Séminaire FONISEN

A quick introduction to Surrogate modeling with **SMT**

joseph.morlier@isae-supaero.fr

P. Saves and R. Lafage and

N. Bartoli , T. Lefevbre,

R. Charayron…

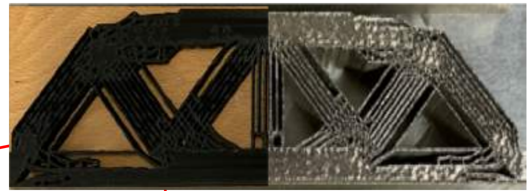
ONERA THE FRENCH AEROSPACE LAB

ISAE Institut Supérieur de l'Aéronautique et de l'Espace SUPAERO

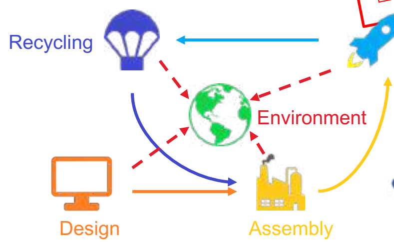
About Me? https://ica.cnrs.fr/en/author/jmorlier/

- 6 PhDs, 3 MsCs

Digital fabrication



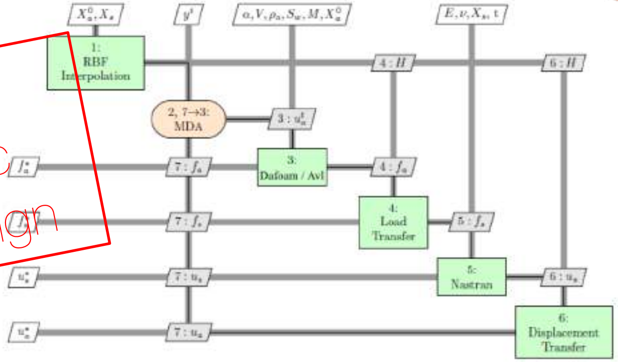
EcoOptimization



Brightway

openMDAO

MDO for Aeroelastic Wing design

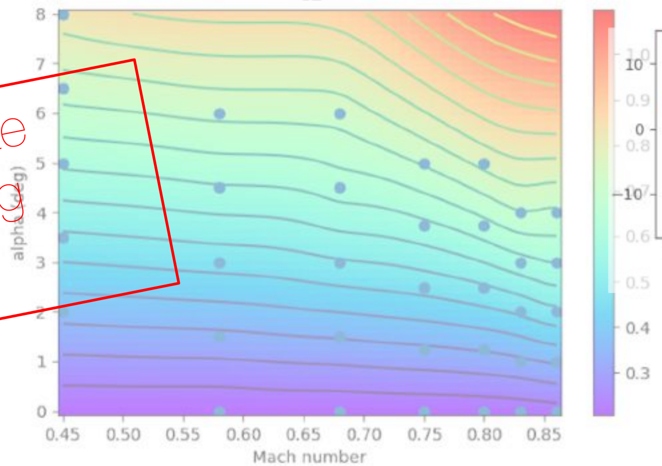


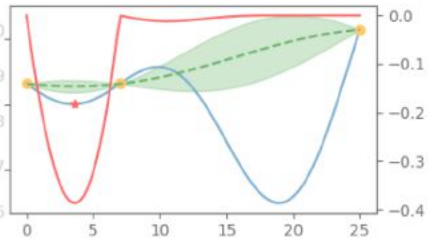
LCA4MDAO (**needed LCA database ecoinvent)**

https://github.com/mid2SUPAERO/LCA4MDAO

https://github.com/SMTorg/SMT

Surrogate modeling AI4E





? Can we Learn from

data $y=f(X)$ to predict engineering (costly) QoI

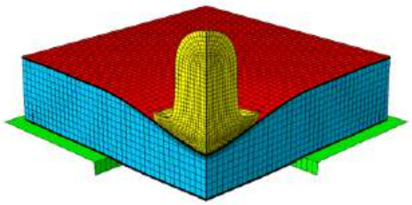
BTW, How to start with an open source Python toolbox?

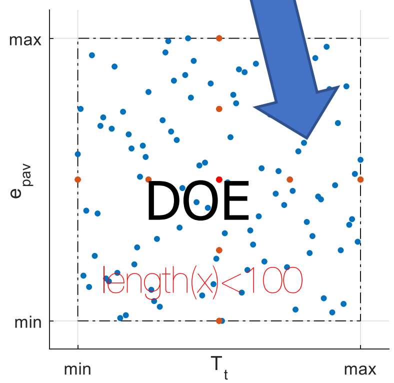
What is Surrogate Modeling ?

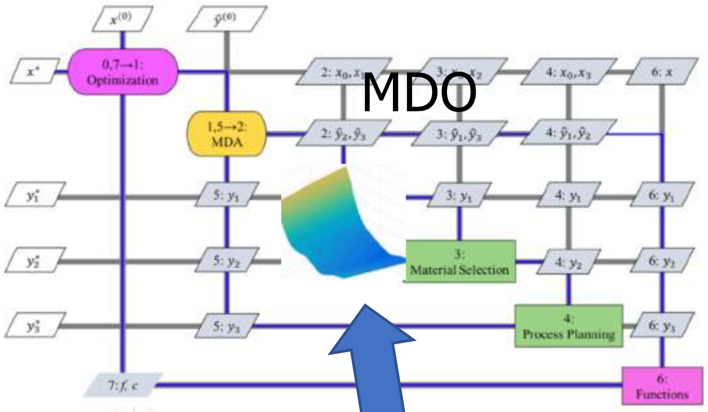
A. Forrester, A. Sobester, and A. Keane. "Engineering Design via Surrogate Modelling: A Practical Guide". Coll. John Wiley & Sons (2008).

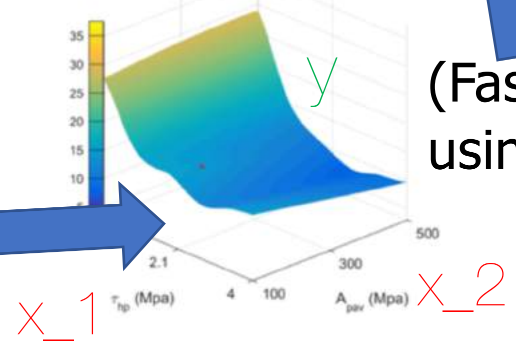
A surrogate model of a function is an **approximation of Expensive Computer simulation**: It's a supervised learning process in AI.

(Expensive) simulation









(Fast) Surrogate using SMT y=f(X)

As the surrogate is less costly to evaluate it can be **used as a "fast" code in a <u>Multidisciplinary Design Optimization loop</u>**. [or do Uncertainty Quantification or do Bayesian Optimization etc…]

Au programme

- **Part1: GP aka Kriging**
- **Part2: About SMT**
- **Part3: Engineering applications**

Part1 : GP aka Kriging

Mathematical foundations

ML vs Engineering

Kriging (Pionneer)	Gaussian Processes (link with AI)
Developed by Daniel Krige – 1951; formalized by Georges Mathéron in the 60's (Mines Paris)	Neural network with infinite neurons tend to Gaussian Process 1994

Krige, D. G., 1951, A statistical approach to some basic mine valuation problems on the Witwatersrand: J. Chem. Metal. Min. Soc. South Africa, v. 52, p. 119–139.

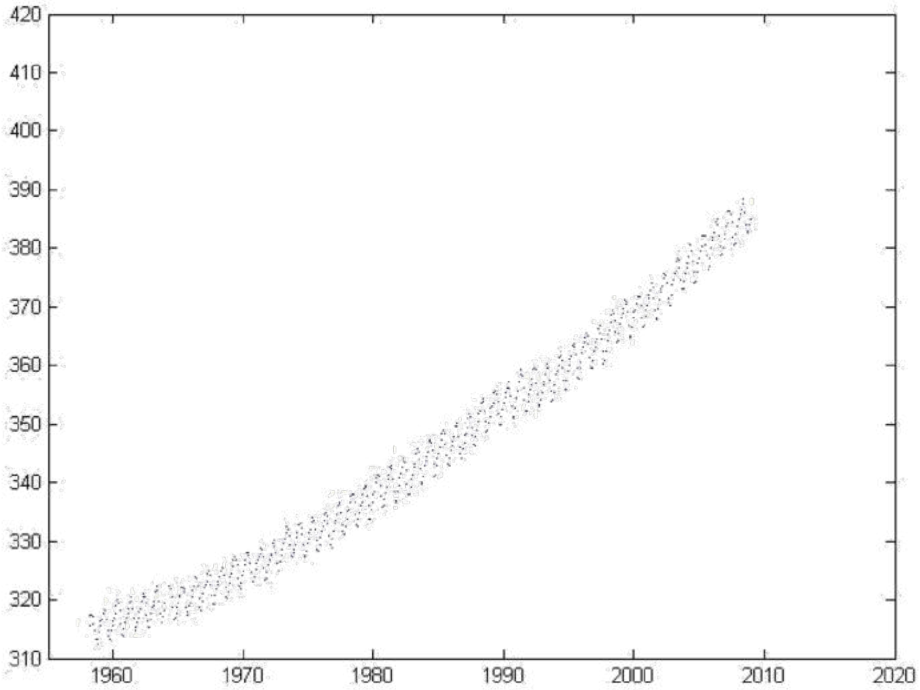
Matheron, G., 1963b, Principles of geostatistics: Economic Geol., v. 58, p. 1246–1266.

Neal, R. Priors for infinite networks. Tech. rep., University of Toronto, 1994.

Williams, C. K. I., and Rasmussen, C. E. Gaussian processes for regression. *Advances in Neural Information Processing Systems 8* (1996), 514–520.

http://extrapolated-art.com

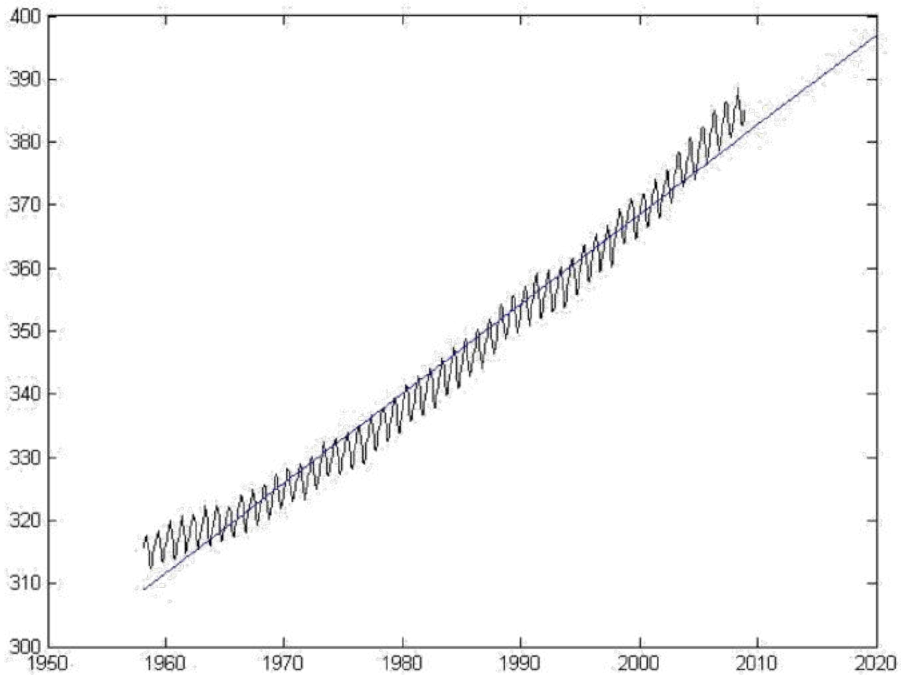
Limit of linear models for prediction



Month-wise data of Co2 concentration in atmosphere at Hawaii

Image Source: http://mlg.eng.cam.ac.uk/teaching/4f13/1314/

Example – Linear Regression

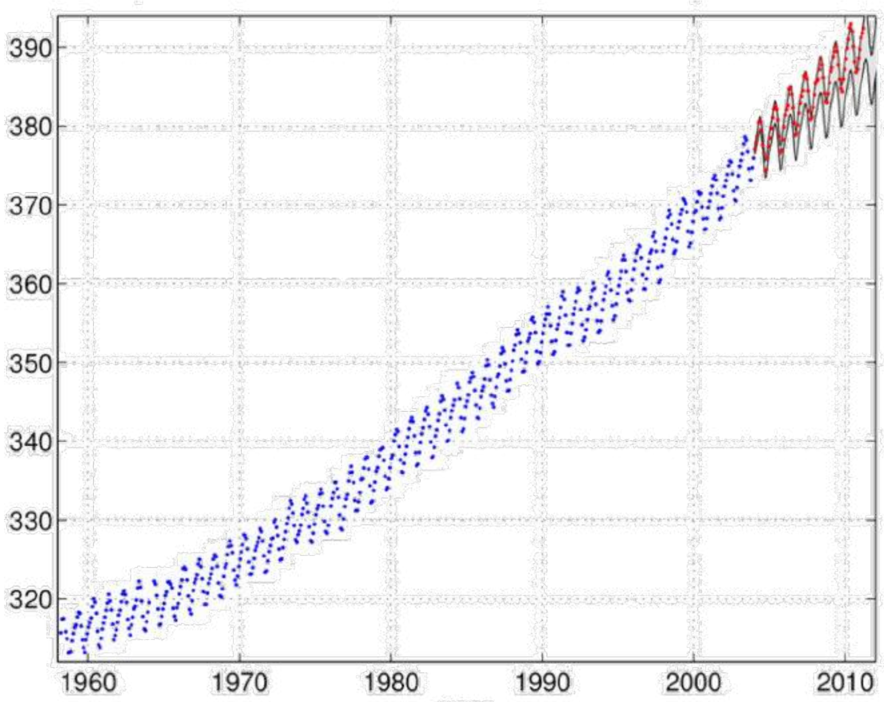


Should we choose a polynomial?
What degree of polynomial should we choose? (overfitting)
For a given degree, what parameters of polynomial should we choose

Image Source: http://mlg.eng.cam.ac.uk/teaching/4f13/1314/

Example – GP

https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_gpr_co2.html



Predicted variance after year 2005 in grey, real data-points in red

Image Source: http://mlg.eng.cam.ac.uk/teaching/4f13/1314/

Main steps to build a surrogate model

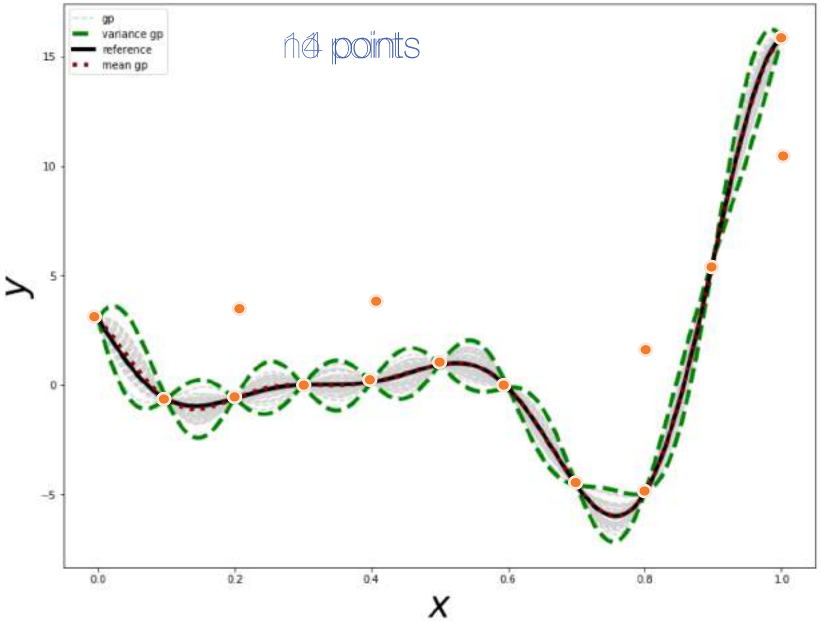
1. Black box to replace by a surrogate model

$$y(x^{(i)}) = f(x^{(i)}) \text{ with } x^{(i)} \in \mathbb{R}^d$$

where f is the model (expensive, free-derivative, noise-free) such as CFD code or FEM code

C. Demay, B. Iooss, L. Le Gratiet, A. Marrel, Model selection based on validation criteria for gaussian process regression: An application with highlights on the predictive variance, 2022, Quality and Reliability Engineering International.

Gaussian process or Kriging model



$x \in \mathbb{R}^d$

$y(x) \in \mathbb{R}$

DOE (LHS)

Gaussian process characterized by:

- its mean (or trend) $\mu(x) \in \mathbb{R}$
- its covariance Kernel $k(x, x') \in \mathbb{R}$

$\hat{y}(x)$

Estimation of hyperparameters θ_i (i=1,…,d) by MLE

Covariance matrix

+ information provided by data

Prediction $\hat{y}$ and variance $\hat{\sigma}^2$ estimations at a new point

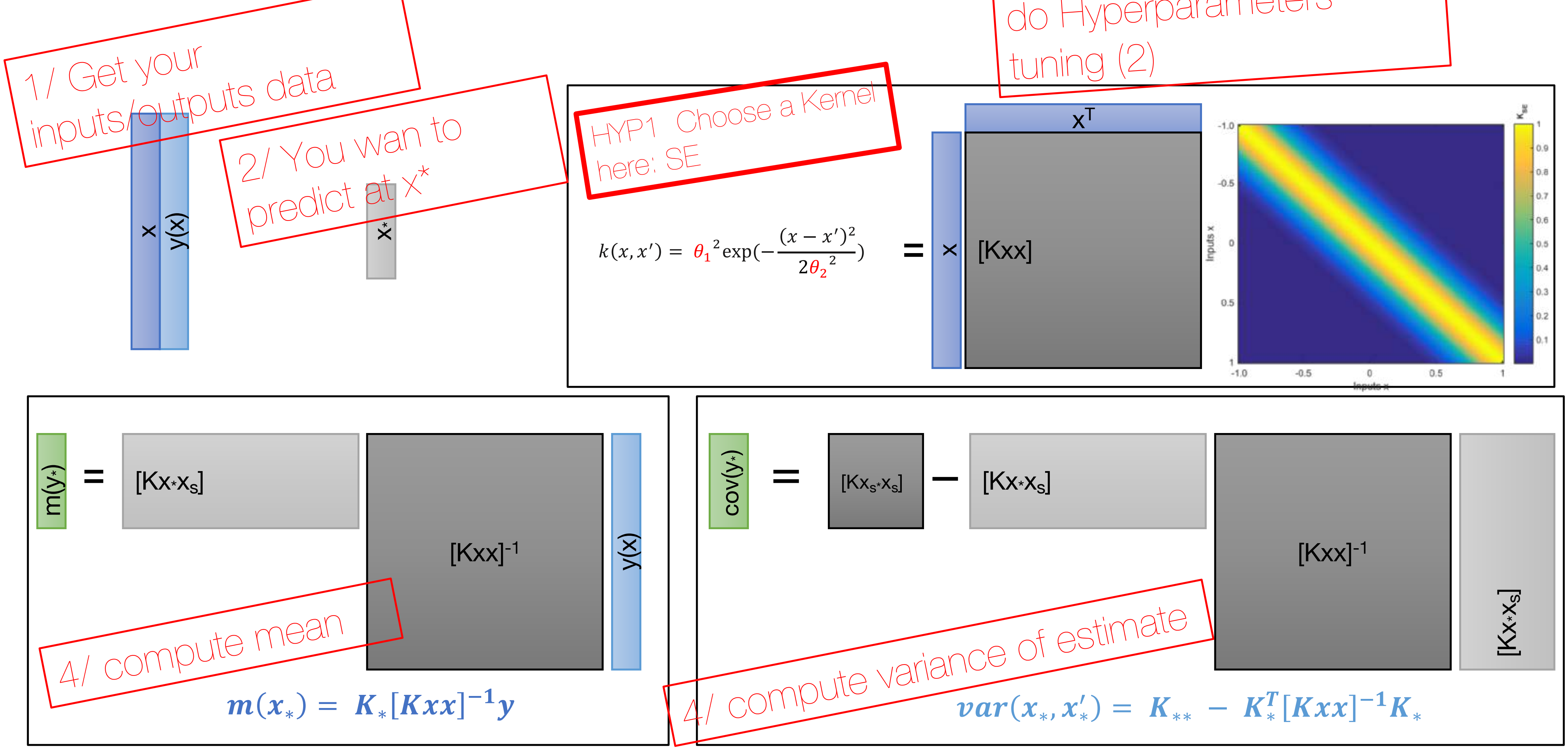
Quantification of the uncertainties in these estimates

$$f(x) \Rightarrow Y(x) = \mathcal{N}(\hat{y}(x), s^2(x))$$

- ➔ Hyperparameters tuning
- ➔ Number of hyperparameters increases with the dimension d (number of design variables)
- ➔ Curse of dimensionality

D. G. Krige, A statistical approach to some basic mine valuation problems on the Witwatersrand, 1951, Journal of the Southern African Institute of Mining and Metallurgy.

Matrix view of Gaussian Process



posterior_mean = covXXs @ np.linalg.inv(covXX_noisy) @ y

posterior_cov = covXsXs − covXXs @ np.linalg.inv(covXX_noisy) @

Need deeper understanding?

The Art of Gaussian Processes: Classical and Contemporary

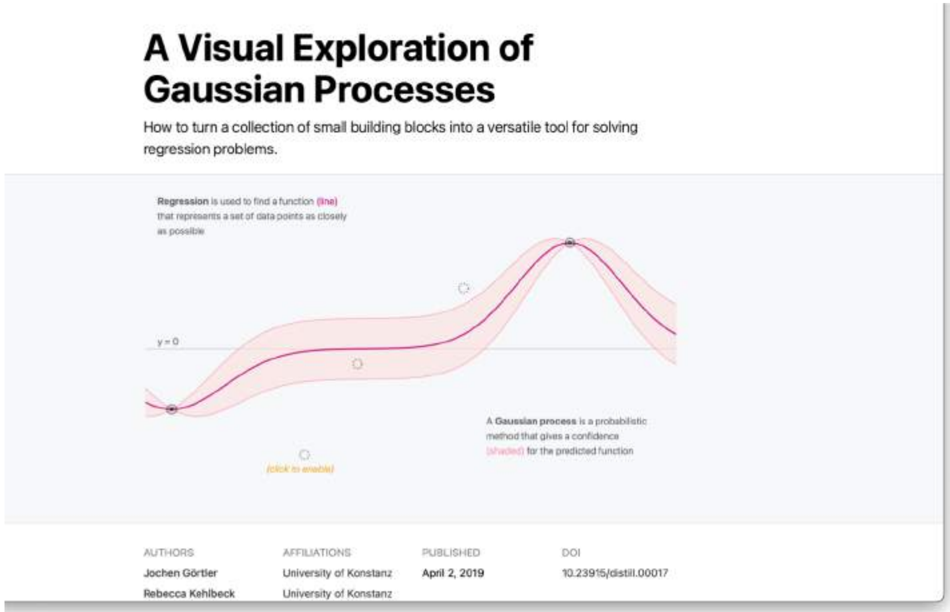
César Lincoln C. Mattos[1] Felipe Tobar[2]

[1]Department of Computer Science
Federal University of Ceará
Fortaleza, Ceará, Brazil
cesarlincoln@dc.ufc.br

[2]Initiative for Data & Artificial Intelligence
University of Chile
Santiago, Chile
ftobar@dim.uchile.cl

2021

https://neurips.cc/virtual/2021/tutorial/21890



https://distill.pub/2019/visual-exploration-gaussian-processes/

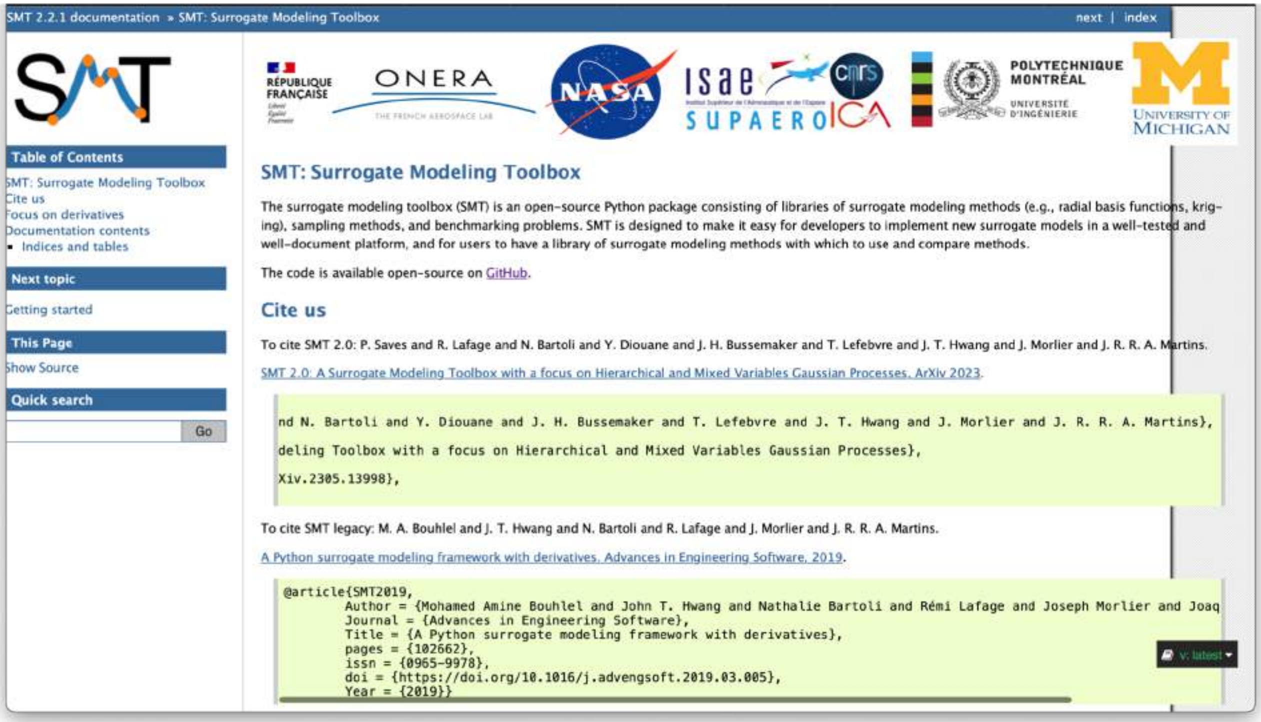
Part2: About SMT

FAQ

X_0 in practice
Since 2017

https://smt.readthedocs.io/en/latest

https://github.com/SMTorg/smt



Required packages

SMT depends on the following modules: numpy, scipy, scikit-learn, pyDOE2 and Cython.

Installation

If you want to install the latest release

```
pip install smt
```

or else if you want to install from the current master branch

```
pip install git+https://github.com/SMTOrg/smt.git@master
```

Usage

For examples demonstrating how to use SMT, you can take a look at the tutorial notebook or go to the 'smt/examples' folder.

The very First question:

Q1: Do you want to create a surrogate model or to do Global Optimization?

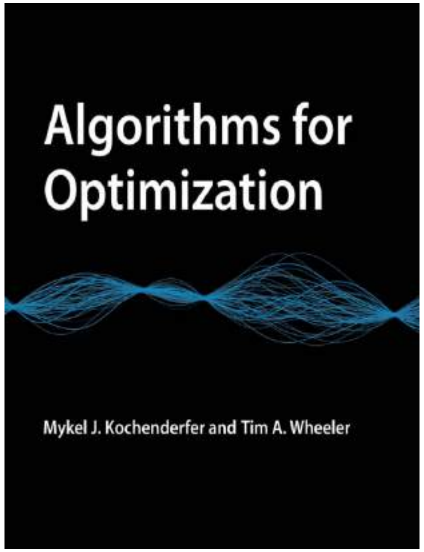
The Second question:

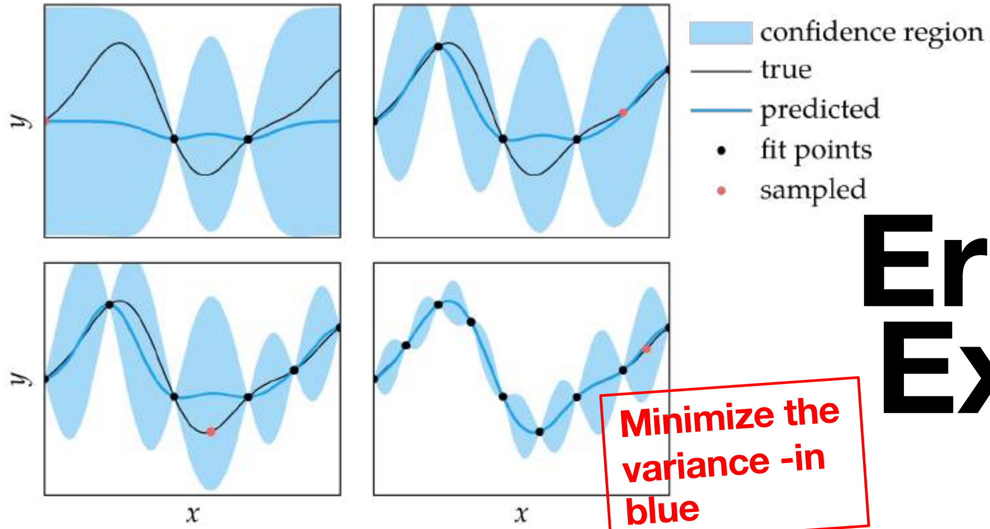
Q2: Are your data cheap or expensive to evaluate?

For this:
you need to understand one key concept

Given a surrogate model (GP) with both prediction and confidence interval, an optimization procedure must balance the search for the expected optimal point and decreasing uncertainty

In other words, the algorithm must ***balance exploitation with exploration to find the global optimum***





I want the most precise surrogate

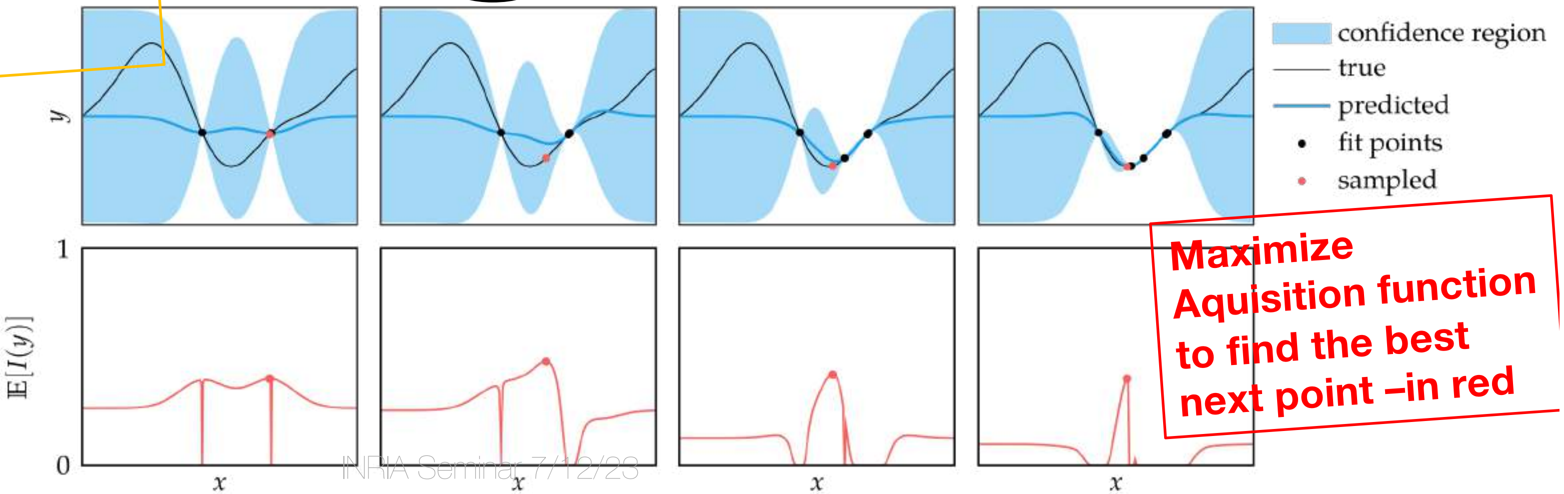
Error-Based Exploration

Minimize the variance -in blue

VS

I want a global optimum at fixed budget

Efficient Global Optimization

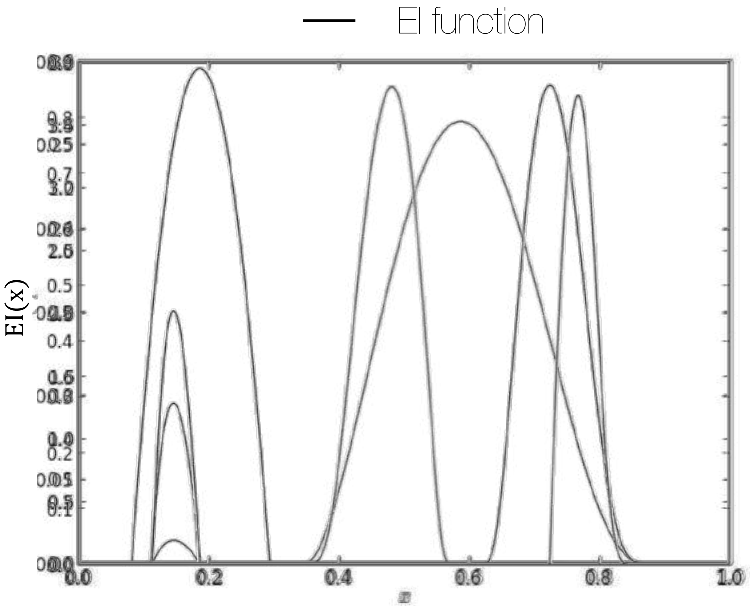


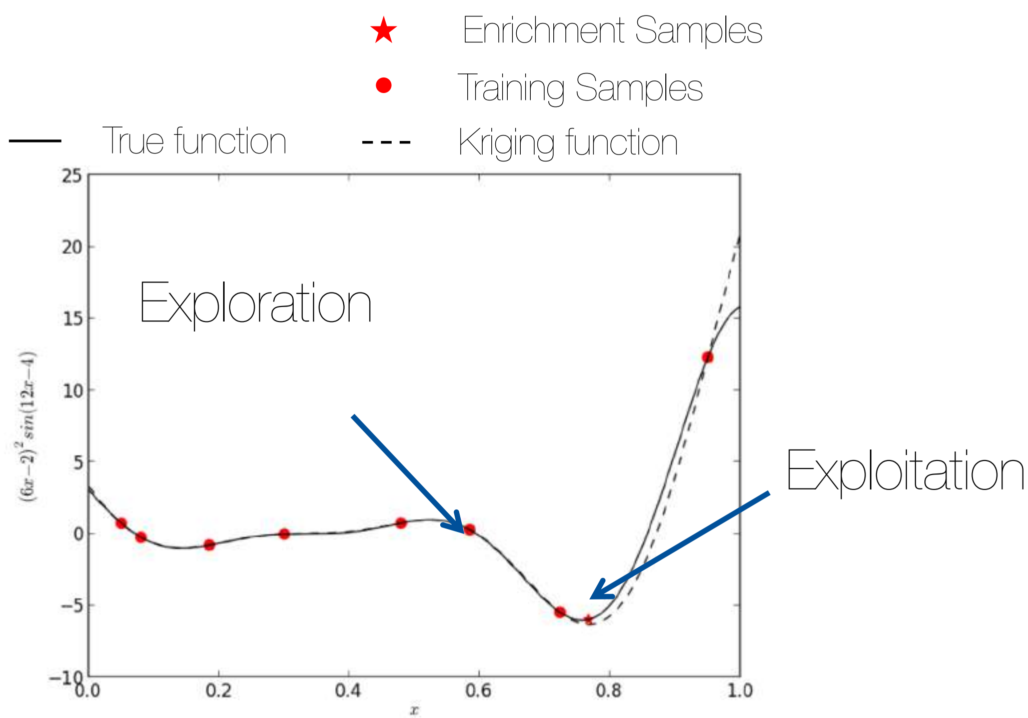
Maximize Aquisition function to find the best next point –in red

Efficient Global Optimization: Illustration

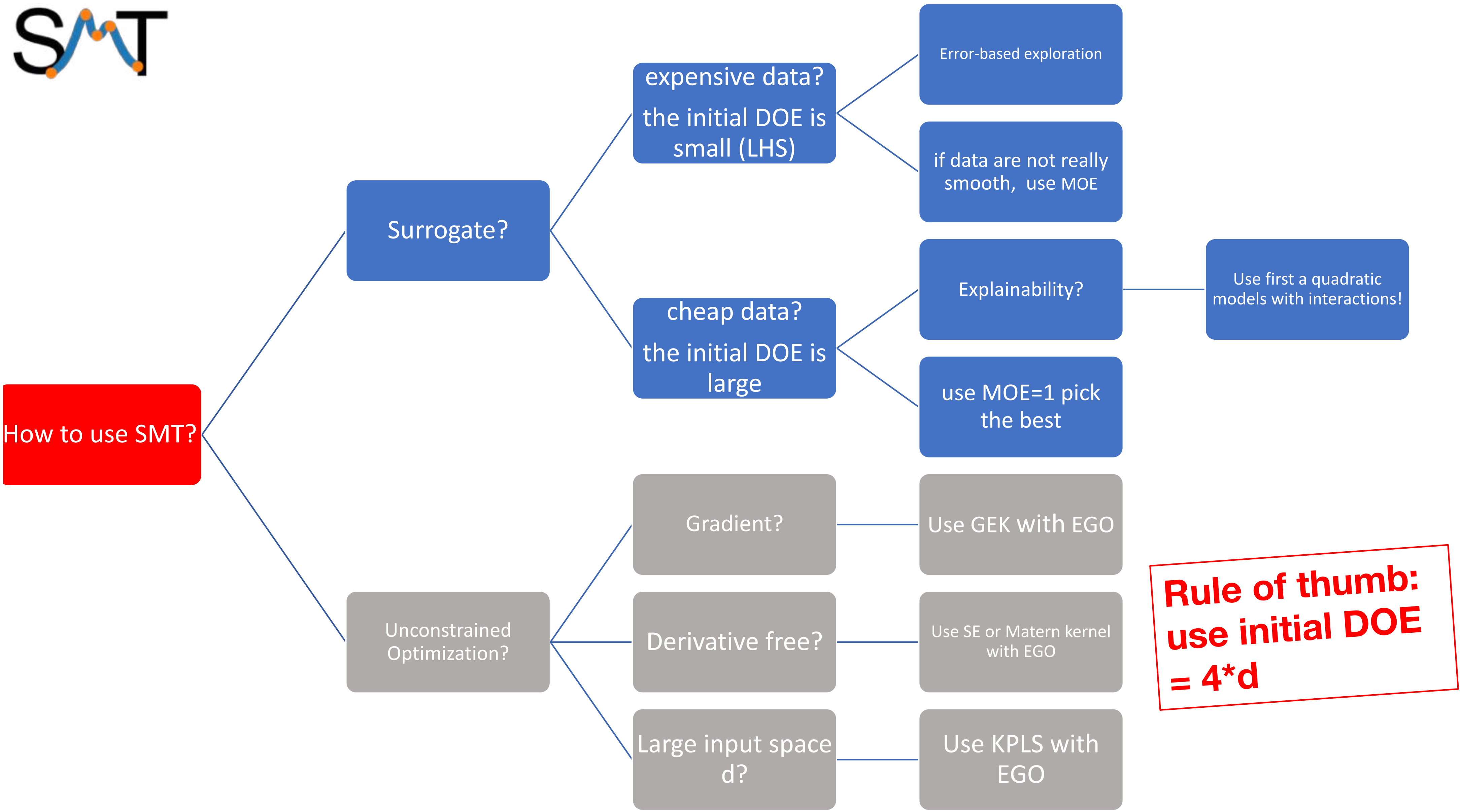
$$\begin{cases} \min \ (6x-2)^2 \sin(12x-4) \\ s.t. \\ 0 \le x \le 1 \end{cases}$$

- 4 points for the initial DOE
  - 6 iterations





N. Bartoli, Optimisation adaptative basée sur les métamodeles, HdR, Université Toulouse III.


How to use SMT?
Surrogate?
expensive data?
the initial DOE is small (LHS)
Error-based exploration
if data are not really smooth, use MOE
cheap data?
the initial DOE is large
Explainability?
Use first a quadratic models with interactions!
use MOE=1 pick the best
Unconstrained Optimization?
Gradient?
Use GEK with EGO
Derivative free?
Use SE or Matern kernel with EGO
Large input space d?
Use KPLS with EGO
Rule of thumb: use initial DOE = 4*d

Next question:
Can you handle different kinds of variable types?

What kind of variable types are available in SMT 2.0 ?

- Continuous
- Discrete
- Categorical
- Hierarchical
- Mixed

https://automl.github.io/ConfigSpace/main/index.html

Mixed-type:

One-hot encoding:

Red: [1,0,0] Green: [0,1,0] Blue: [0,01]

Categorical, continuous, discrete, binary variables

Categorical: Red, Green, Blue

Continuous: [0,1]

Discrete: 1,2,3,4,5

Binary: Yes, No

Tuning hyperparameters for deep neural network

$$= f(\overbrace{\underbrace{x_1}_{\substack{\text{learning rate} \\ \in [1e^{-6},\ 1e^{-1}]}}, \underbrace{x_2}_{\substack{\text{weight decay} \\ \in [1e^{-6},\ 1e^{-1}]}}}^{\text{continuous variables}} , \overbrace{\underbrace{x_3}_{\substack{\text{optimiser type} \\ \in \{SGD, Adam, \ldots\}}}, \underbrace{x_4}_{\substack{\text{activation type} \\ \in \{tanh, sigmoid, \ldots\}}}}^{\text{categorical variables}})$$

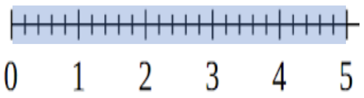
Multiple categorical - each categorical has multiple options

Models to handle mixed variables

Hybrid variables

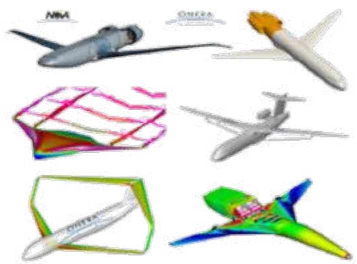
Variables types:

- Continuous (x) Ex: wing length



- Integer (z) Ex: winglet number

- Categorical (u) Ex: Plane shape / material properties


[1]

Categorical variables: n variables, n=2
u1= shape
u2= color

Levels: L_i levels for I in 1,..n, $L_1=3$, $L_2=2$
Levels(u1)= square, circle, rhombus
Levels(u2)= blue, red

Categories: $\prod_{i=1}^{n} L_i$, 2*3=6
- Blue square
- Blue circle
- Blue rhombus
- Red square
- Red circle
- Red rhombus

6 possibilities

Mixed variables Kriging: $x \in [a,b]$ $x \in \{0,1,\dots,N\}$ $x \in \{blue, red, green\}$

$$k(x,x',\Theta) = k_{con}(x_{con}, x'_{con}, \theta_{con}) k_{int}(x_{int}, x'_{int}, \theta_{int}) \underbrace{k_{cat}(x_{cat}, x'_{cat}, \theta_{cat})}$$

$$\boxed{k_{cat}^{CR}(x_{cat}, x'_{cat}, \theta_{cat}) = \prod_{i=1}^{d_{cat}} k_{cont,i}(e_{x_{cat,i}}, e_{x'_{cat,i}}, \theta_{cat,i})}$$

SO RELAXED DIMENSION IS now REALLY

… BIGGER

Encoding x-mixed → continuous space (bigger) x_encoded →Optimze x*_encoded → Decoding x*_engineering

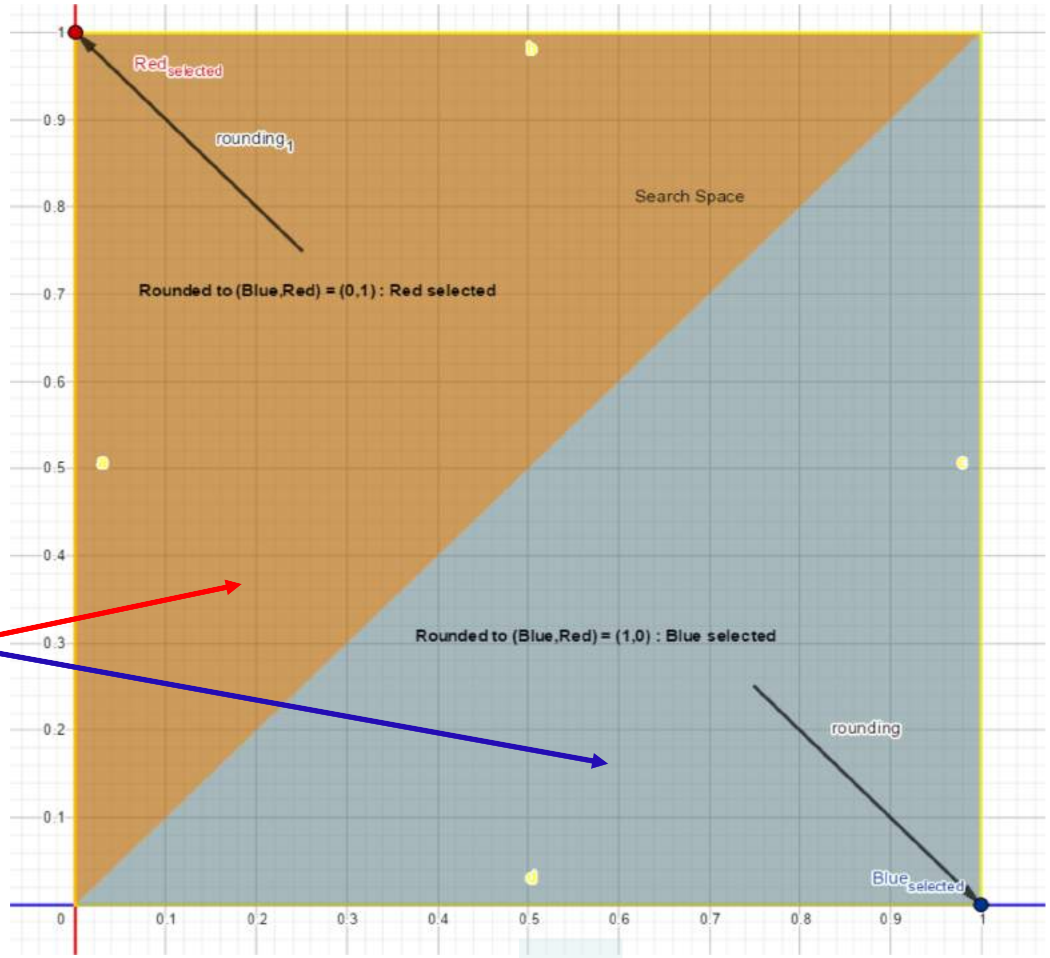
SoA- Continuous relaxation

Example with 1 categorical variable and two levels

- Red color
- Blue color

➔ One-hot encoding: Categorical variable replaced by two continuous variables denoted by X_1 and X_2

- If $X_1 > X_2 \Rightarrow e_{c_1^b} = (1., 0.) \Rightarrow$ Blue color
- If $X_1 < X_2 \Rightarrow e_{c_1^r} = (0., 1.) \Rightarrow$ Red color



$\boldsymbol{n}$ relaxed dimension

$\boldsymbol{x^r}, \boldsymbol{x^s} \in \mathbb{R}^{\boldsymbol{n}}$

A continuous kernel

$$k(x^r, x^s, \theta^{cont}) = \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

E. C. Garrido-Merchán, D. Hernández-Lobato, Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes, 2020, Neurocomputing.

Mixed GP

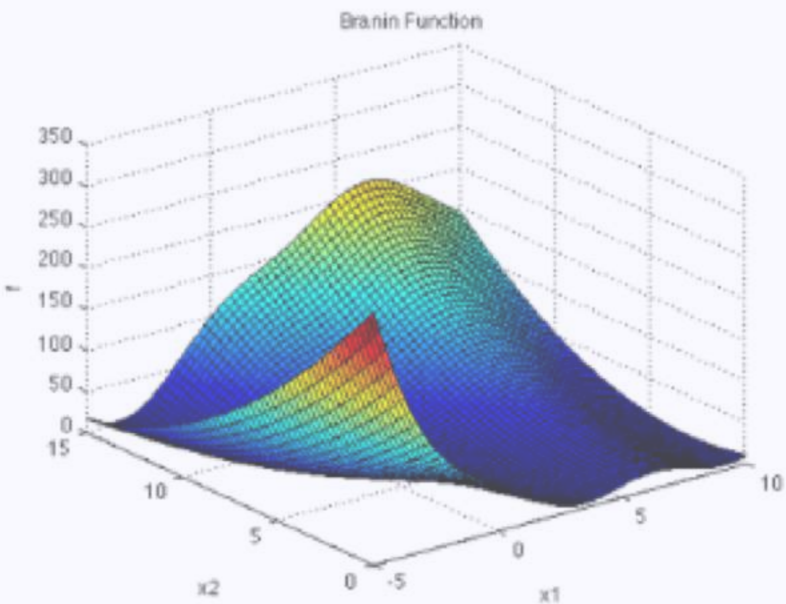
Validation problem n = 2

- **Variable types: continuous and categorical with 10 levels. n' = 11**

$$f(x,z)=\begin{cases}\cos(3.6\pi(x-2))+x-1 & \text{if} \quad z=1,\\ 2\cos(1.1\pi\exp(x))-\frac{x}{2}+2 & \text{if} \quad z=2,\\ \cos(2\pi x)+\frac{1}{2}x & \text{if} \quad z=3,\\ x\left(\cos(3.4\pi(x-1))-\frac{x-1}{2}\right) & \text{if} \quad z=4,\\ -\frac{x^2}{2} & \text{if} \quad z=5,\\ 2\cos(\frac{\pi}{4}\exp(-x^4))^2-\frac{x}{2}+1 & \text{if} \quad z=6,\\ x\cos(3.4\pi x)-\frac{x}{2}+1 & \text{if} \quad z=7,\\ x(-\cos\left(7\frac{\pi}{2}x\right)-\frac{x}{2})+2 & \text{if} \quad z=8,\\ -\frac{x^5}{2}+1 & \text{if} \quad z=9,\\ -\cos(5\frac{\pi}{2}x)^2\sqrt{x}-\frac{\ln(x+0.5)}{2}-1.3 & \text{if} \quad z=10.\end{cases}$$

Branin 2D with

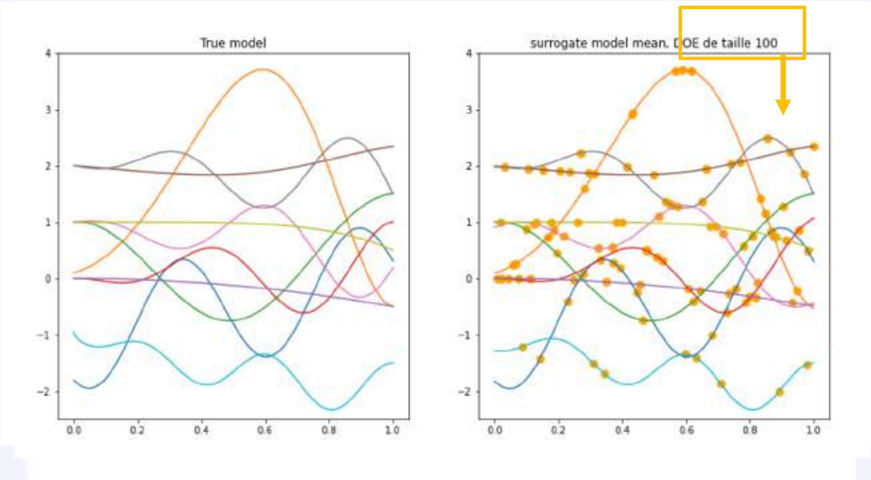
X_1 as an integer in [-5,10] and X_2 continuous in [0,15]



Mixed GP

Validation problem n = 2

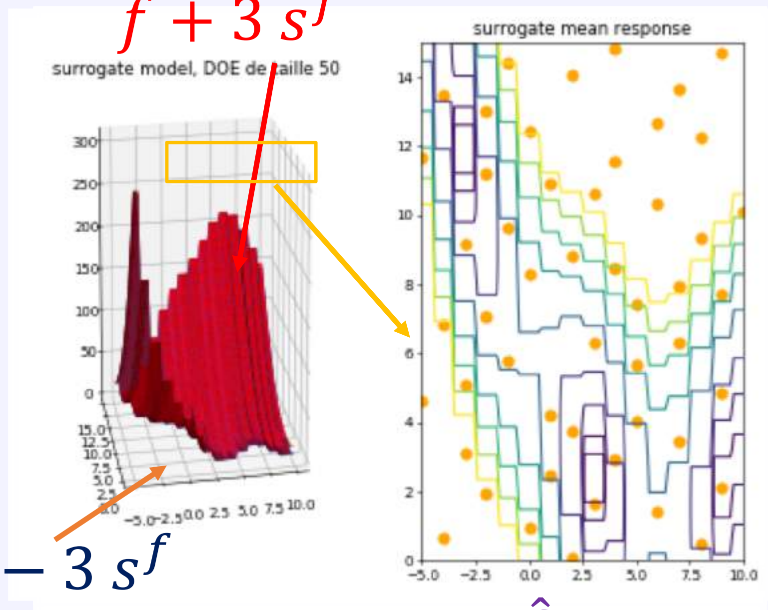
- Variable types: continuous and categorical with 10 levels. n' = 11



Validation problem n = 2

- Variable types: continuous and integer

$\hat{f} + 3\, s^f$



$\hat{f} \; - 3\, s^f$

Integer Branin $\hat{f}$

P. Saves, Y. Diouane, N. Bartoli, T. Lefebvre, J. Morlier, A mixed-categorical correlation kernel for Gaussian process, 2023, Neurocomputing.

Part3: Engineering applications

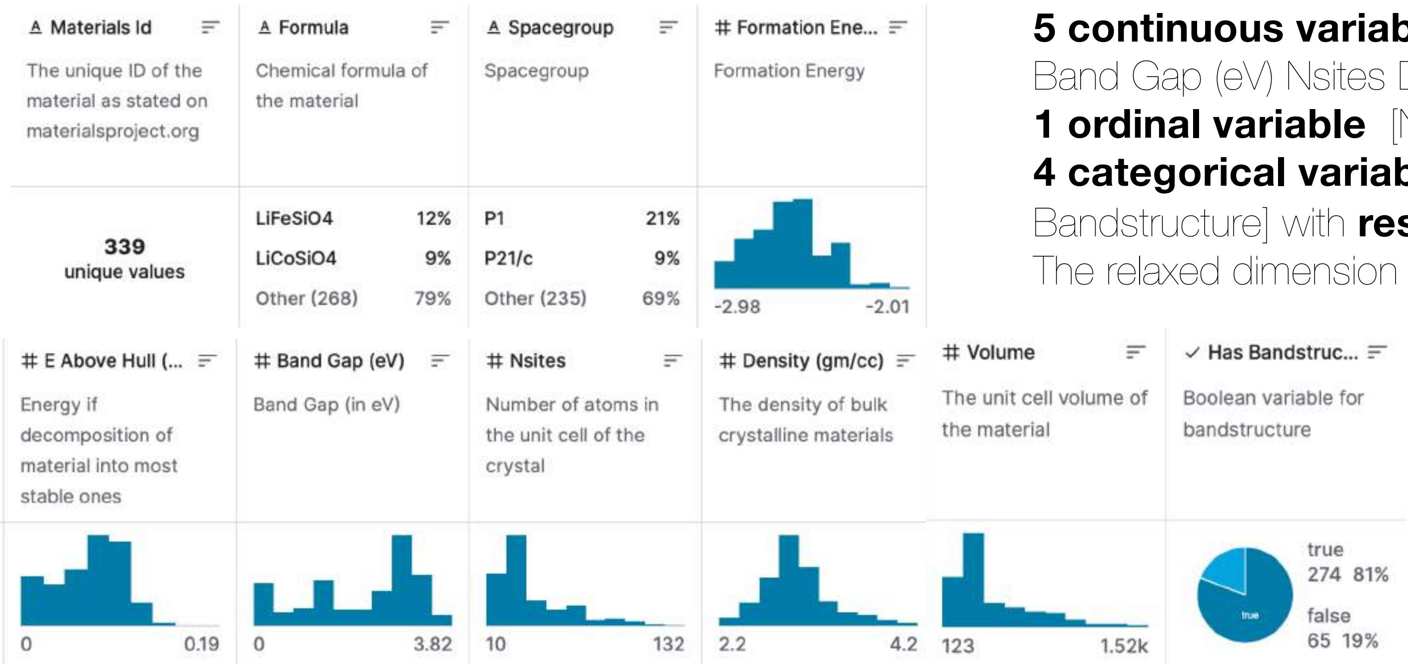
Most of them are opensource

Example: Accelerating Material discovery

Predict y the crystal structure type (monoclinic, orthorhombic, triclinic)

from x Lithium-ion physical and chemical compound information

i.e. learn from learning database y=f(x)



The function inputs are:

5 continuous variables [Formation Energy (eV) E Above Hull (eV) Band Gap (eV) Nsites Density (gm/cc) Volume]

1 ordinal variable [Nsites]

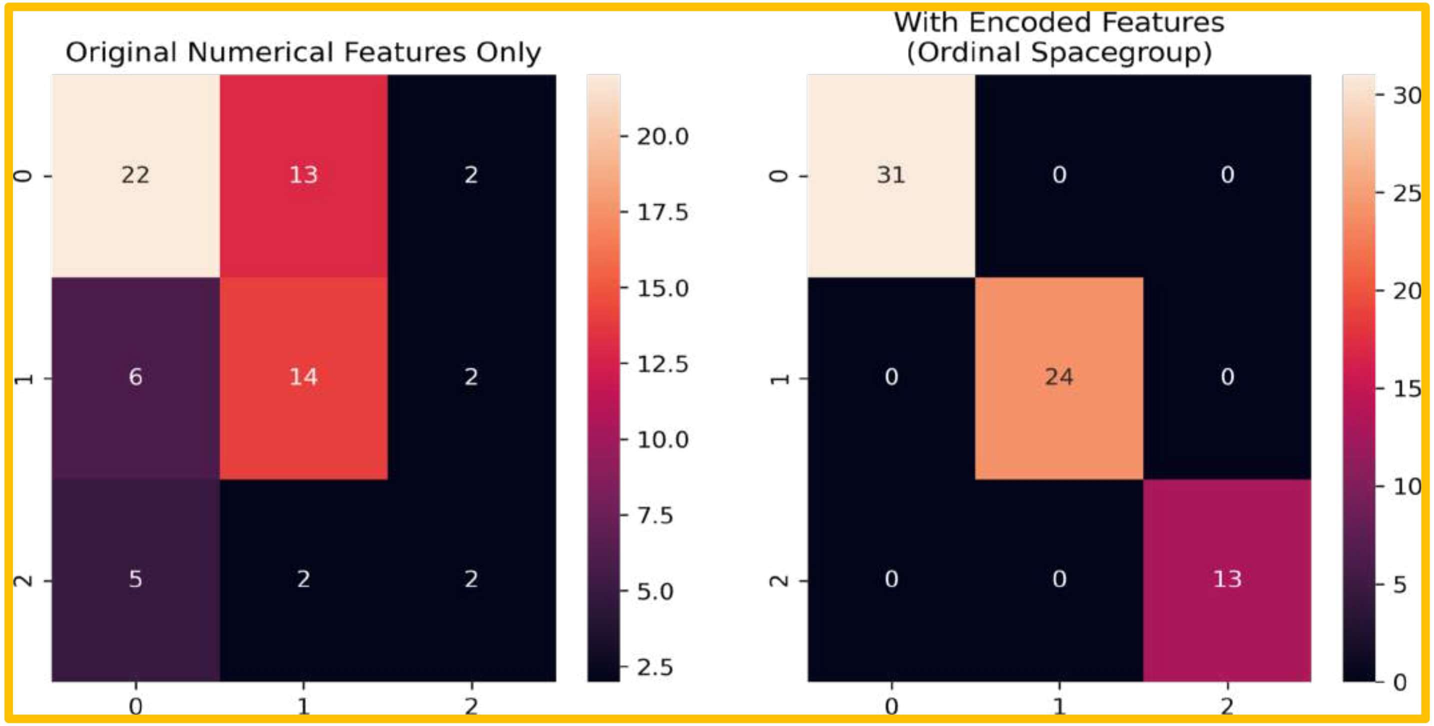
4 categorical variables [Materials Id Formula Spacegroup Has Bandstructure] with **respectively: 339, 114 and 44 levels.**

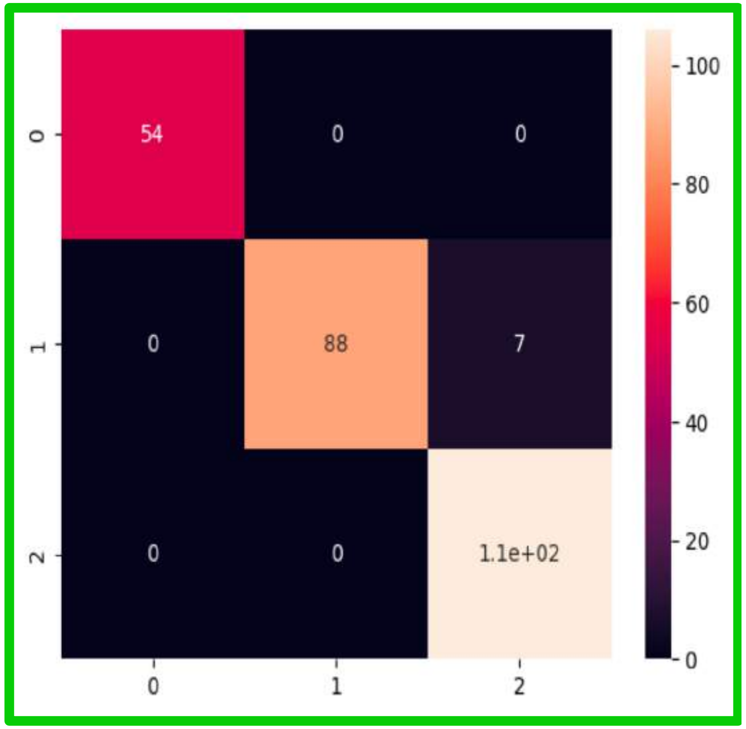
The relaxed dimension by **one-hot-encoding is 503**

… REALLY HIGH DIMENSION

Agrawal, D., Crystal System Properties for Li-ion batteries, Properties of Li-ion silicate to predict the crystal system class of the battery, Kaggle, March 2020.

Results: Accelerating Material discovery

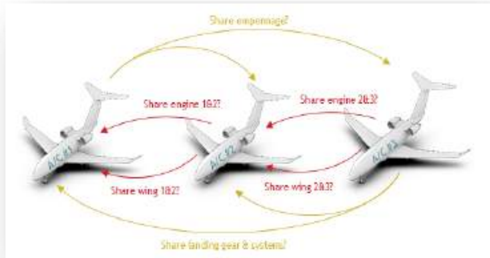




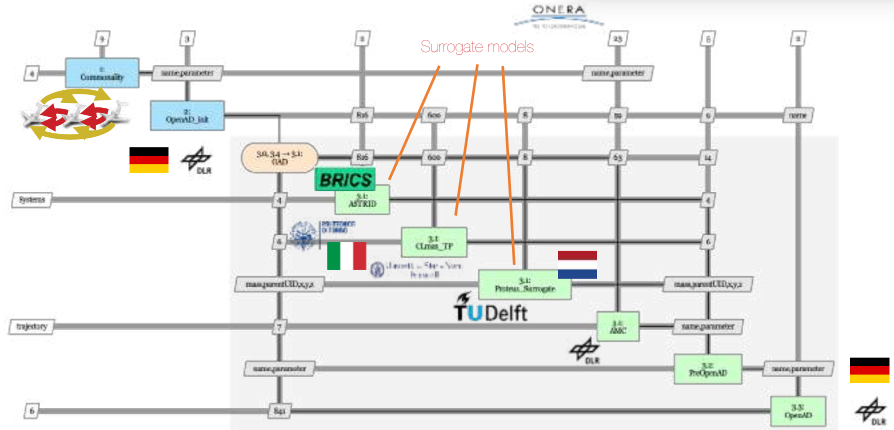
scikit-learn with <u>80/20</u>
dtree w/wo specific features (*sf*)

SMT with <u>10/90 (!!)</u> wo sf

Example in BO with hierarchical variables

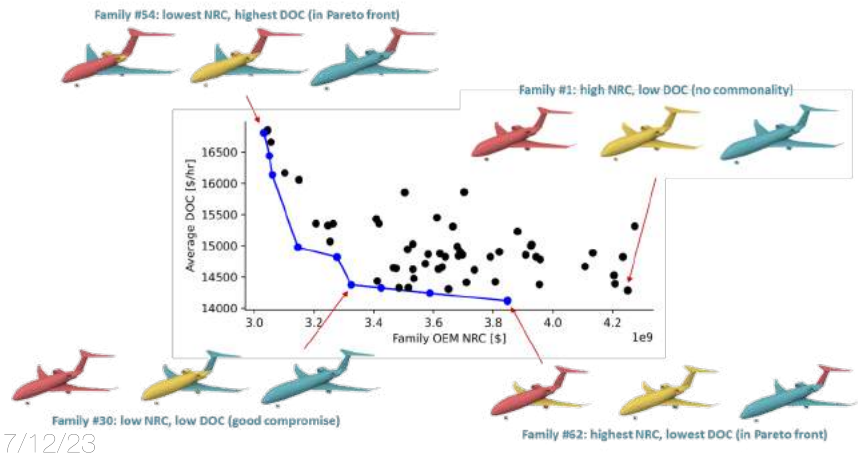


Select the best family with commonality choices (wing, engine, empennage)	
Objectives	**2 obj: Min (Direct Operating Costs, OEM Non – Recurring Costs)**
Design variables	**10 categorical: commonality with 2 levels 9 continuous: Leading edge sweep, rear spar location, Wing t/c for each family member**
Constraints	**2 ineq. : Balanced Field Length, Landing Field Length**



1. Handling of hierarchical variables using imputation method (mean value used when the var. is inactive)
2. DOE and offline optimization to fine tune the optimization parameter

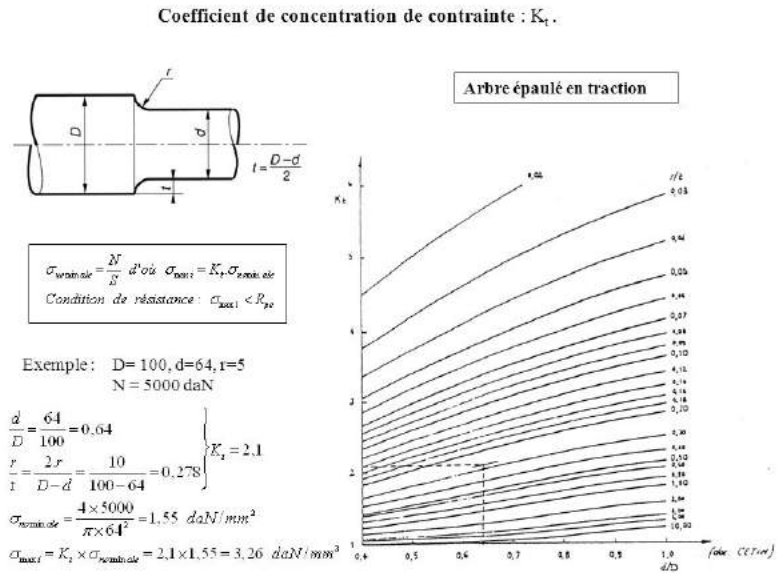
AGILE 4.0



Conclusions

Surrogate is the new abacus

and more



Conclusions

- SMT 1.0: A Python surrogate modeling framework **with derivatives** (2019)
- SMT 2.0: A Surrogate Modeling Toolbox with a focus on Hierarchical and **Mixed Variables** Gaussian Processes (2023)
- Explainability, . . .

https://doi.org/10.1016/j.advengsoft.2019.03.005

https://doi.org/10.1016/j.advengsoft.2023.103571

Documentation:
https://smt.readthedocs.io/en/latest/

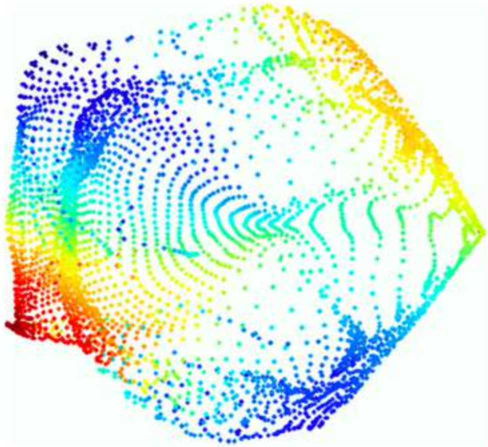
Code:
https://github.com/SMTorg/SMT

How to start:
https://github.com/SMTorg/smt/tree/master/tutorial#readme

Open in Colab

Mixed Optimization in 200 dimensions?

- High dimension causes problem for optimization.
  - Statistical challenge: the search space grows exponentially
  - Computational challenge: global optimizers fail to return an optimum within limited time and resource



--> refactoring & opensourcing our Constrained Bayesian Optimization code based on SMT called SEGOMOE

GPyOpt
Tune your algorithms and your design wetlab experiment

botorch
Bayesian optimization in PyTorch (by pytorch)

Ax
Adaptive Experimentation Platform (by facebook)

scikit learn

https://scikit-learn.org/stable/modules/gaussian_process.html

https://scikit-optimize.github.io/stable/auto_examples/bayesian-optimization.html

SEGOMOE is available Through a webservice
https://github.com/whatsopt/wopsego

joseph.morlier@isae-supaero.fr
nathalie.bartoli@onera.fr

MERCI

ONERA: P. Saves and R. Lafage and N. Bartoli , T. Lefevbre, R. Charayron…

ISAE-SUPAERO: R. Rufato Careira, P. Saves, R. Charayron

Polytechnique Montréal: Y. Diouane

DLR: J. H. Bussemaker

UCSD: J. T. Hwang

UoM: J. R. R. A. Martins

SMT 1.0 coding due to Mohamed Amine Bouhlel (INTEL)