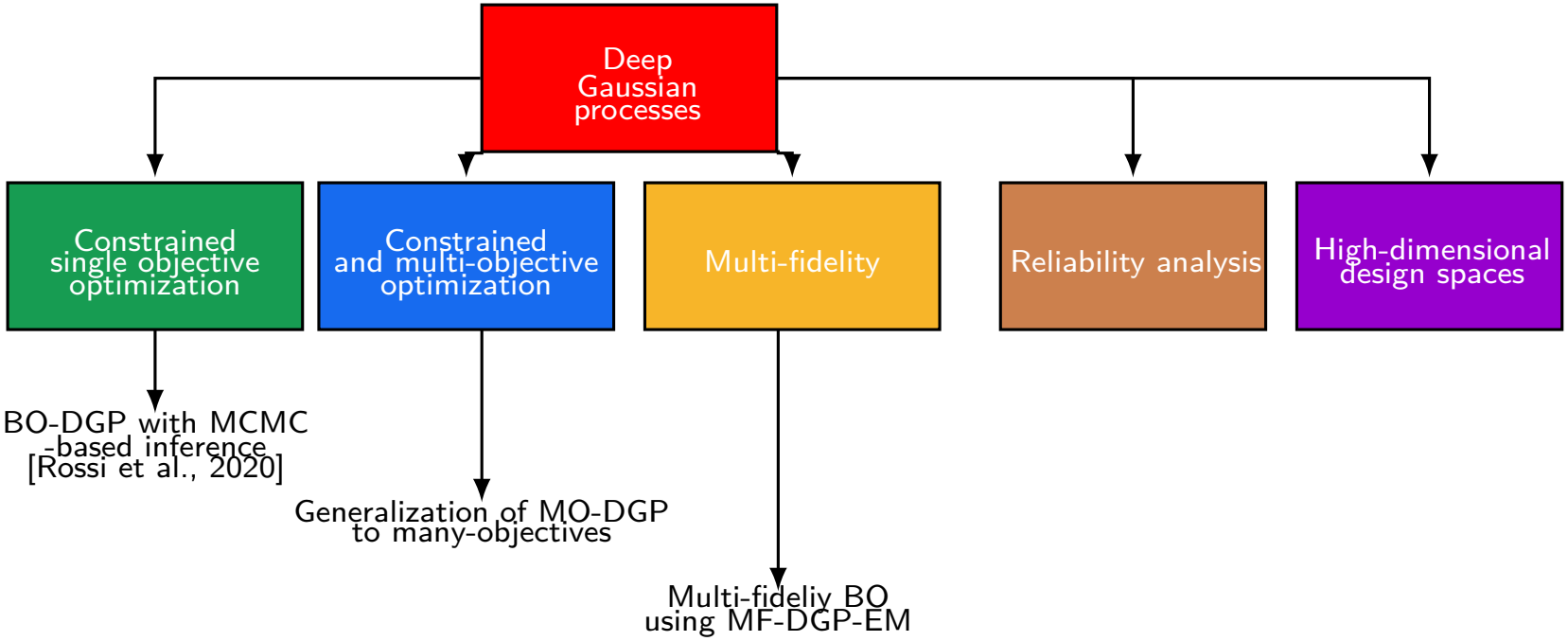
Perspectives

Horizontal extensions



Perspectives

Horizontal extensions



Publications

Journal papers :

- **Hebbal, A.**, Brevault, L., Balesdent, M., Talbi, E. G., & Melab, N. Bayesian optimization using deep Gaussian processes with applications to aerospace system design. In *Optimization and Engineering*, 1-41. Springer, 2020.
- **Hebbal, A.**, Brevault, L., Balesdent, M., Talbi, E. G., & Melab, N. Multi-fidelity modeling with different input domain definitions using deep Gaussian processes. In *Structural and Multidisciplinary Optimization Journal*. (Accepted).
- Brevault, L., Balesdent, M., & **Hebbal, A**. Multi-objective multidisciplinary design optimization approach for partially reusable launch vehicle design. In *Journal of Spacecraft and Rockets*, 57(2), 373-390, 2020.
- Brevault, L., Balesdent, M., & **Hebbal, A**. Overview of Gaussian process based multi-fidelity techniques with variable relationship between fidelities, application to aerospace systems. In *Aerospace Science and Technology*, 106339. Vol.107, Elsevier, 2020.

Book chapters :

- Brevault, L., Pelamatti, J., **Hebbal, A.**, Balesdent, M., Talbi, E. G., & Melab, N. MDO Related Issues : Multi-objective and mixed continuous/discrete optimization. In *Aerospace System Analysis and Optimization in Uncertainty* (pp. 321-358). Springer, Cham, 2020.
- Brevault, L., Balesdent, M., & **Hebbal, A**. Expendable and reusable launch vehicle design. In *Aerospace System Analysis and Optimization in Uncertainty* (pp. 421-476). Springer, Cham, 2020.

Publications

Communications : Présentées :

- **Hebbal, A.**, Brevault, L., Balesdent, M., Talbi, E., & Melab, N. Multi-fidelity modeling using DGPs : Improvements and a generalization to varying input space dimensions. *4th workshop on Bayesian Deep Learning (NeurIPS 2019)*, Vancouver, Canada.
- **Hebbal, A.**, Brevault, L., Balesdent, M., Talbi, E., & Melab, N. A deep Gaussian Process based model for multi-objective optimization. *The 13th International Conference on Multiple Objective Programming and Goal Programming (MOPGP)*, 2019, Marrakesh, Morocco.
- **Hebbal, A.**, Brevault, L., Balesdent, M., Talbi, E. G., & Melab, N. Multi-objective optimization using deep Gaussian processes : application to aerospace vehicle design. In *AIAA Scitech 2019 Forum* (p. 1973), San Diego, USA.
- Brevault, L., Balesdent, M., **Hebbal, A.**, & Patureau De Mirand, A. Surrogate model-based multi-objective MDO approach for partially eeusable launch vehicle design. In *AIAA Scitech 2019 Forum* (p. 0704), San Diego, USA.
- **Hebbal, A.**, Brevault, L., Balesdent, M., Talbi, E., and Melab, N., Bayesian Optimization using deep Gaussian processes for non-stationary problems. *Program Gaspard Monge (PGMO) days 2018*, Paris, France.
- **Hebbal, A.**, Brevault, L., Balesdent, M., Taibi, E. G., & Melab, N. Efficient global optimization using deep Gaussian processes. In *IEEE Congress on evolutionary computation (CEC)* 2018 (pp. 1-8). Rio de Janeiro, Brazil
- **Hebbal, A.**, Brevault, L., Balesdent, M., Taibi, E. G., & Melab, N. Multi-disciplinary design multi-objective optimization of aerospace vehicles using surrogate models. *International workshop on Optimization and Learning : Challenges and Applications* 2018. Alicante, Spain.

erci pour votre attention !

Bibliography I

Amari, S.-I. (1998).
Natural gradient works efficiently in learning.
Neural computation, 10(2) :251–276.

Bartoli, N., Lefebvre, T., Dubreuil, S., Olivanti, R., Bons, N., Martins, J., Bouhlel, M. A., and Morlier, J. (2017).
An adaptive optimization strategy based on mixture of experts for wing aerodynamic design optimization.
In *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 4433.

Bartoli, N., Lefebvre, T., Dubreuil, S., Olivanti, R., Priem, R., Bons, N., Martins, J. R., and Morlier, J. (2019).
Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design.
Aerospace Science and Technology, 90 :85–102.

Bui, T., Hernández-Lobato, D., Hernandez-Lobato, J., Li, Y., and Turner, R. (2016).
Deep gaussian processes for regression using approximate expectation propagation.
In *International Conference on Machine Learning*, pages 1472–1481.

Choi, S., Alonso, J., Kroo, I., and Wintzer, M. (2004).
Multi-fidelity design optimization of low-boom supersonic business jets.
In *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 4371.

Cutajar, K., Bonilla, E. V., Michiardi, P., and Filippone, M. (2017).
Random feature expansions for deep Gaussian processes.
In *International Conference on Machine Learning*, pages 884–893. PMLR.

Bibliography II

Cutajar, K., Pullin, M., Damianou, A., Lawrence, N., and González, J. (2019).
Deep gaussian processes for multi-fidelity modeling.
arXiv preprint arXiv :1903.07320.

Dai, Z., Damianou, A., González, J., and Lawrence, N. (2015).
Variational auto-encoded deep gaussian processes.
arXiv preprint arXiv :1511.06455.

Damianou, A. and Lawrence, N. (2013).
Deep gaussian processes.
In *Artificial Intelligence and Statistics*, pages 207–215.

Emmerich, M. T., Giannakoglou, K. C., and Naujoks, B. (2006).
Single-and multiobjective evolutionary optimization assisted by gaussian random field metamodels.
IEEE Transactions on Evolutionary Computation, 10(4) :421–439.

Gramacy, R. B. and Lee, H. K. H. (2008).
Bayesian treed Gaussian process models with an application to computer modeling.
Journal of the American Statistical Association, 103(483) :1119–1130.

Haas, T. C. (1990).
Kriging and automated variogram modeling within a moving window.
Atmospheric Environment. Part A. General Topics, 24(7) :1759–1769.

Haibin, Y., Chen, Y., Low, B. K., Jaillet, P., and Dai, Z. (2019).
Implicit posterior variational inference for deep Gaussian processes.
In *Advances in Neural Information Processing Systems*, pages 14502–14513.

Bibliography III

Havasi, M., Hernández-Lobato, J. M., and Murillo-Fuentes, J. J. (2018).
Inference in deep Gaussian processes using stochastic gradient Hamiltonian Monte Carlo.
In *Advances in neural information processing systems*, pages 7506–7516.

Higdon, D., Swall, J., and Kern, J. (1999).
Non-stationary spatial modeling.
Bayesian statistics, 6(1) :761–768.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998).
Efficient global optimization of expensive black-box functions.
Journal of Global optimization, 13(4) :455–492.

Kennedy, M. C. and O'Hagan, A. (2001).
Bayesian calibration of computer models.
Journal of the Royal Statistical Society : Series B (Statistical Methodology), 63(3) :425–464.

Le Gratiet, L. and Garnier, J. (2014).
Recursive co-kriging model for design of computer experiments with multiple levels of fidelity.
International Journal for Uncertainty Quantification, 4(5).

Li, W., Chen, S., Jiang, Z., Apley, D. W., Lu, Z., and Chen, W. (2016).
Integrating Bayesian calibration, bias correction, and machine learning for the 2014 Sandia verification and validation challenge problem.
Journal of Verification, Validation and Uncertainty Quantification, 1(1).

Loh, W.-L. et al. (1996).
On latin hypercube sampling.
The annals of statistics, 24(5) :2058–2080.

Bibliography IV

Marmin, S. and Filippone, M. (2018).
Variational calibration of computer models.
arXiv preprint arXiv :1810.12177.

Paciorek, C. J. and Schervish, M. J. (2006).
Spatial modelling using a new class of nonstationary covariance functions.
Environmetrics, 17(5) :483–506.

Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N., and Karniadakis, G. E. (2017).
Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling.
Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences, 473(2198) :20160751.

Rasmussen, C. E. (2004).
Gaussian processes in machine learning.
In *Advanced lectures on machine learning*, pages 63–71. Springer.

Rasmussen, C. E. and Ghahramani, Z. (2002).
Infinite mixtures of gaussian process experts.
In *Advances in neural information processing systems*, pages 881–888.

Rossi, S., Heinonen, M., Bonilla, E., Shen, Z., and Filippone, M. (2020).
Rethinking sparse gaussian processes : Bayesian approaches to inducing-variable approximations.
arXiv preprint arXiv :2003.03080.

Salimbeni, H. and Deisenroth, M. (2017).
Doubly stochastic variational inference for deep gaussian processes.
arXiv preprint arXiv :1705.08933.

Bibliography V

Sasena, M. J. (2002).
Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations.
PhD thesis, University of Michigan Ann Arbor, MI.

Shah, A. and Ghahramani, Z. (2016).
Pareto frontier learning with expensive correlated objectives.
In *International Conference on Machine Learning*, pages 1919–1927.

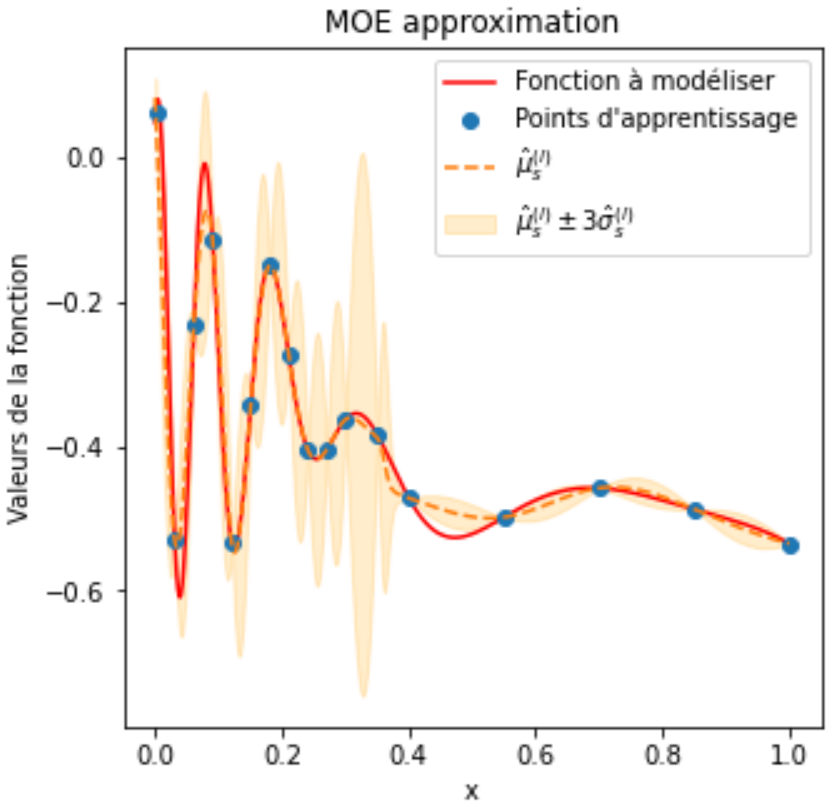
Snoek, J., Swersky, K., Zemel, R., and Adams, R. P. (2014).
Input warping for Bayesian optimization of non-stationary functions.
In *International Conference on Machine Learning*, pages 1674–1682.

Tao, S., Apley, D. W., Chen, W., Garbo, A., Pate, D. J., and German, B. J. (2019).
Input mapping for model calibration with application to wing aerodynamics.
AIAA Journal, pages 2734–2745.

Titsias, M. (2009).
Variational learning of inducing variables in sparse gaussian processes.
In *Artificial Intelligence and Statistics*, pages 567–574.

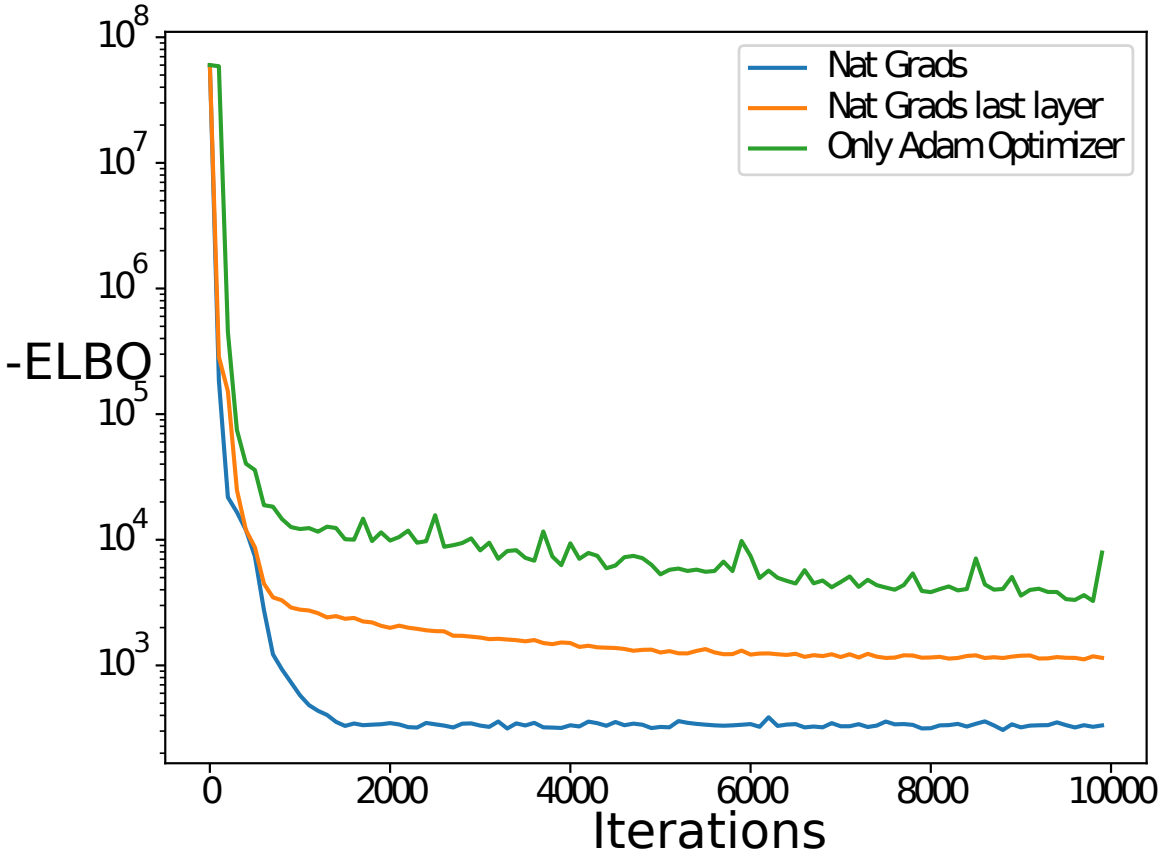
Xiong, Y., Chen, W., Apley, D., and Ding, X. (2007).
A non-stationary covariance-based kriging method for metamodelling in engineering design.
International Journal for Numerical Methods in Engineering, 71(6) :733–756.

MoE



MoE on the modified Xiong function

ELBO convergence



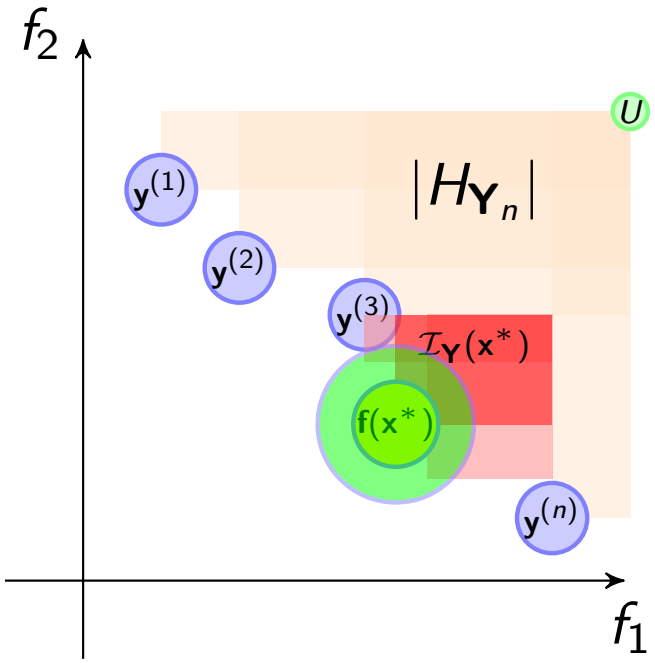
Evidence Lower Bound convergence with different training approaches.

Expected Hyper-Volume Improvement

Expected Hyper-volume improvement [Emmerich et al., 2006]

The expected hyper-volume improvement is the mathematical expected improvement of the hyper-volume by adding a candidate to the sample

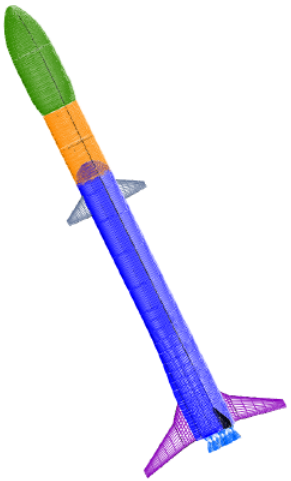
$$EHVI(\mathbf{x}^*) = \mathbb{E}_{p(\mathbf{f}(\mathbf{x}^*)|\mathbf{Y},\mathbf{x}^*,\mathbf{X})}\left[\mathcal{I}_{\mathbf{Y}}(\mathbf{x}^*)\right]$$



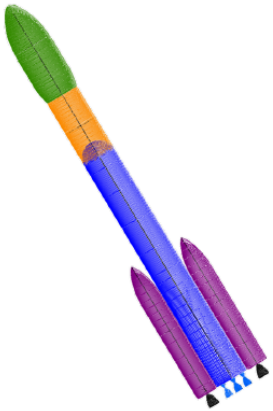
$$p(\mathbf{f}(\mathbf{x}^*)|\mathbf{Y},\mathbf{x}^*,\mathbf{X})\mathrm{d}\mathbf{f}(\mathbf{x}^*) = \prod_{i=1}^{n_o} p(f_i(\mathbf{x}^*)|\mathbf{y}_i,\mathbf{x}^*,\mathbf{X})\mathrm{d}f_i(\mathbf{x}^*)$$

$|H_{\mathbf{Y}_n}|$: The dominated hyper-volume by $\mathbf{Y}_n$

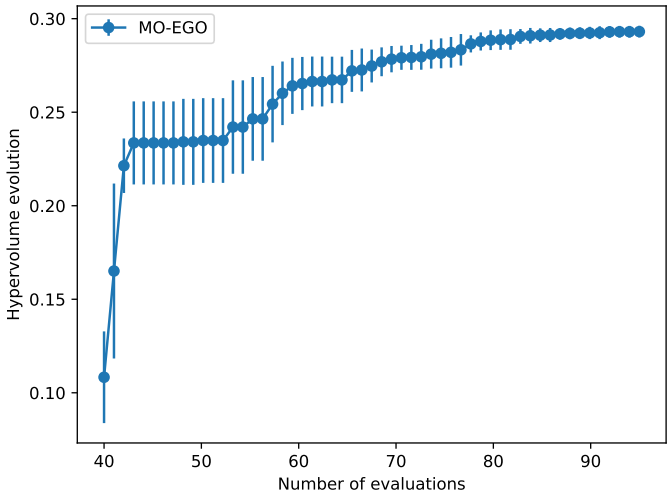
$\mathcal{I}_{\mathbf{Y}}(\mathbf{x}^*)$: The improvement in hyper-volume by adding the evaluation of $\mathbf{x}^*$ to $\mathbf{Y}_n$.



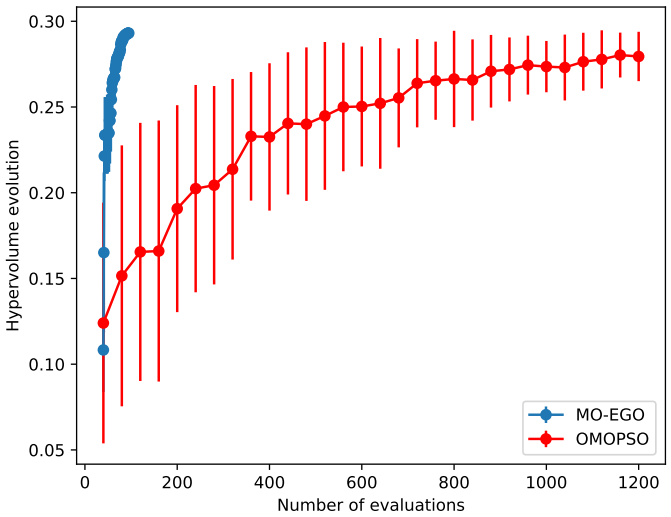
SSO mission, with recovery, refurbishment and reuse of the launch vehicle first stage.



GTO mission, in expandable mode with no recovery of the vehicle first stage. Non stationary objective function

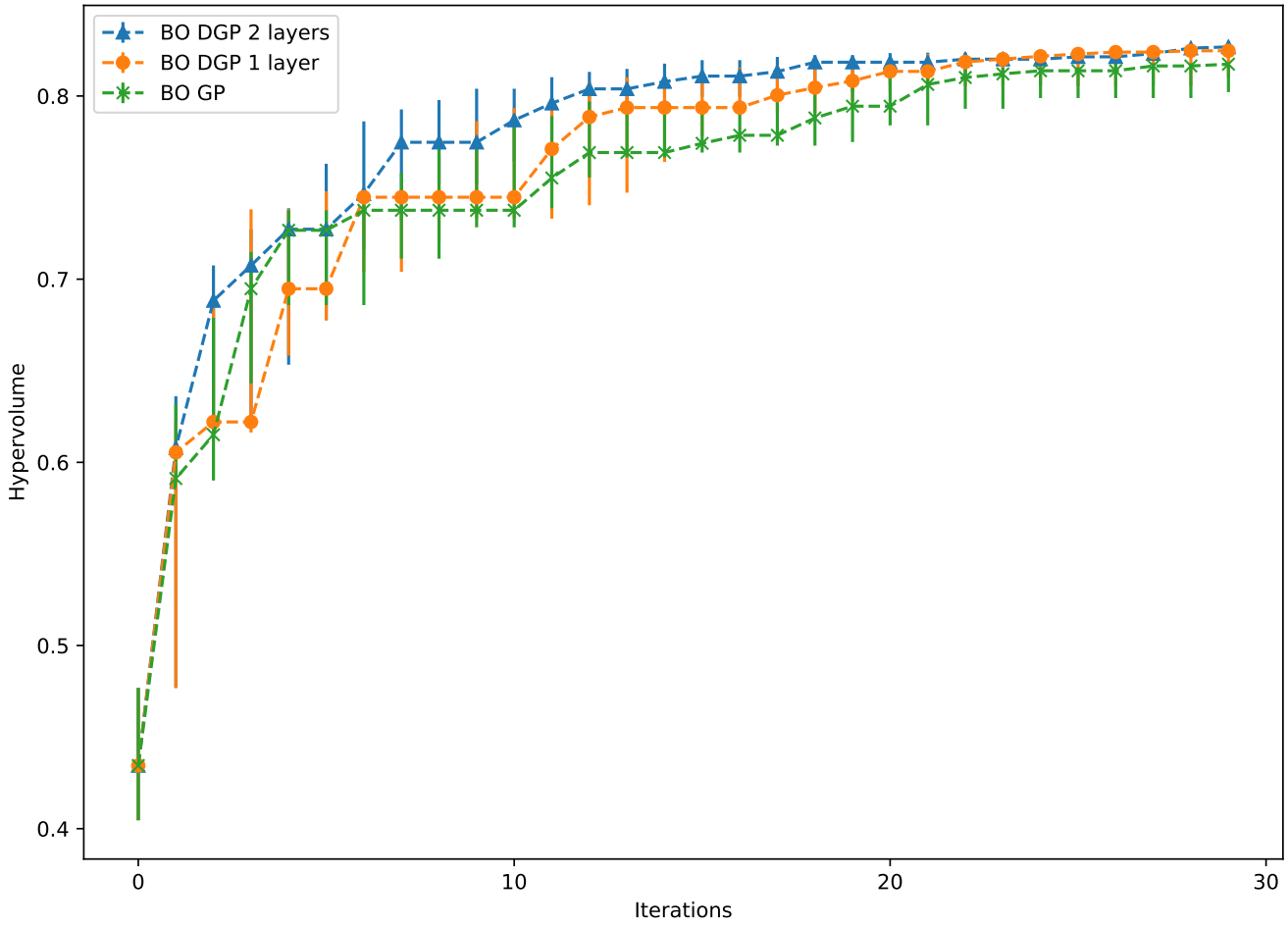


Hypervolume evolution of MO-EGO



Hypervolume evolution of OMOPSO

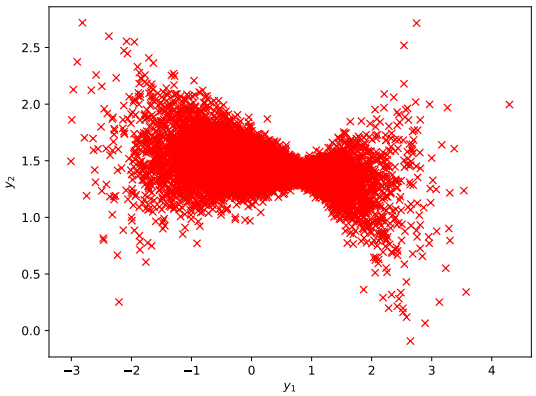
MO-BO with DGPs

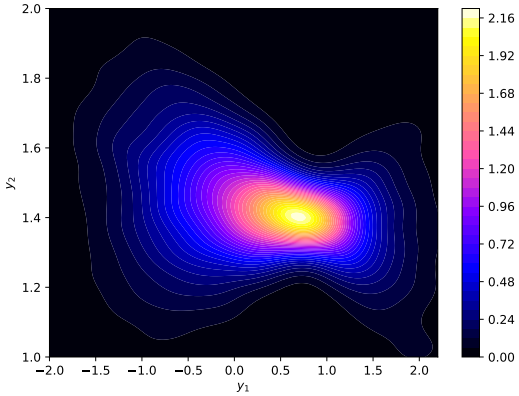


Convergence plot of BO with different architecture of DGPs and a regular GP.

EHVI with correlated objectives

The predictive distribution of the model is estimated using a Kernel density estimation with a Gaussian kernel and a diagonal Bandwidth function **B**.





The EHVI comes back to the average over s analytically tractable EHVI :

$$EHVI(\mathbf{x}) \approx \mathbb{E}_{\mathcal{N}(\mathbf{f}(\mathbf{x})|\mathbf{f}^{[t]},\mathbf{B})}\left[\mathcal{I}_{\mathbf{Y}}(\mathbf{x})\right]$$

s samples drawn from the posterior predictive distribution $(\mathbf{f}^{[1]}, \ldots, \mathbf{f}^{[s]})$

$$\hat{p}(\mathbf{f}(\mathbf{x})) = \frac{1}{s}\sum_{j=1}^{s}\prod_{i=1}^{n_o}\phi_{\mathcal{N}(f_{[i]}^{[t]},\mathbf{B}_{ii})}\left(f_{[i]}(\mathbf{x})\right)$$

Kursawe MO case

$$
\begin{array}{ll}
\min & [f_1(\mathbf{x}), f_2(\mathbf{x})] \\
\text{s.t.} & -5 \leq x_i \leq 5 \qquad\qquad i = 1, \ldots, 3 \\
\text{with} & f_1(\mathbf{x}) = \sum_{i=1}^{2} \left(-10 \exp\left(-0.2\sqrt{x_i^2 + x_{i+1}^2}\right)\right) \\
\text{and} & f_2(\mathbf{x}) = \sum_{i=1}^{3} \left(|x_i|^{0.8} + 5 \sin\left(x_i^3\right)\right)
\end{array}
\tag{1}
$$

Model	**average HV**	**HV standard deviation**
MO-DGP	0.350	**0.060**
GPs	**0.372**	0.072
LMC	0.273	0.078
NSGA-II	0.173	0.0322

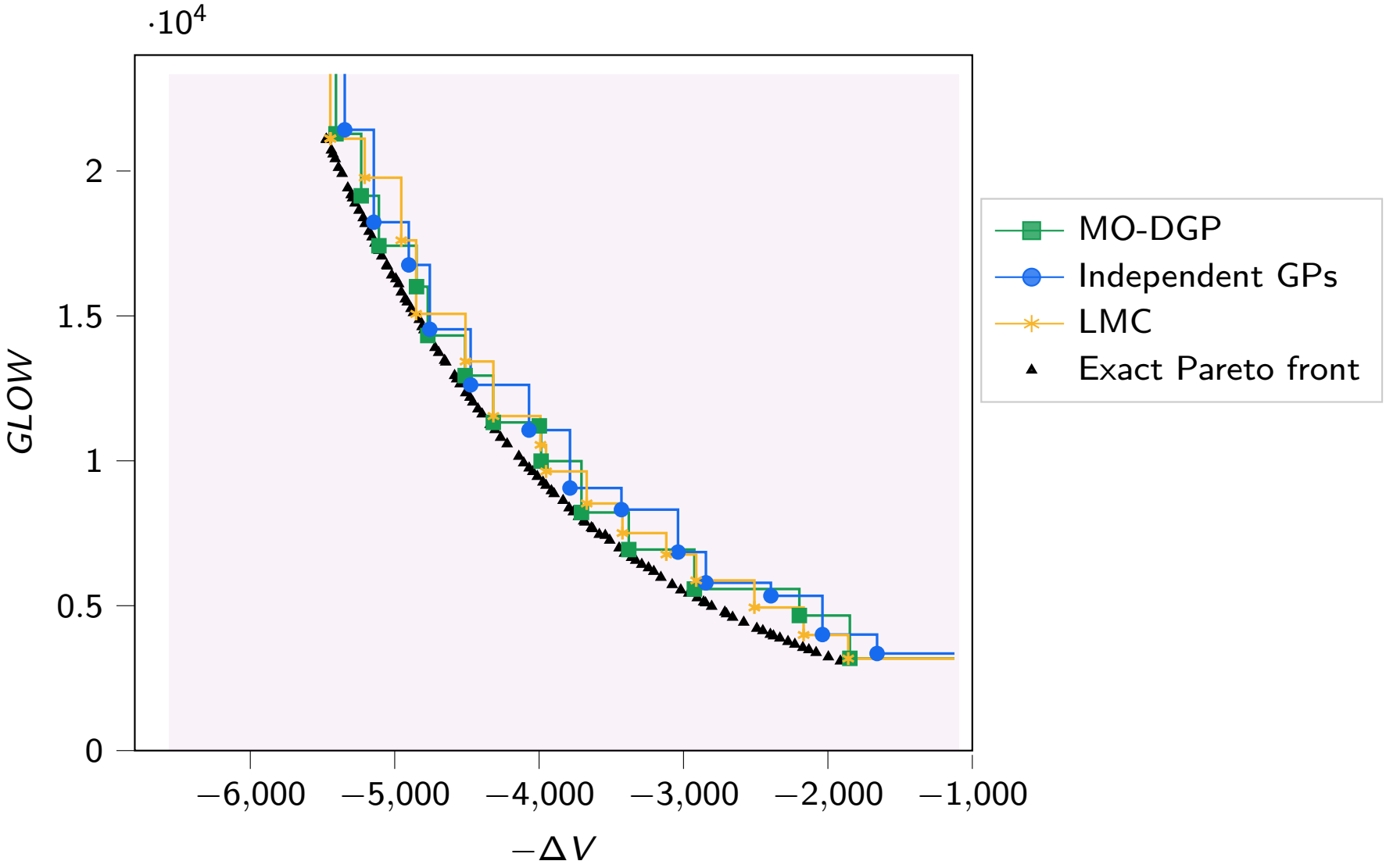
TABLE – Number of parameters and time needed for training of each model for Problem 1 with 30 LF two-dimensional input data and three scenarios on the available HF information (4, 6, and 8 input data for , 4-dimensional HF). The computational time is on a Tesla P100 GPU and using Tensorflow.

Analytical Problem 1			
HF DoE size	**Algorithms**	Number of parameters	Time for training (seconds)
4 data points	HF model	6	$\approx 0.5s$
	BC model	6	$\approx 0.5s$
	MF-DGP-EM	673	$\approx 232s$
6 data points	HF model	6	$\approx 0.55s$
	BC model	6	$\approx 0.55s$
	MF-DGP-EM	754	$\approx 245s$
8 data points	HF model	6	$\approx 0.62s$
	BC model	6	$\approx 0.62s$
	MF-DGP-EM	855	$\approx 280s$

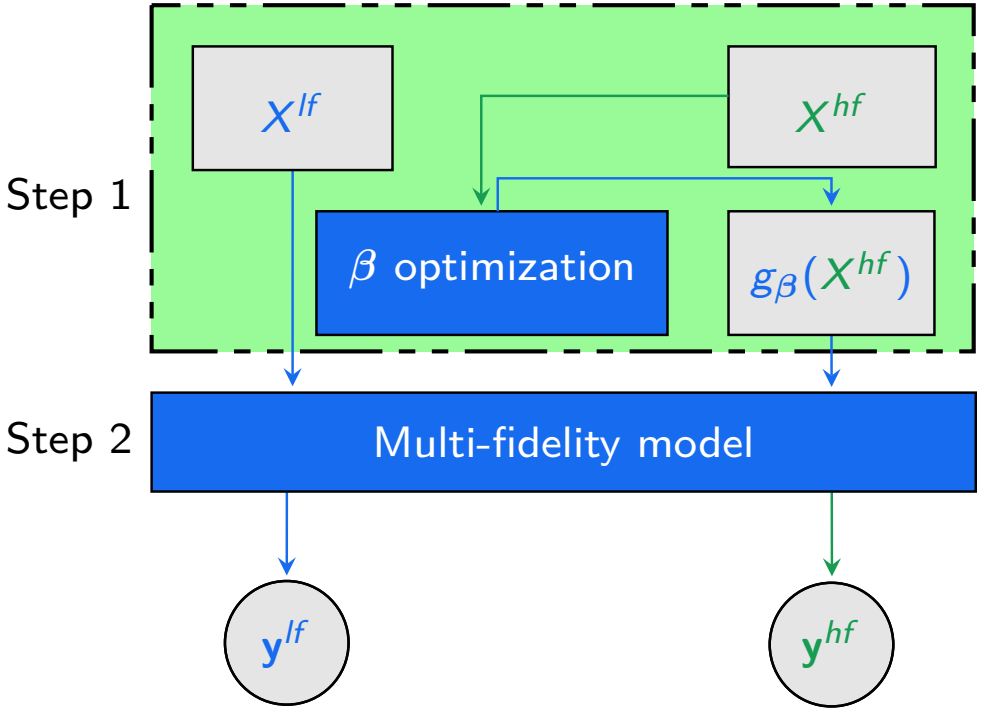
TABLE – Number of parameters and time needed for training of each model for the aerodynamic problem with 120 LF 6-dimensional input data and three scenarios on the available HF information (10, 15, and 20 input data for a 12-dimensional the HF). The computational time is on a Tesla P100 GPU and using Tensorflow.

Aerodynamic problem			
HF DoE size	**Algorithms**	Number of parameters	Time for training (seconds)
10 data points	HF model	14	$\approx 1.8s$
	BC model	14	$\approx 1.8s$
	MF-DGP-EM	9221	$\approx 500s$
15 data points	HF model	14	$\approx 0.55s$
	BC model	14	$\approx 0.55s$
	MF-DGP-EM	10251	$\approx 530s$
20 data points	HF model	14	$\approx 3.4s$
	BC model	14	$\approx 3.4s$
	MF-DGP-EM	11606	$\approx 580s$

Application to industrial test case : design of aerospace vehicle



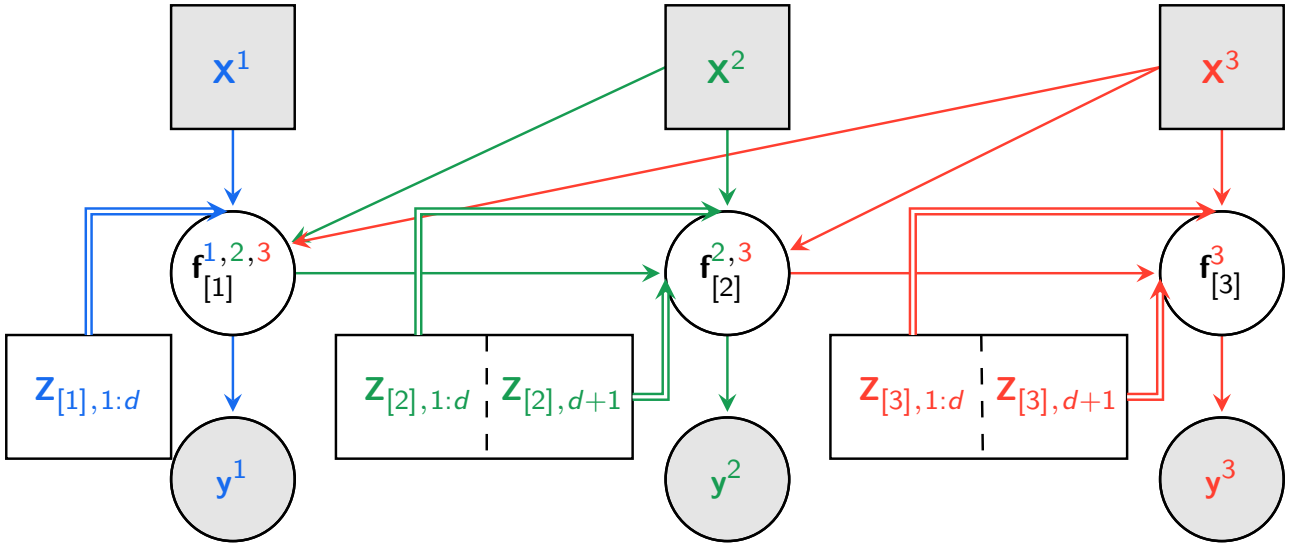
Input Mapping Calibration



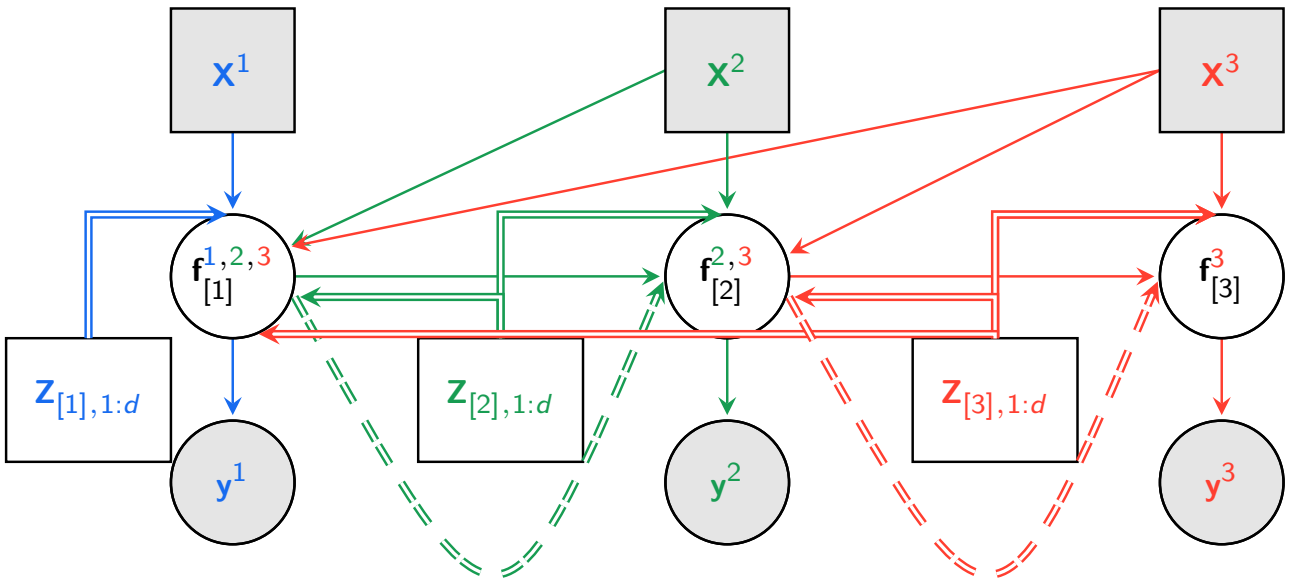
Classic approach for multi-fidelity with varying input spaces [Tao et al., 2019].

- Use of the exact low-fidelity model,
- Disjoint optimization of the mapping parameters and the multi-fidelity model,
- Training only on the lower fidelity input space.

Optimization of the inducing inputs



Optimization of the inducing inputs



$\hat{f}_{[1]}\left(\mathbf{Z}_{[i],1:d}\right)$ $\hat{f}_{[2]}\left(\left[\mathbf{Z}_{[i],1:d}, \hat{f}_{[1]}(\mathbf{Z}_{[i],1:d})\right]\right)$