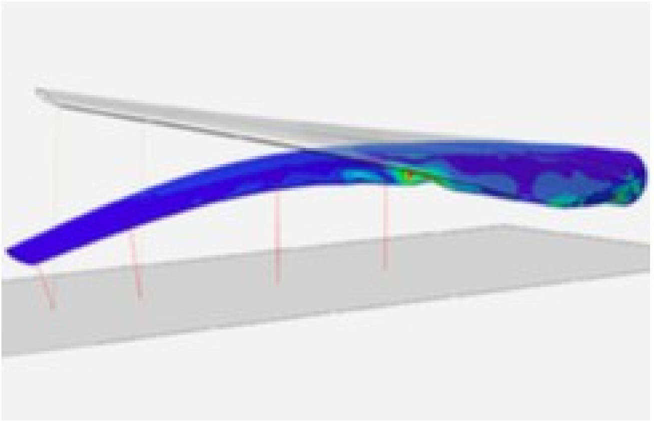
COMPUTATIONAL METHODS FOR UNCERTAINTIES IN FLUIDS AND ENERGY SYSTEMS



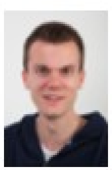
- ✓ Development of efficient **UQ**-forward propagation method
- ✓ **Data-inferred stochastic modeling** when a limited amount of data is available for unsteady non-linear systems
- ✓ **Numerical simulation of fluids (CFD)** for energy application (high number of uncertainties, optimization under uncertainty, etc)

CWI Scientific Computing Group

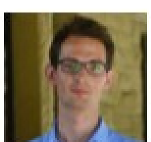
Daan Crommelin
COMMUNES Team Leader

Benjamin Sanderse
Research Scientist, CWI

Anne Engels
PhD Candidate, CWI



Laurent van den Bos
PhD Candidate, CWI



Yous van Halder
PhD Candidate, CWI

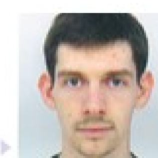
DeFI Team (INRIA SIF, Ecole Polytechnique)

Pietro M Congedo
COMMUNES Team Leader

Olivier Le Maitre
Research Scientist, CNRS

Nassim Razaaly
PhD Candidate, CWI-INRIA

Francois Sanson
PhD Candidate, INRIA



Mickael Rivier
PhD Candidate, INRIA

Some main activities

- Workshop of COMMUNES Team, December 3-4 at CWI, 11 participants.
- Several visits (to INRIA and CWI)
- MS Organization at SIAM UQ 2018
- Joint Paper at Eccomas Conference 2018

- International Conference UQOP on UQ and Optimization, Paris 2019

Clustering-based UQ

Anne Eggels

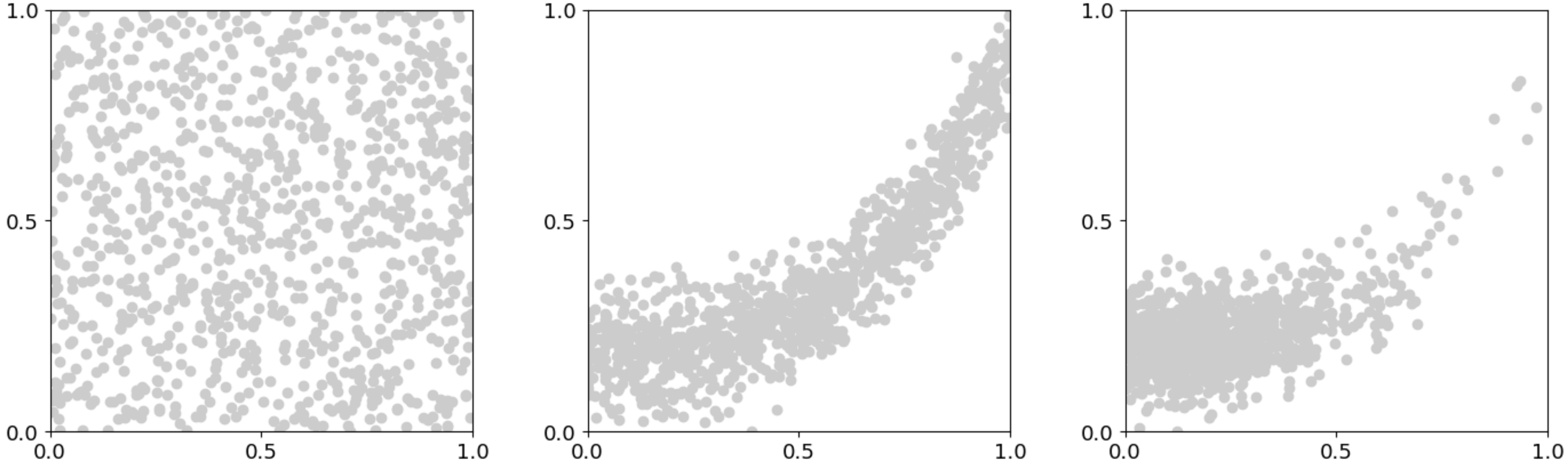
September 26th, 2018

NWO Applied and Engineering Sciences

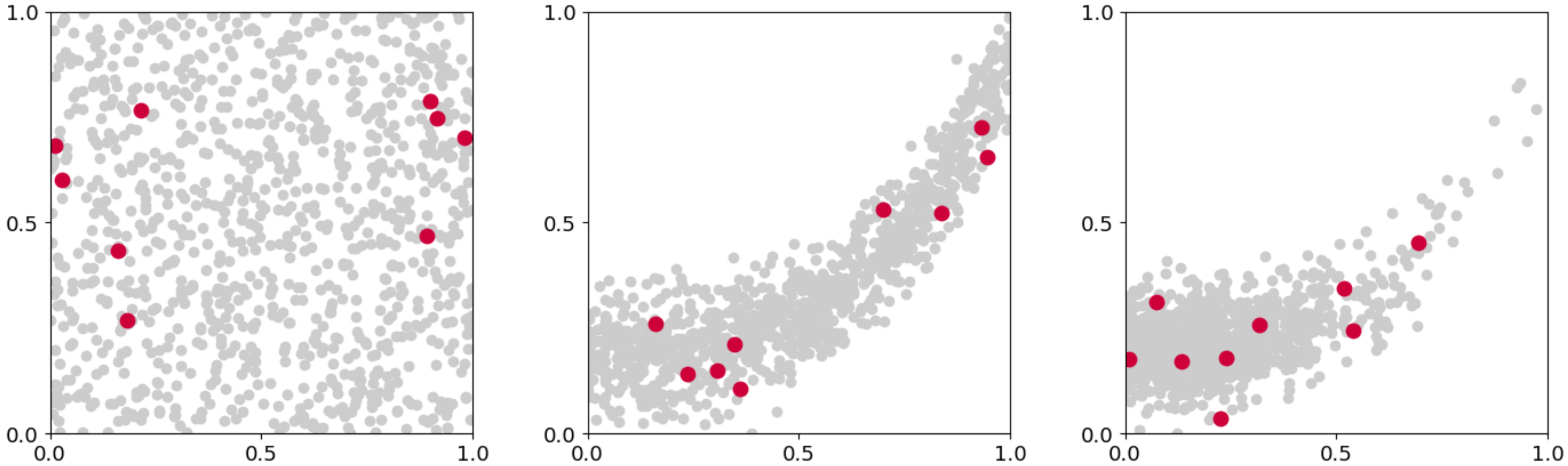
CWI

Uncertainty Quantification

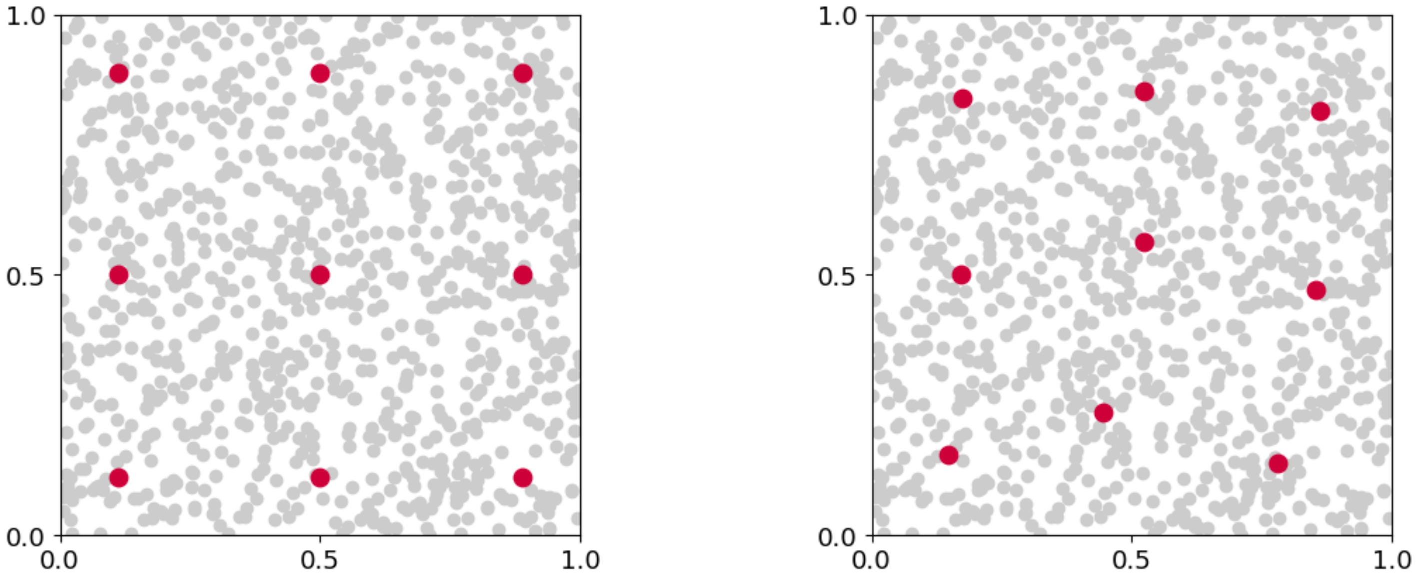
Different types of data



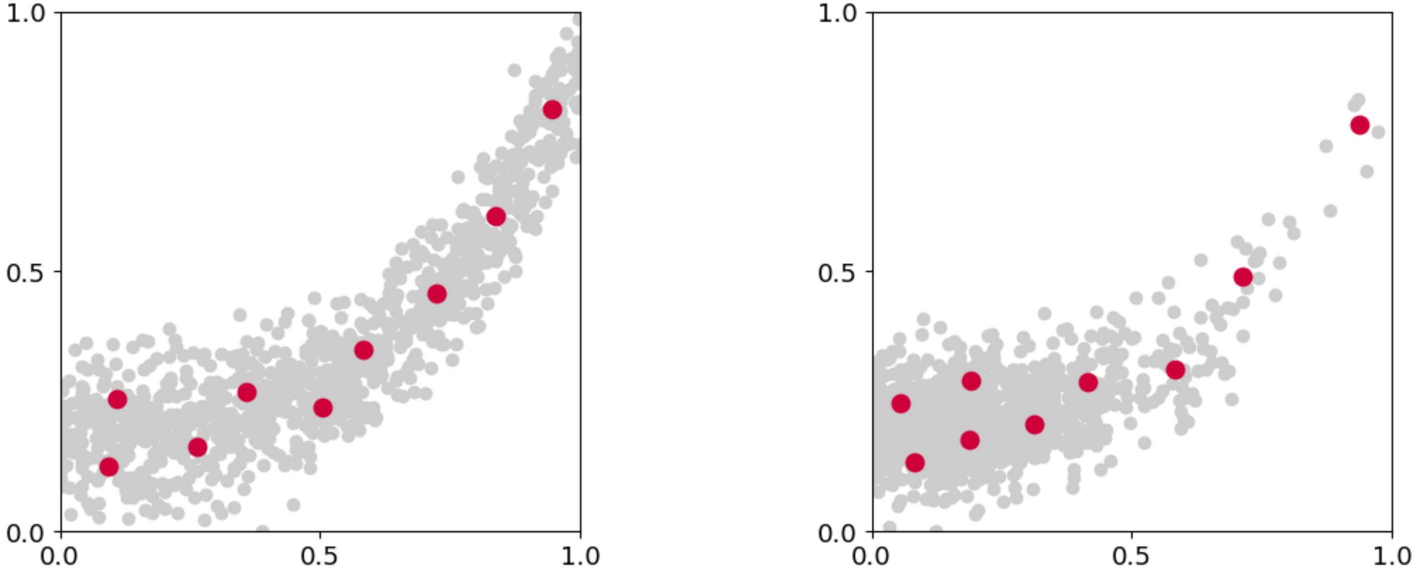
Monte Carlo sampling



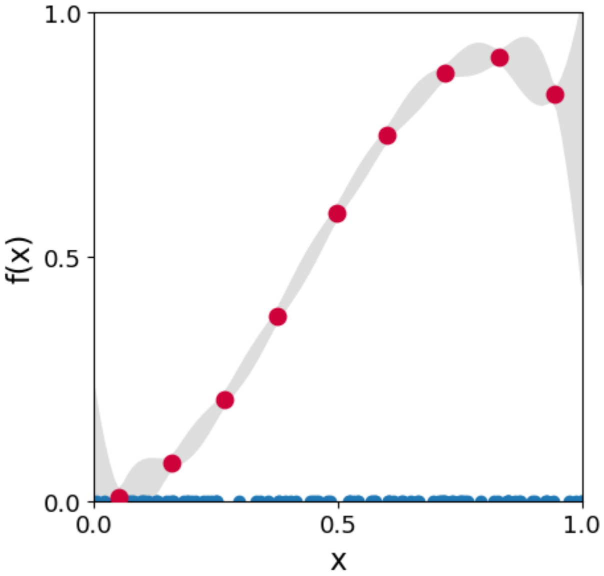
Smart sampling - independent data

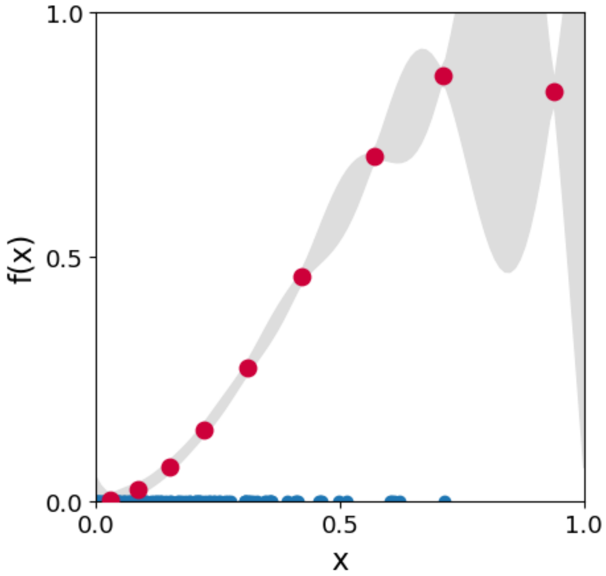


Smart sampling - *k*-means



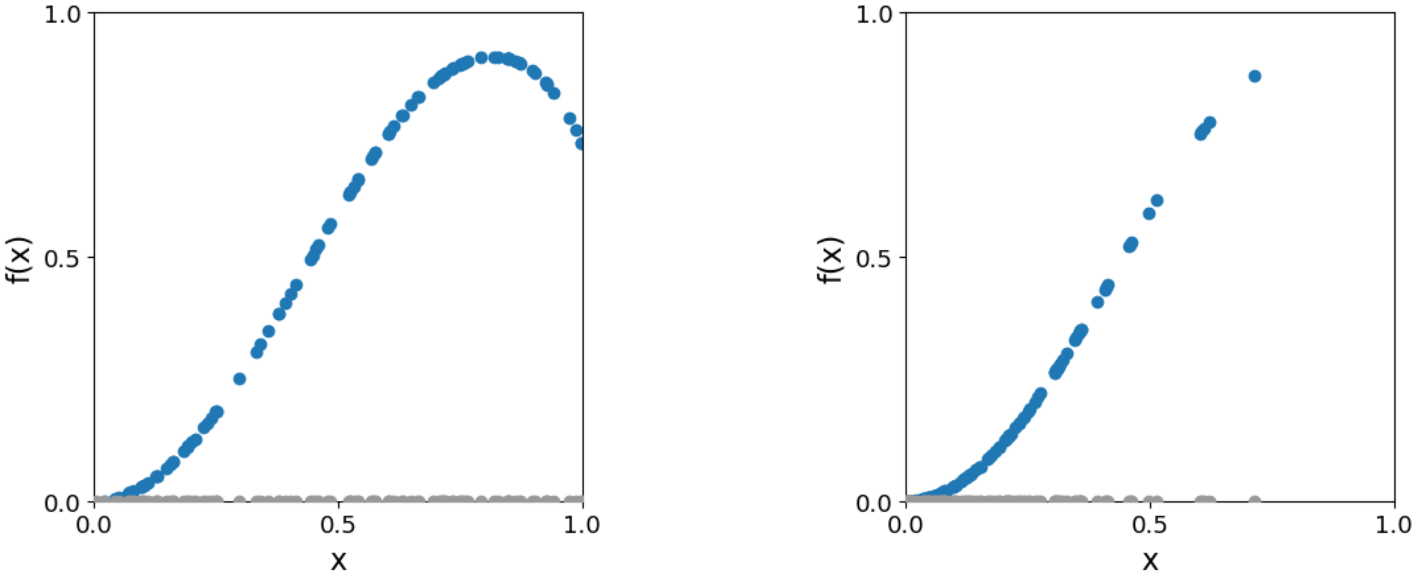
Gaussian processes (GPs)





CWI

Gaussian processes (GPs)



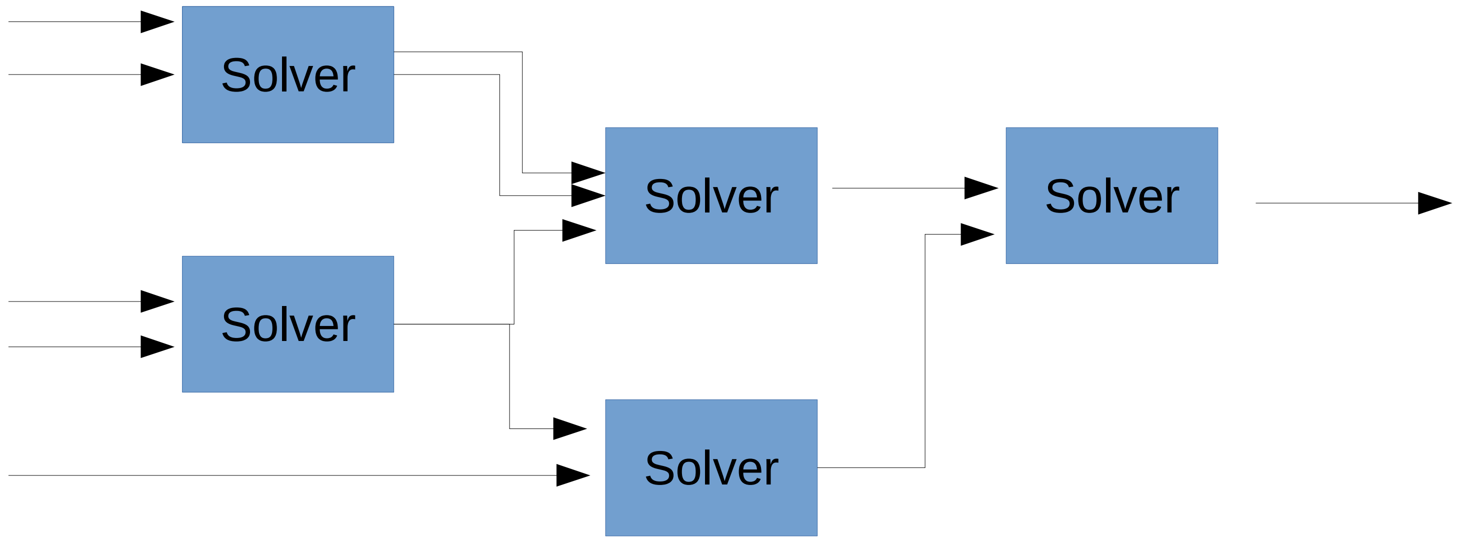
Work in progress

- GPs have length scales
- Rate of variance growth for each variable
- Combine the GP length scales with *k*-means to improve the prediction

Efficient clustering based training set generation for system of solvers

Francois Sanson (Inria BSO)
Anne W. Eggels (CWI)
Olivier Le Maitre (LIMSI)
Dann Crommelin (CWI)
Pietro M. Congedo (Inria Saclay)

Context : forward propagation of uncertainties in a system of solvers



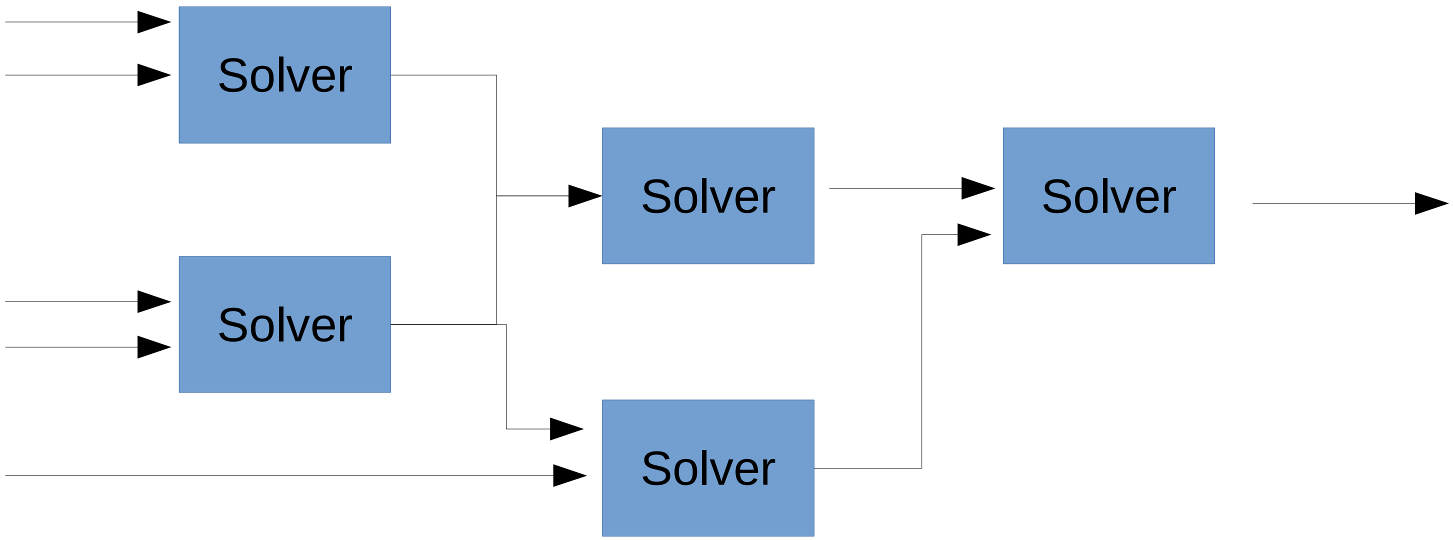
Example : Tsunami simulation, Space object reentry

How to propagate uncertainties through a system of solvers at minimal computational cost ?

How to propagate uncertainties through a system of solvers at minimal computational cost ?

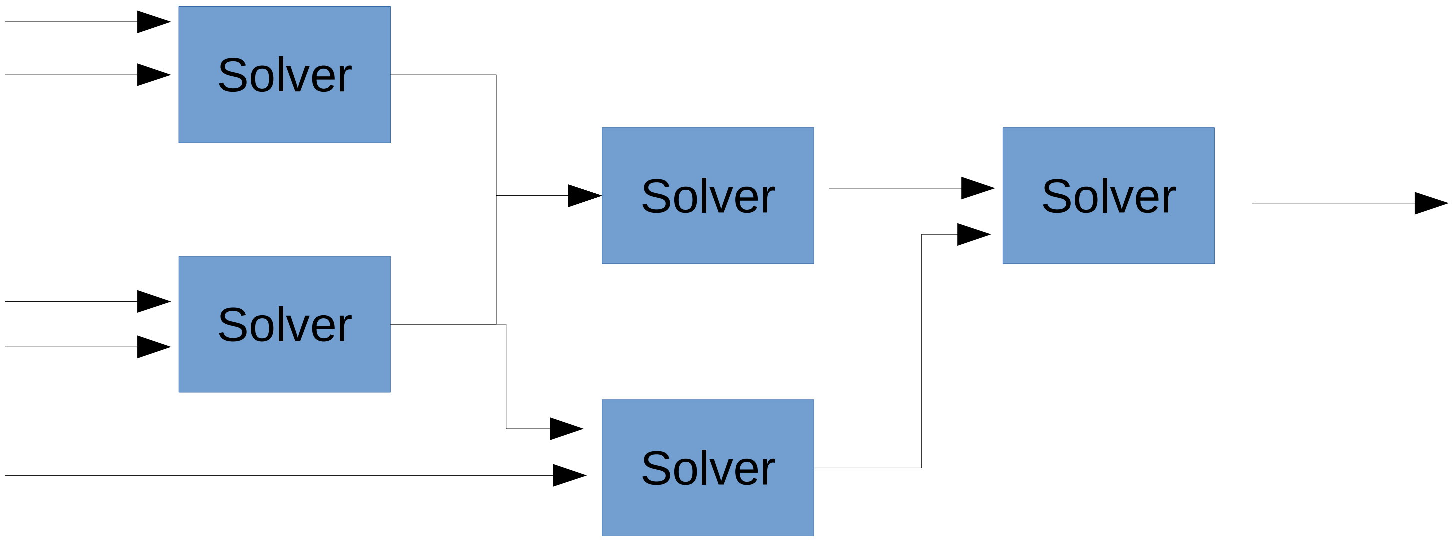
Build a cheap surrogate of the system and use Monte Carlo methods

A naive strategy : the black box approach



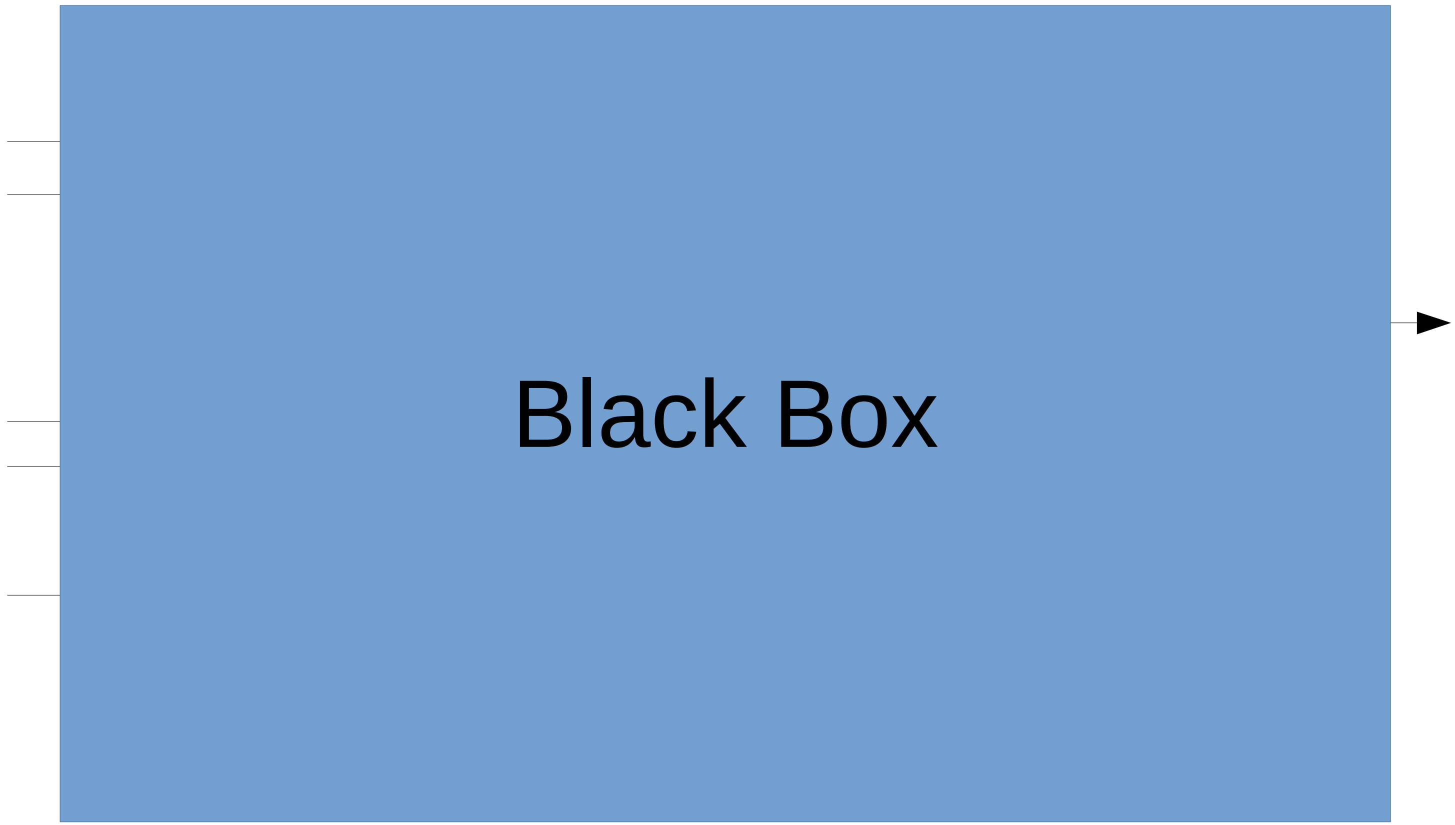
Computationally expensive and not optimal

A naive strategy : the black box approach



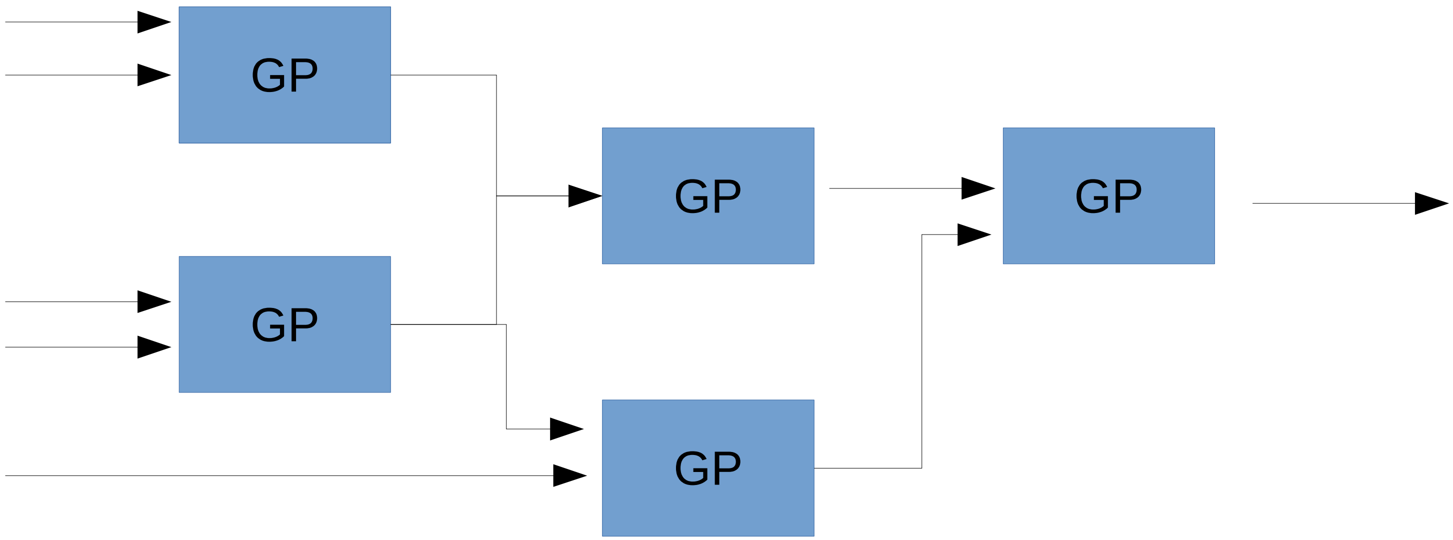
Computationally expensive and not optimal

A naive strategy : the black box approach

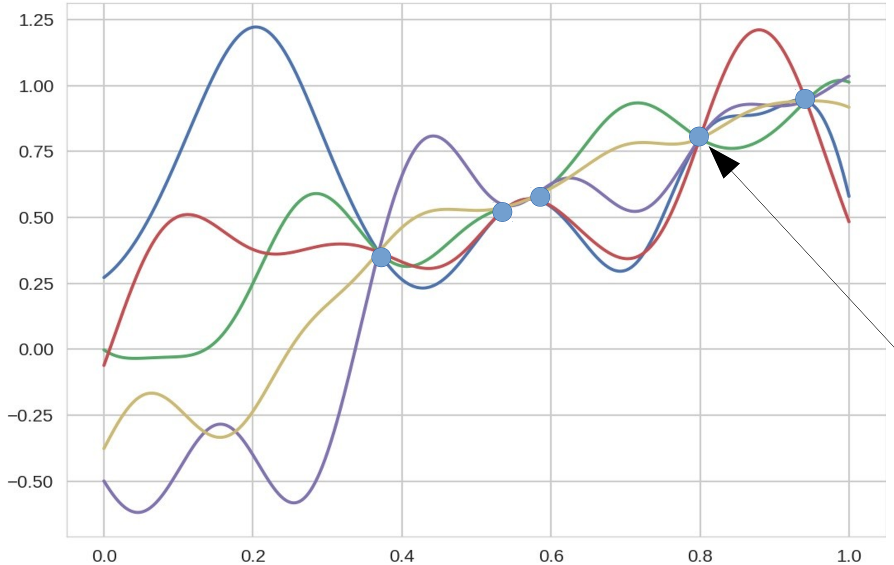


Computationally expensive and not optimal

Our strategy : systems of Gaussian Processes



Predictions with Gaussian processes



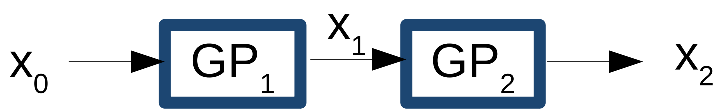
The first two moments can be computed analytically

$$\mu(x) = \mathbf{k}_{\mathcal{A}}(x) K_{\mathcal{A}}^{-1} \mathbf{y}_{\mathcal{A}},$$
$$\sigma^2(x) = k(x,x) - \mathbf{k}_{\mathcal{A}}(x) K_{\mathcal{A}}^{-1} \mathbf{k}_{\mathcal{A}}(x)^T.$$

Training evaluations

Define a distribution of possible functions given observations

Mean predictions with SoGPs



- Composition of average :

$$f_2 \circ f_1(x_0) \approx \mu_2(\mu_1(x_0)) = \mu_2 \circ \mu_1(x_0).$$

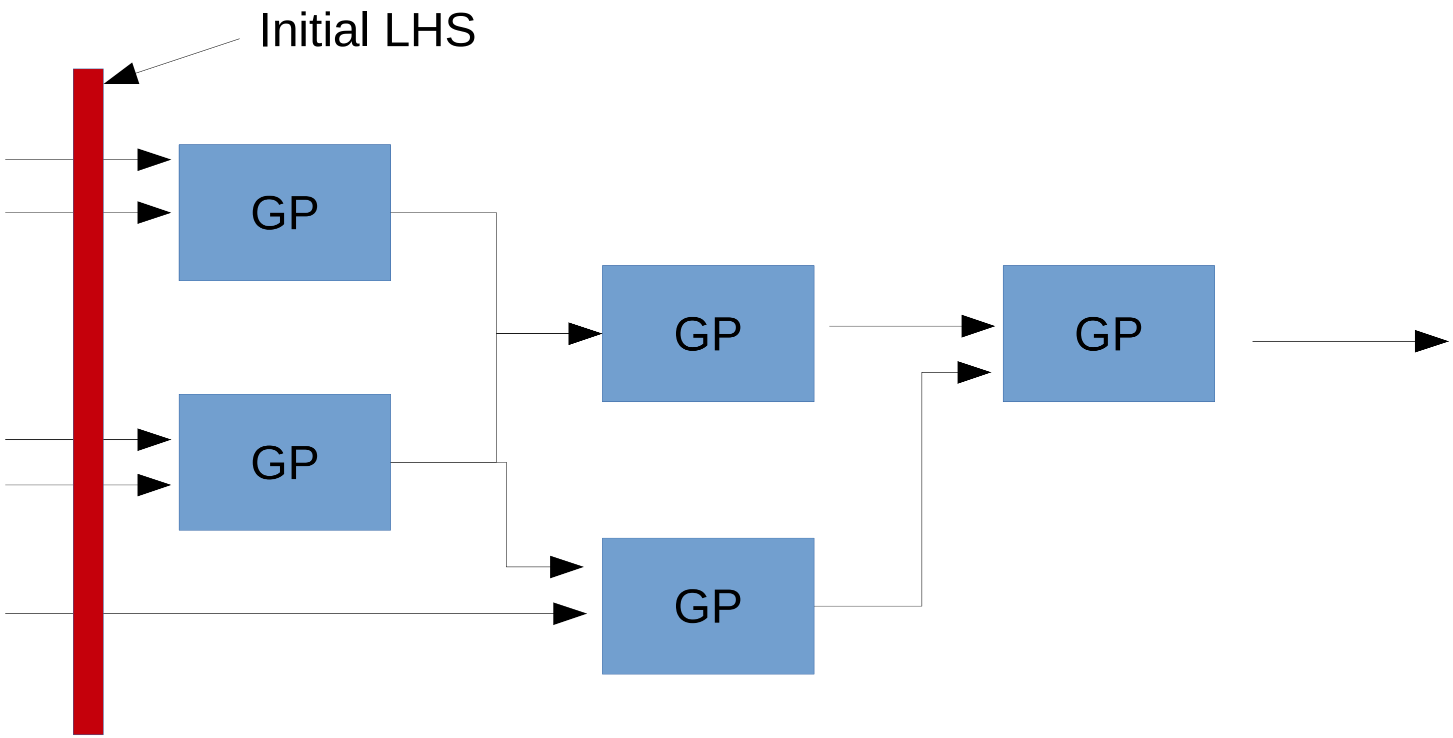
- Average of composition :

$$f_2 \circ f_1(x_0) \approx E\left[\mathrm{G}_2 \circ \mathrm{G}_1(x_0)\right] = \frac{1}{\sqrt{2\pi\sigma_1^2(x_0)}} \int \mu_2(x_1) \exp\left(-\frac{(x_1 - \mu_1(x_0))^2}{2\sigma_1^2(x_0)}\right) dx_1.$$

The integral cannot be computed analytically and have to be approximated (MC,inducing points, Taylor series expansion)

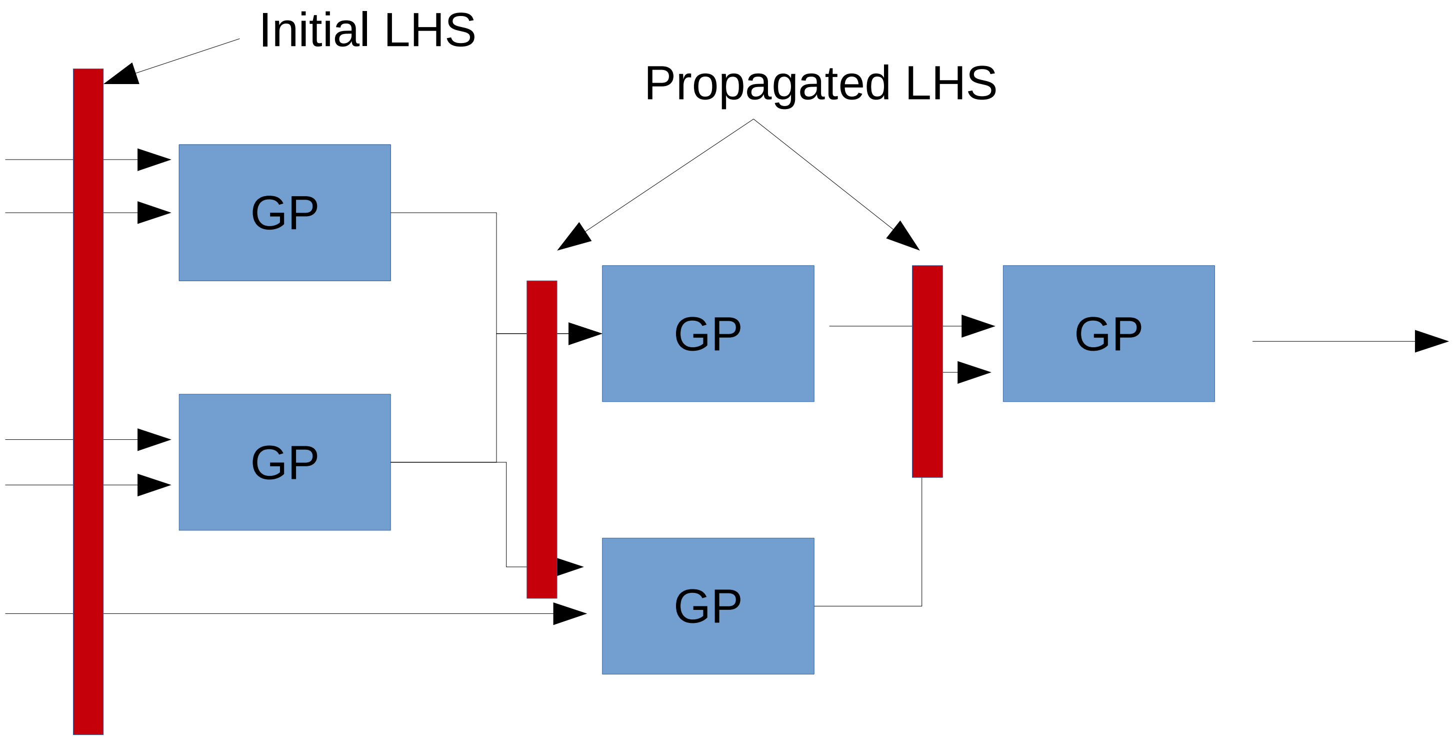
Generating training sets for SoGPs

LHS training



LHS are smart techniques for independent variables, are those properties propagated through the layers ?

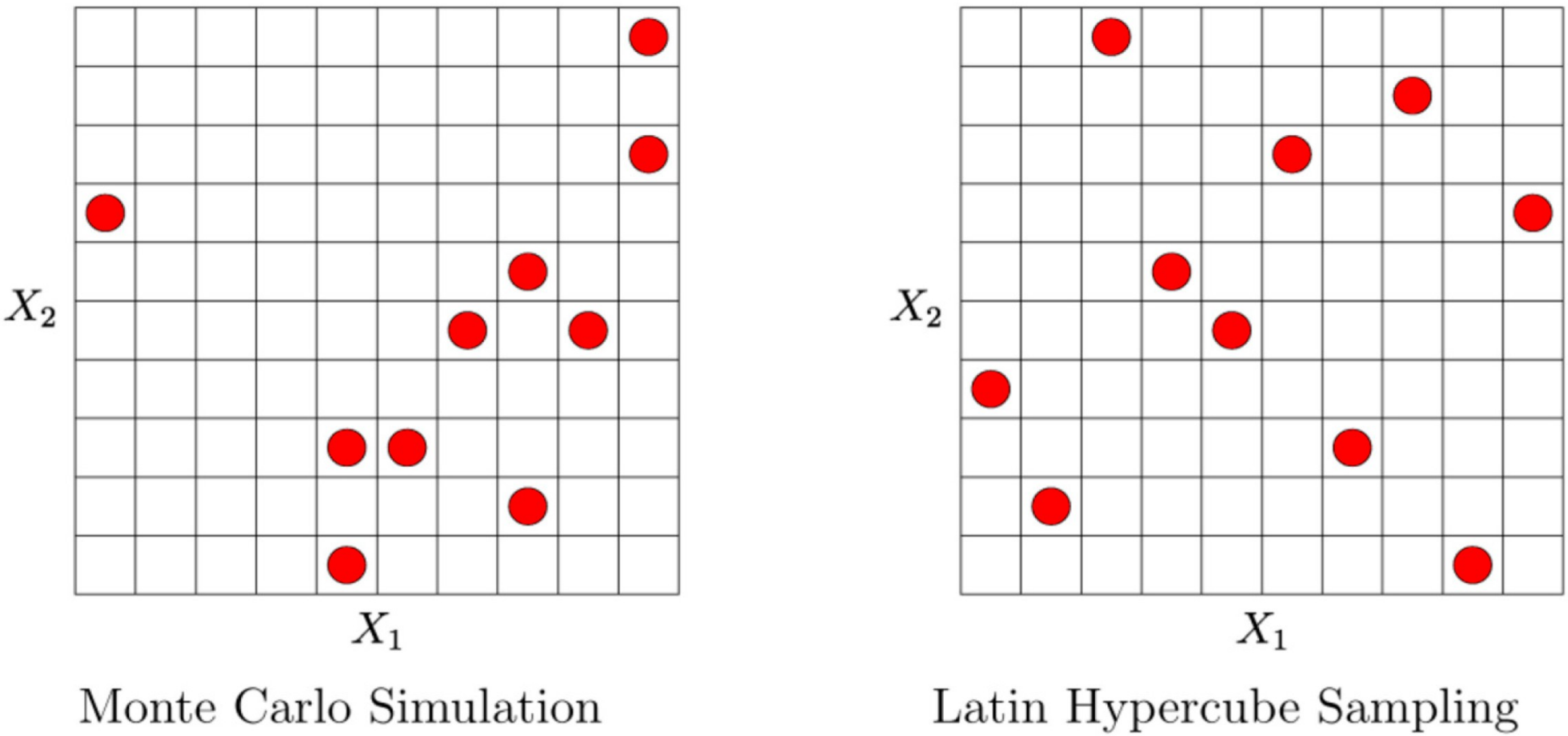
LHS training



LHS are smart techniques for independent variables, are those properties propagated through the layers ?

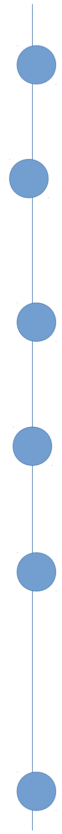
Latin Hyper Cube sampling

No shared rows and columns



Taken from Schultze, V.; Schillig, B.; IJsselsteijn, R.; Scholtes, T.; Woetzel, S.; Stolz, R. An Optically Pumped Magnetometer Working in the Light-Shift Dispersed Mz Mode. Sensors 2017, 17, 561.

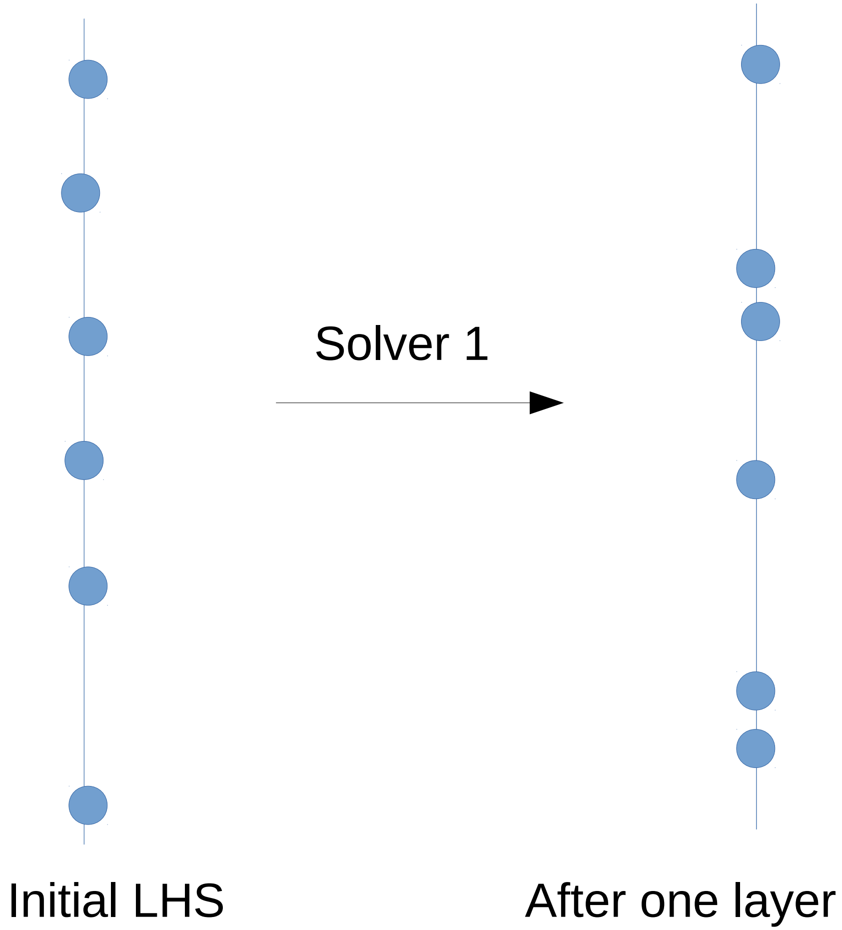
Limitation of LHS training



Initial LHS

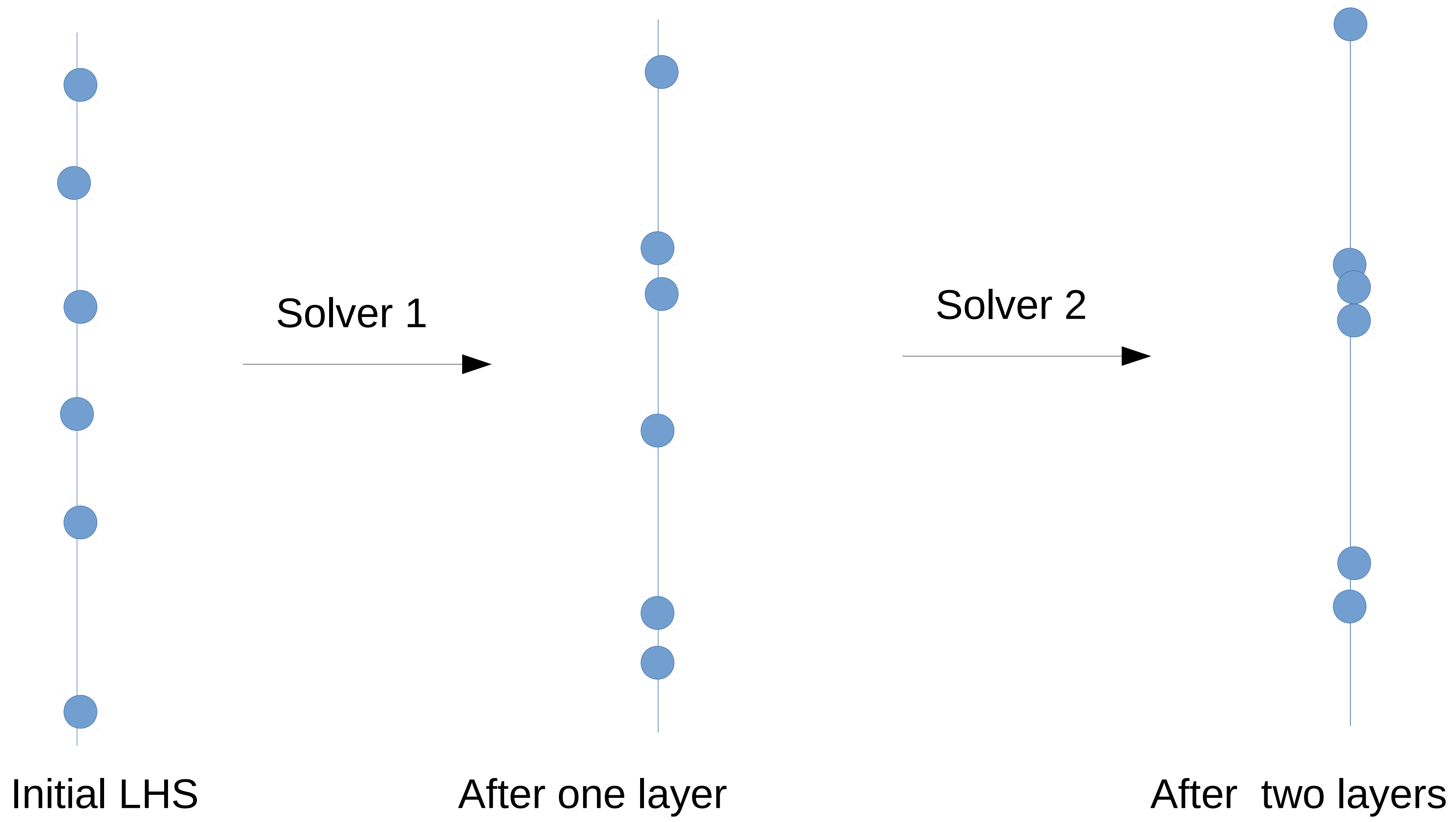
The good properties of LHS : no alignments, good coverage are not assured after propagation

Limitation of LHS training



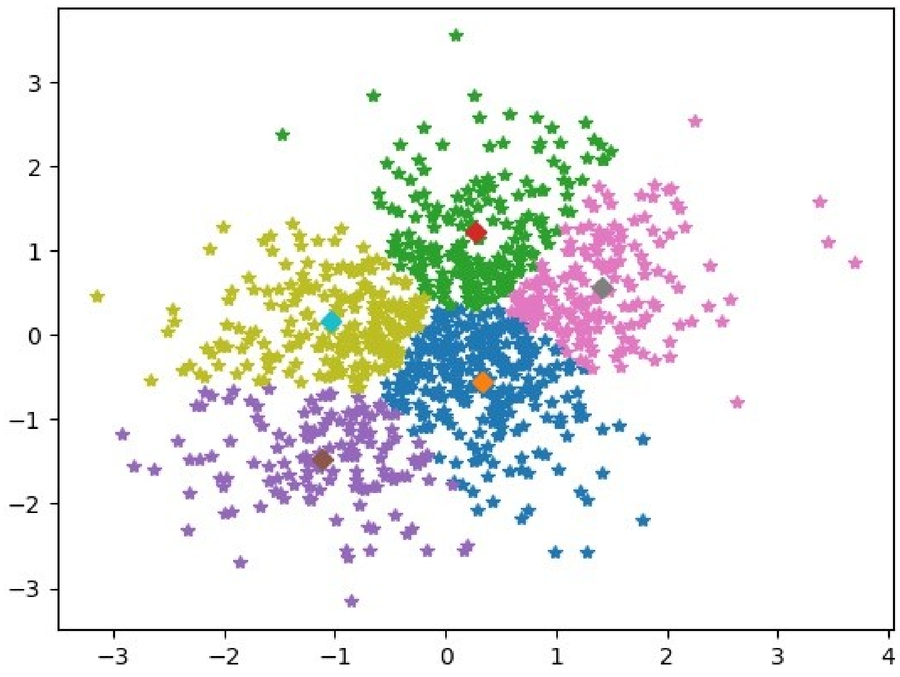
The good properties of LHS : no alignments, good coverage are not assured after propagation

Limitation of LHS training



The good properties of LHS : no alignments, good coverage are not assured after propagation

Use clustering to resample the training set at each layer



Local distorsion :

$$d(X_t, x) = \min_{x' \in X_t} ||x' - x||$$

Distorsion :

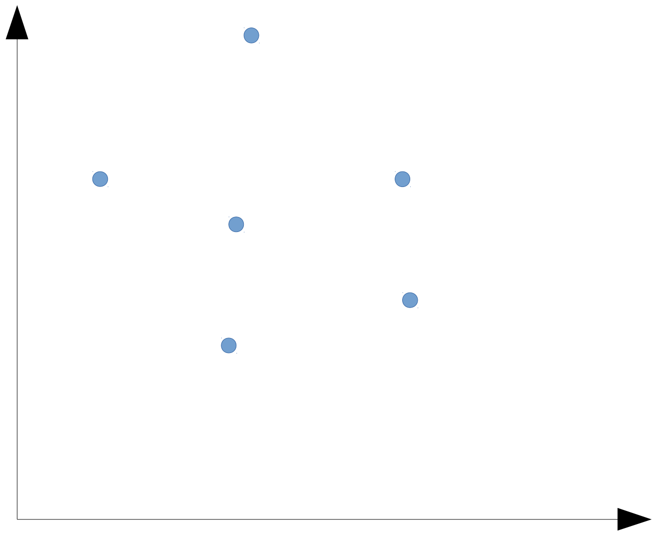
$$\tilde{D}(X_t) = \frac{1}{m} \sum_{x \in X_d} d(X_t, x)$$

Kmeans minimizes for X_t the distorsion

Lekivetz, R., and Jones, B. (2015) Fast Flexible Space-Filling Designs for Nonrectangular Regions. Qual. Reliab. Engng. Int., 31: 829–837. doi: 10.1002/qre.1640.

Lloyd algorithm for cluster construction

1. Initialize the centroids possibly randomly $X_t = (x_{c_1}, ...x_{c_{nt}})$



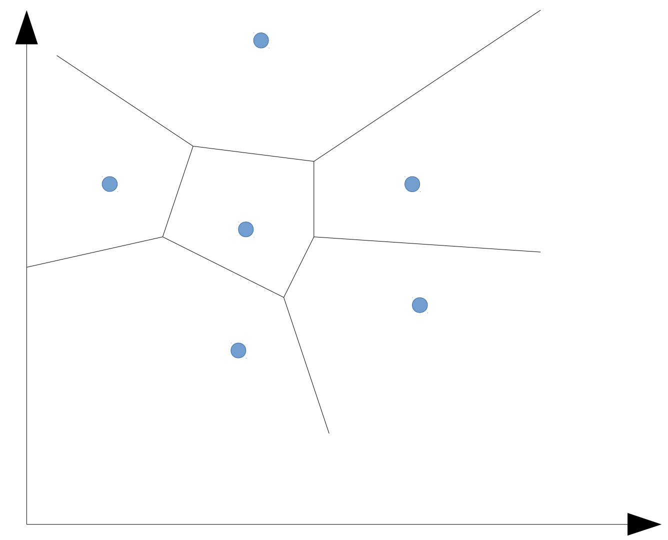
Lloyd algorithm for cluster construction

1. Initialize the centroids possibly randomly

$$X_t = (x_{c_1}, ...x_{c_{nt}})$$

2. Assign each sample of the dataset to its closest centroid to create clusters

$$C_i = \{x : d(x, Xt) = ||x - x_{C_i}||\}$$



Lloyd algorithm for cluster construction

1. Initialize the centroids possibly randomly

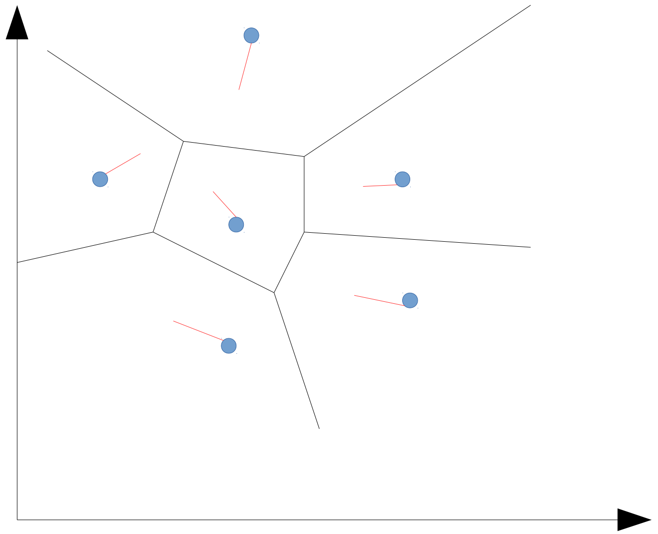
$$X_t = (x_{c_1}, ...x_{c_{nt}})$$

2. Assign each sample of the dataset to its closest centroid to create clusters

$$C_i = \{x : d(x, Xt) = ||x - x_{C_i}||\}$$

3. Redefine the centroids as barycenters of each cluster

$$x_{c_1} = \frac{1}{card(C_i)} \sum_{x \in C_i} x$$



Lloyd algorithm for cluster construction

1. Initialize the centroids possibly randomly

$$X_t = (x_{c_1}, ... x_{c_{nt}})$$

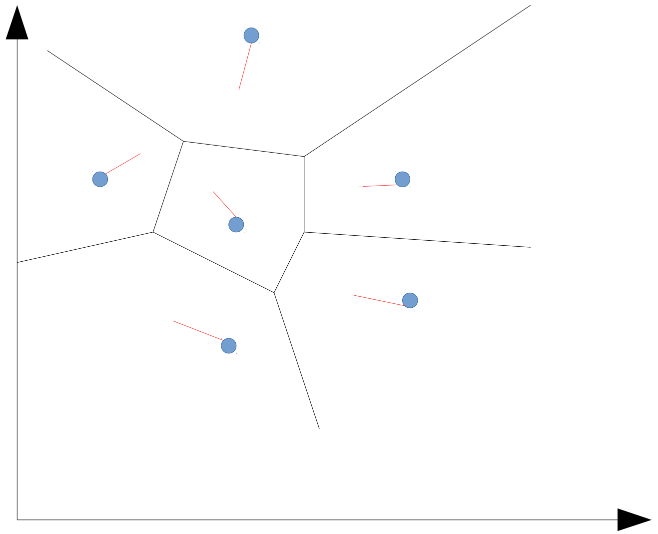
2. Assign each sample of the dataset to its closest centroid to create clusters

$$C_i = \{x : d(x, Xt) = ||x - x_{C_i}||\}$$

3. Redefine the centroids as barycenters of each cluster

$$x_{c_1} = \frac{1}{card(C_i)} \sum_{x \in C_i} x$$

4. Iterate 2 and 3 until convergence



SoGP training with clustering

1. Generate an initial LHS training set

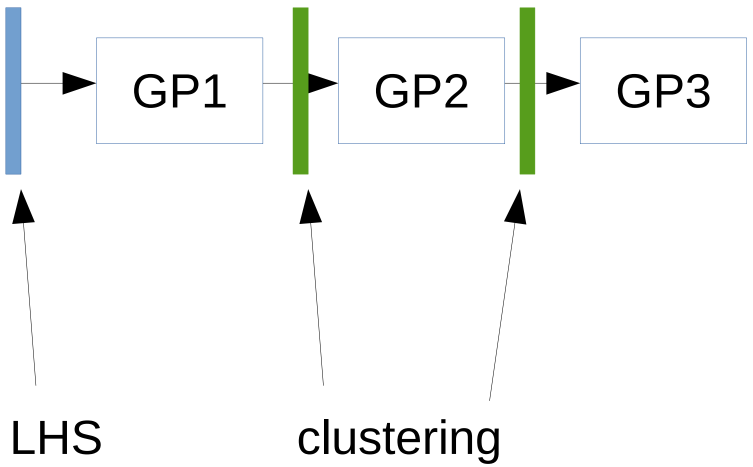
2. Use it to train GP 1

3. Propagate a large number of samples using GP1

4. Use the samples to run a clustering algorithm

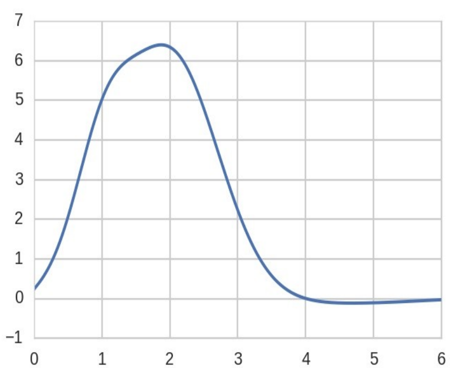
5. Use the centroids of each cluster as training points of GP2 (**alternatively one random sample from each cluster**)

6. Repeat 2 to 5 for GP3

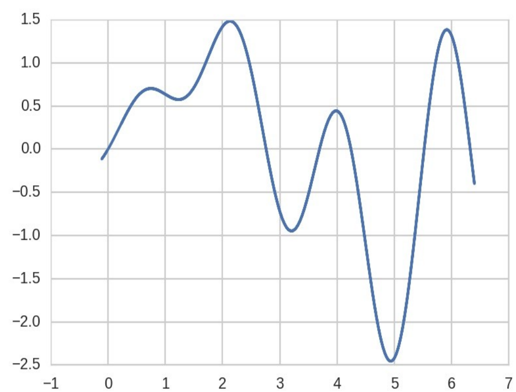


Numerical tests

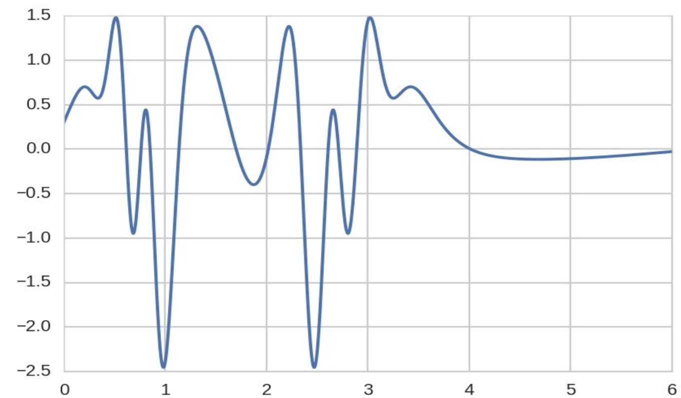
1D test



Solver 1 (f_1)



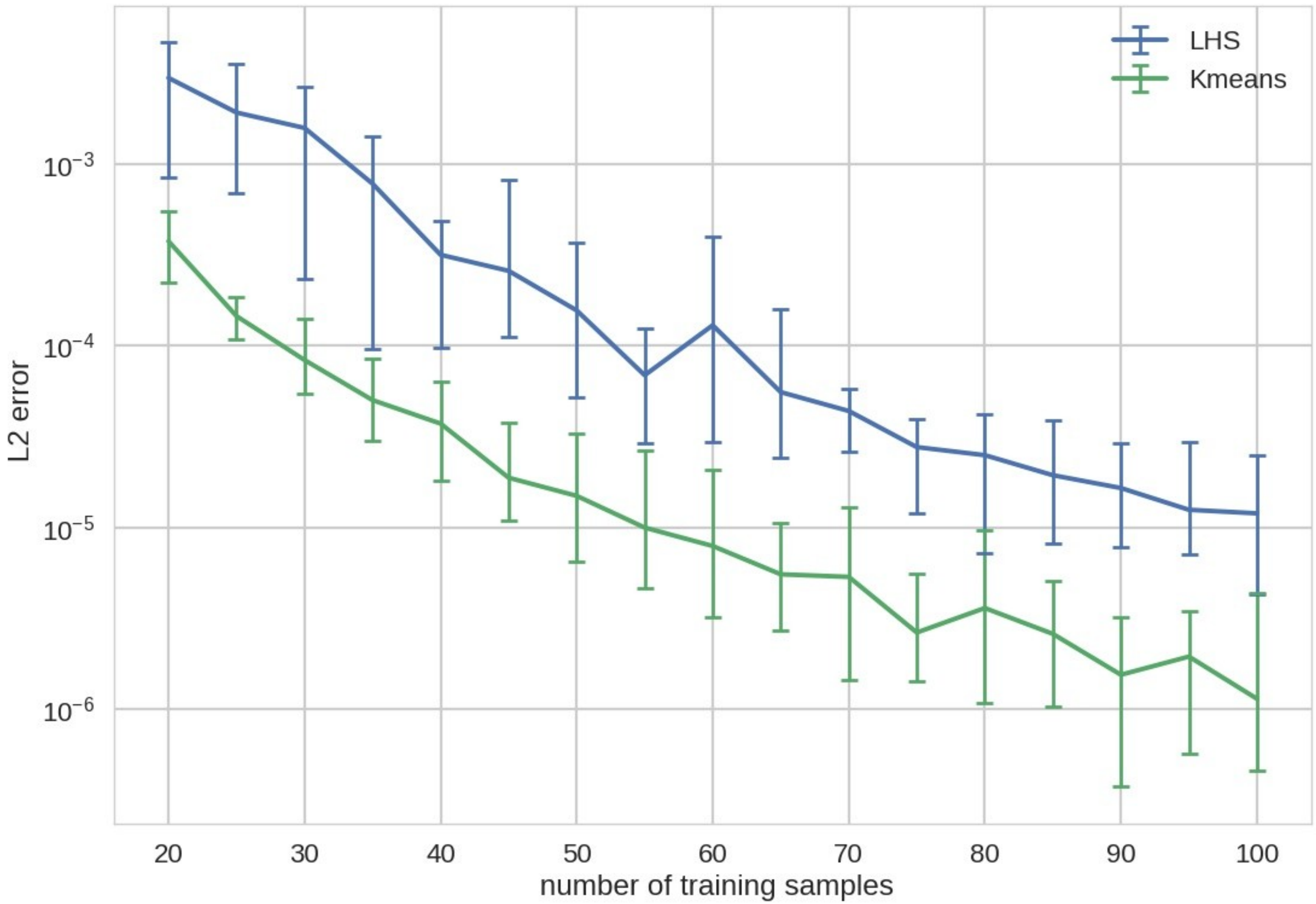
Solver 2 (f_2)



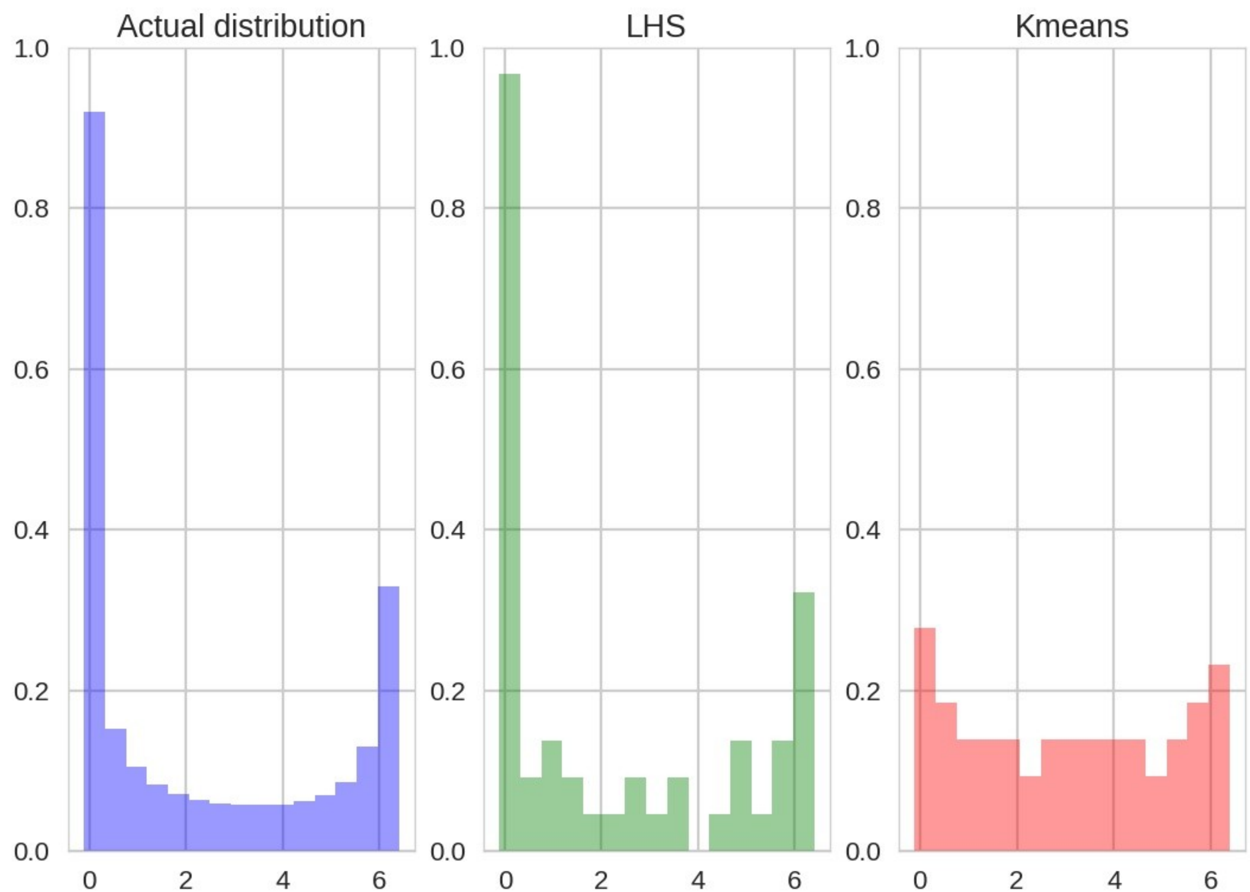
System output ($f_1 o\ f_2$)

system of two solvers

Results

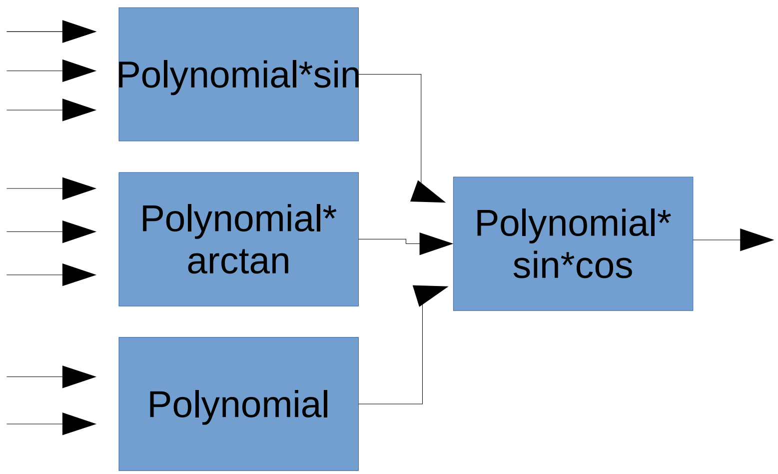


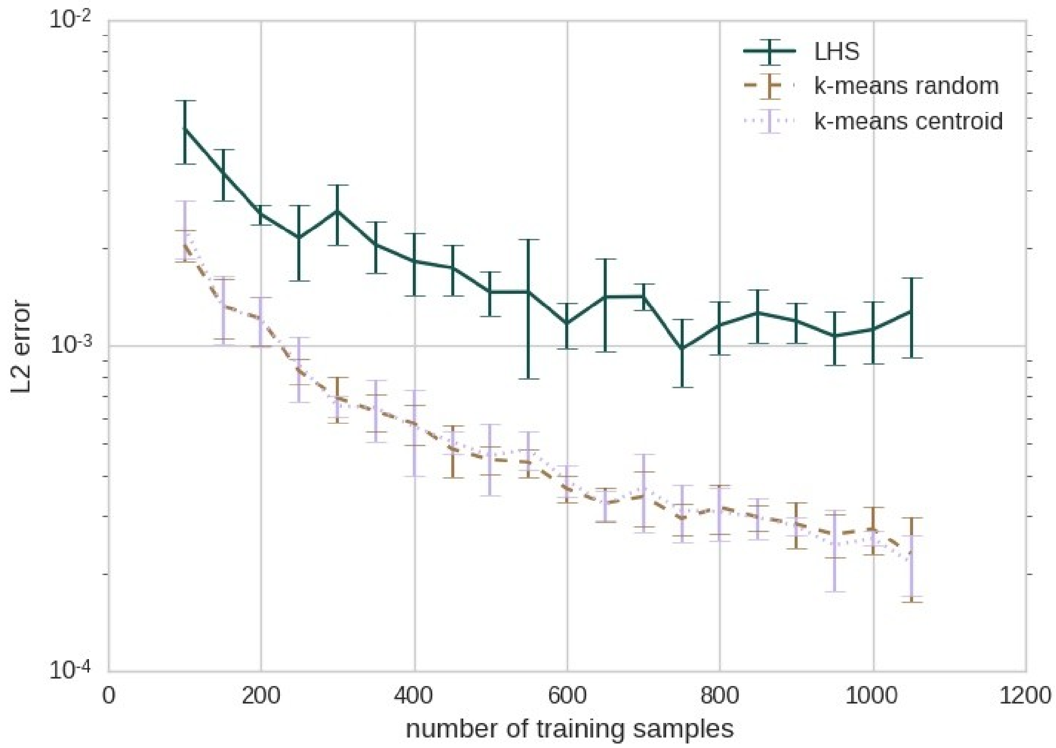
LHS
Kmeans
10^{-3}
10^{-4}
10^{-5}
10^{-6}
L2 error
20
30
40
50
60
70
80
90
100
number of training samples



Centroids do not follow the distribution but present a better coverage

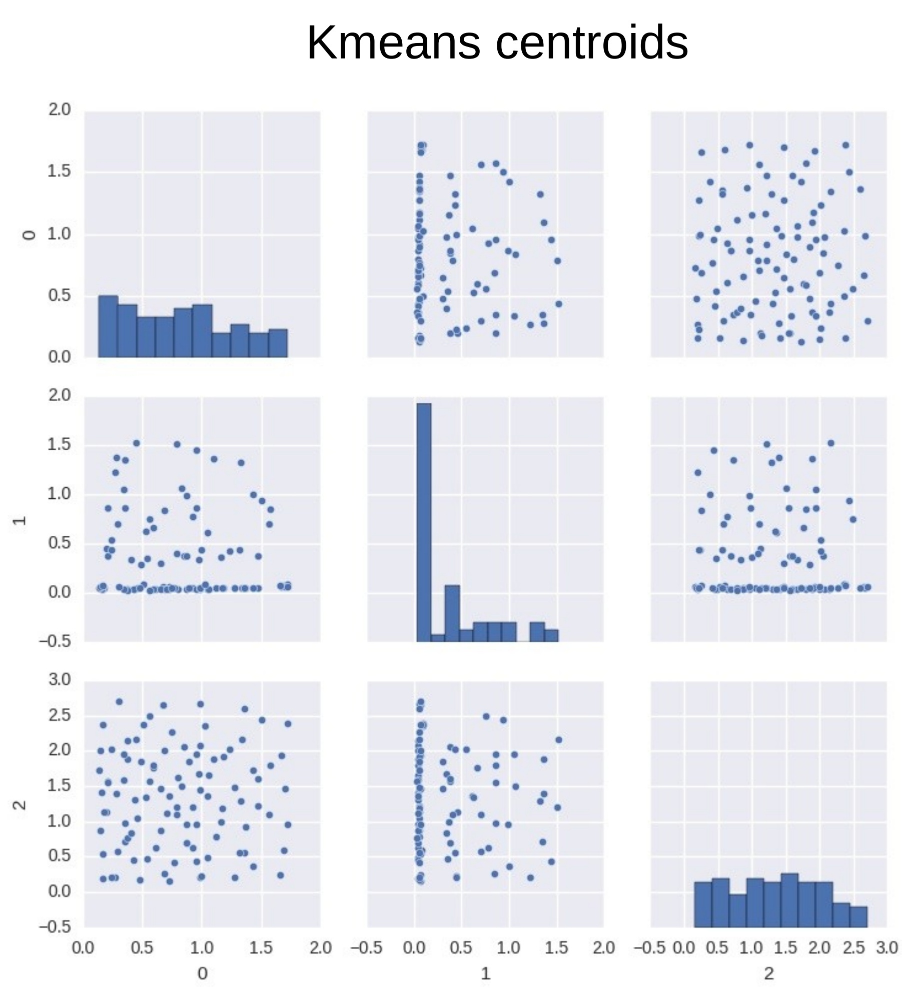
Test case 8D





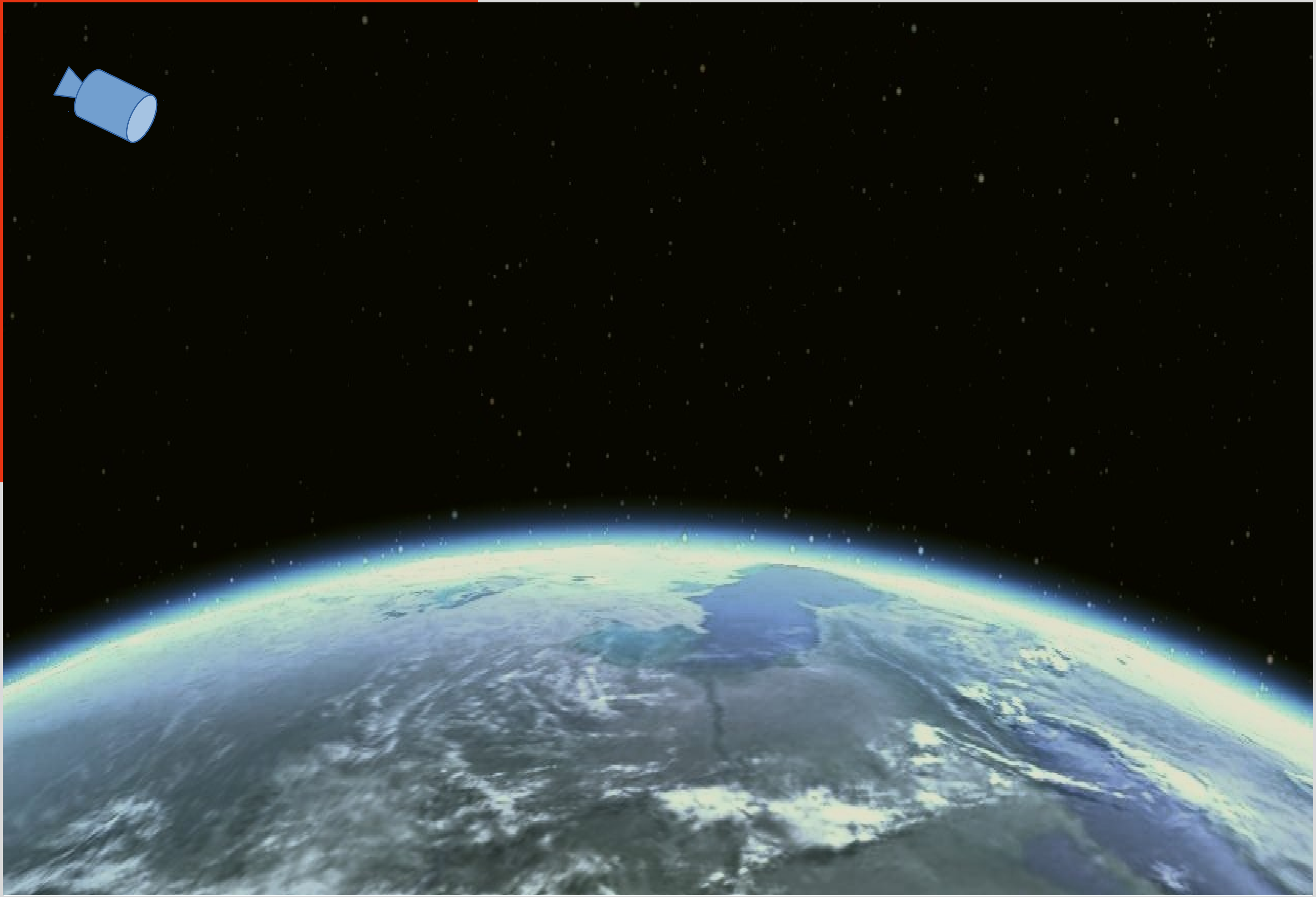
Kmeans centroids or random perform equally well

Scatter matrices of the training sets

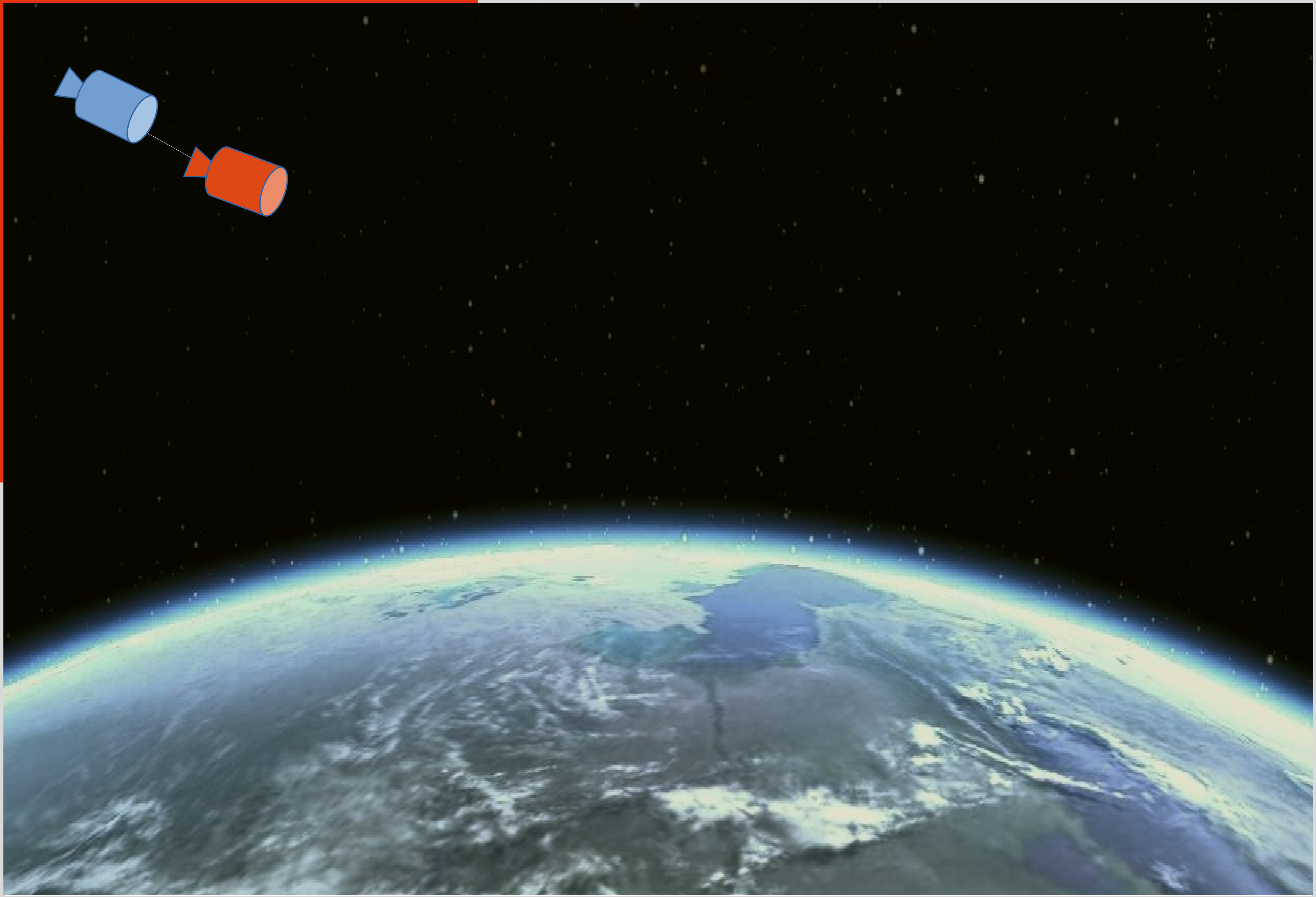


Kmeans avoids the edge of the domain but features alignments

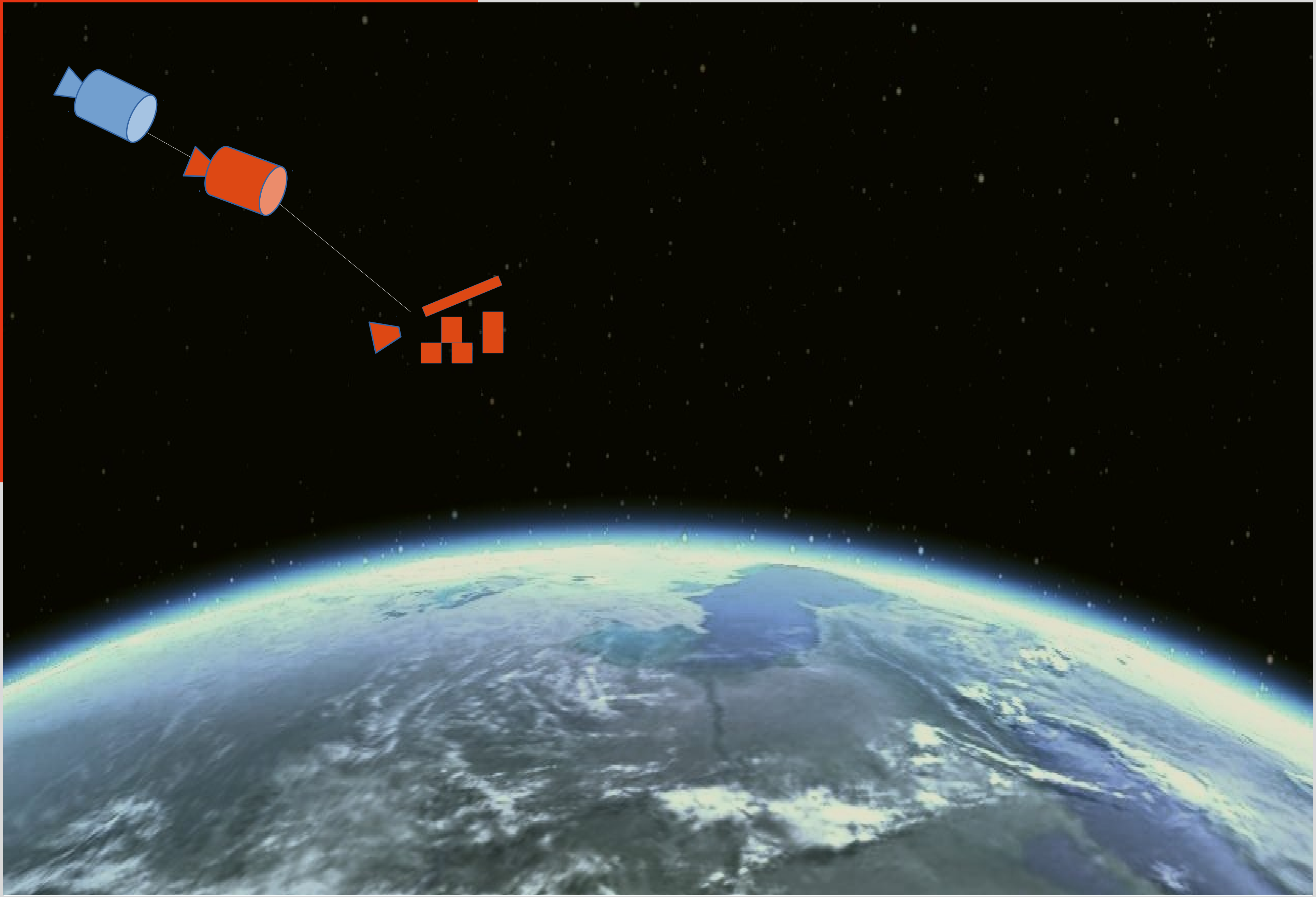
Space Object Breakup Prediction

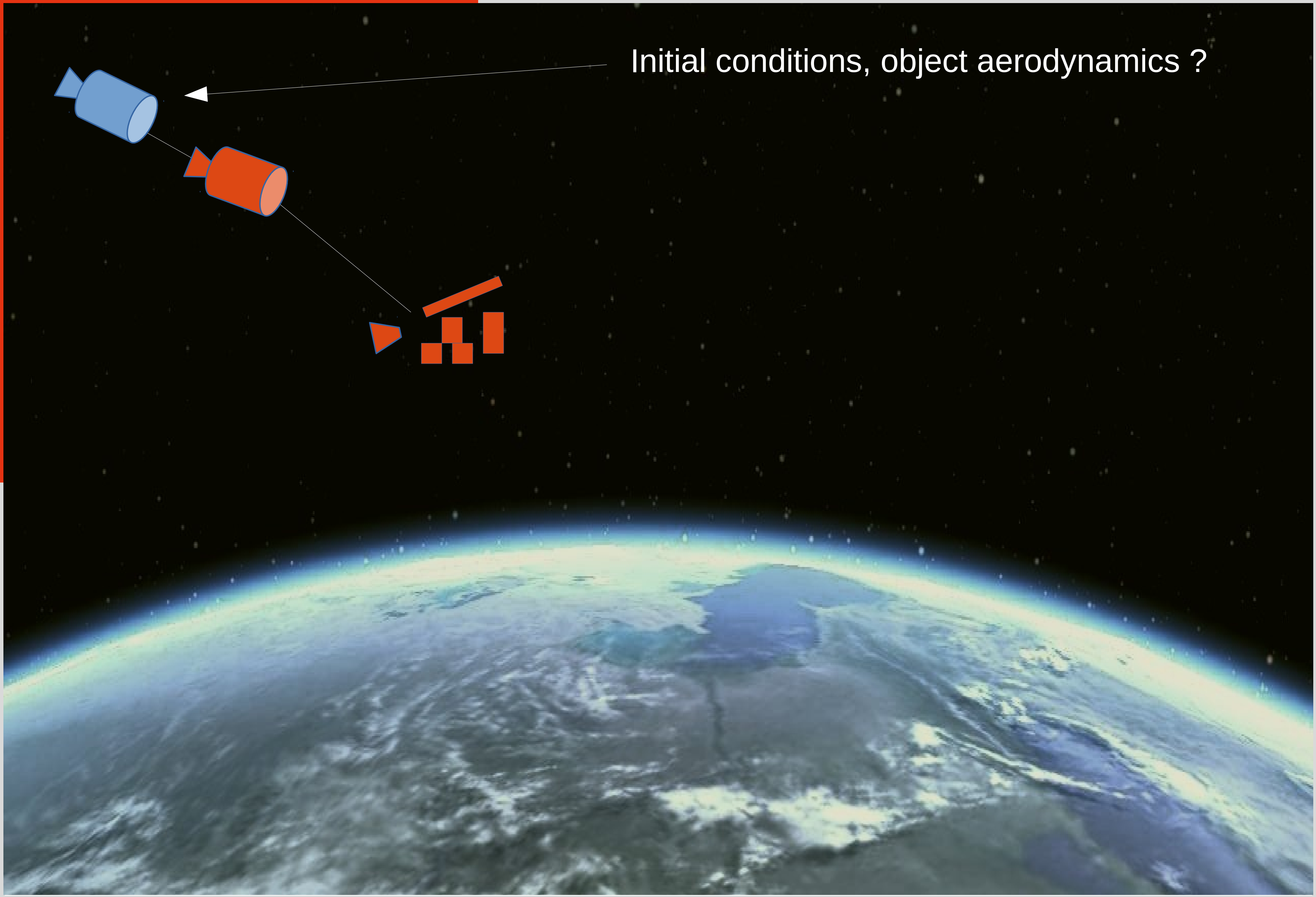


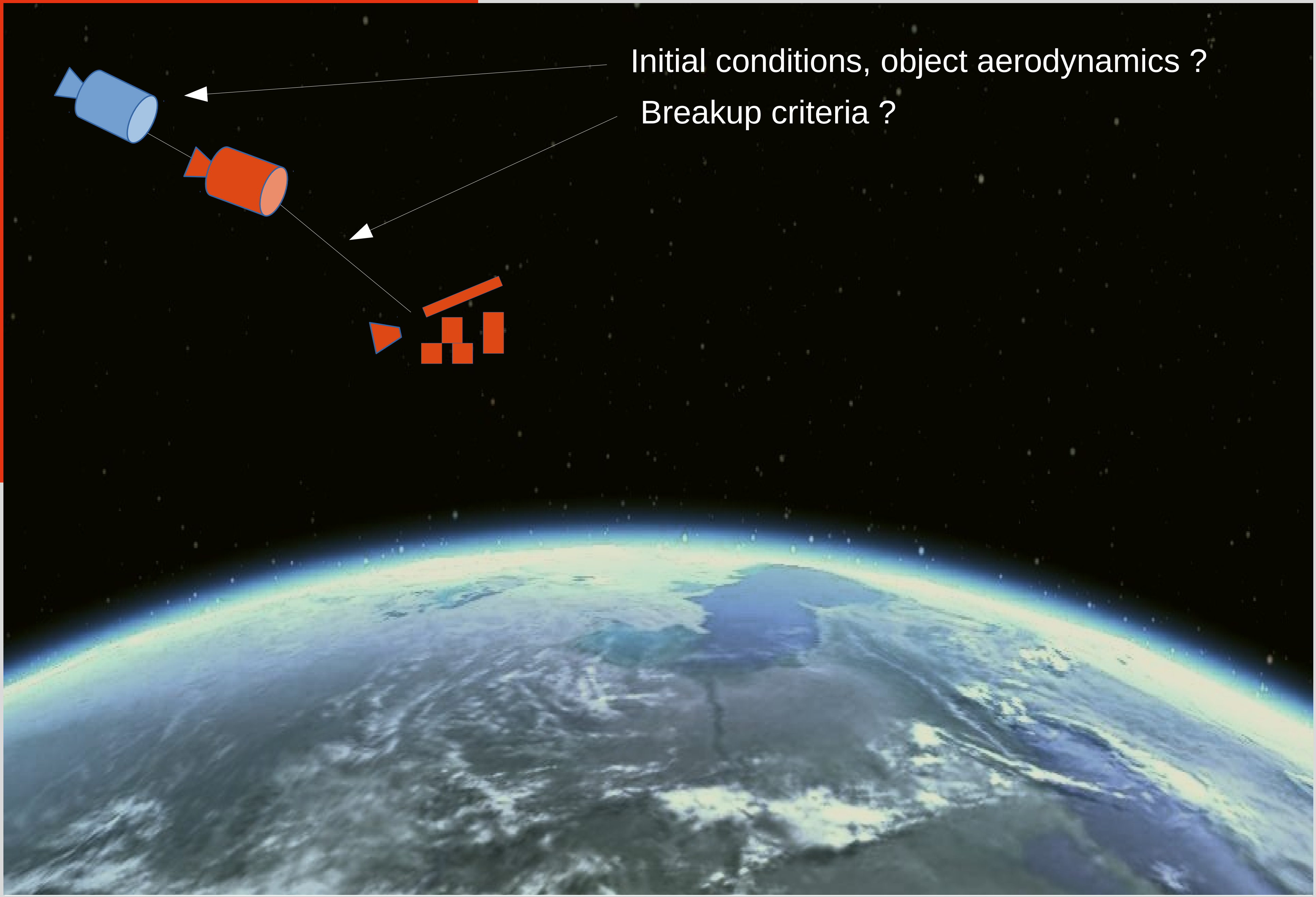
Inria
inventeurs du monde numérique

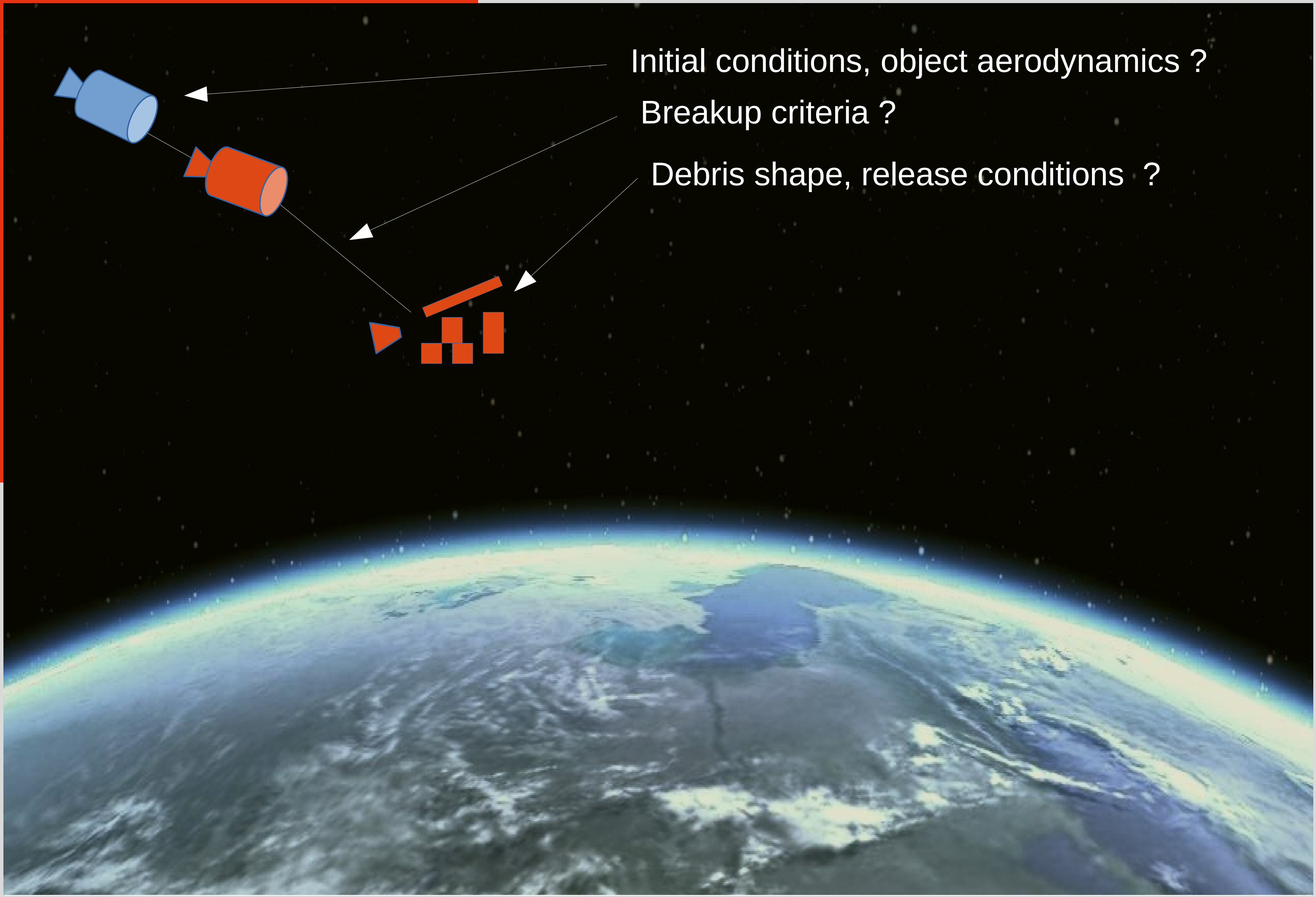


inventeurs du monde numérique




Initial conditions, object aerodynamics ?

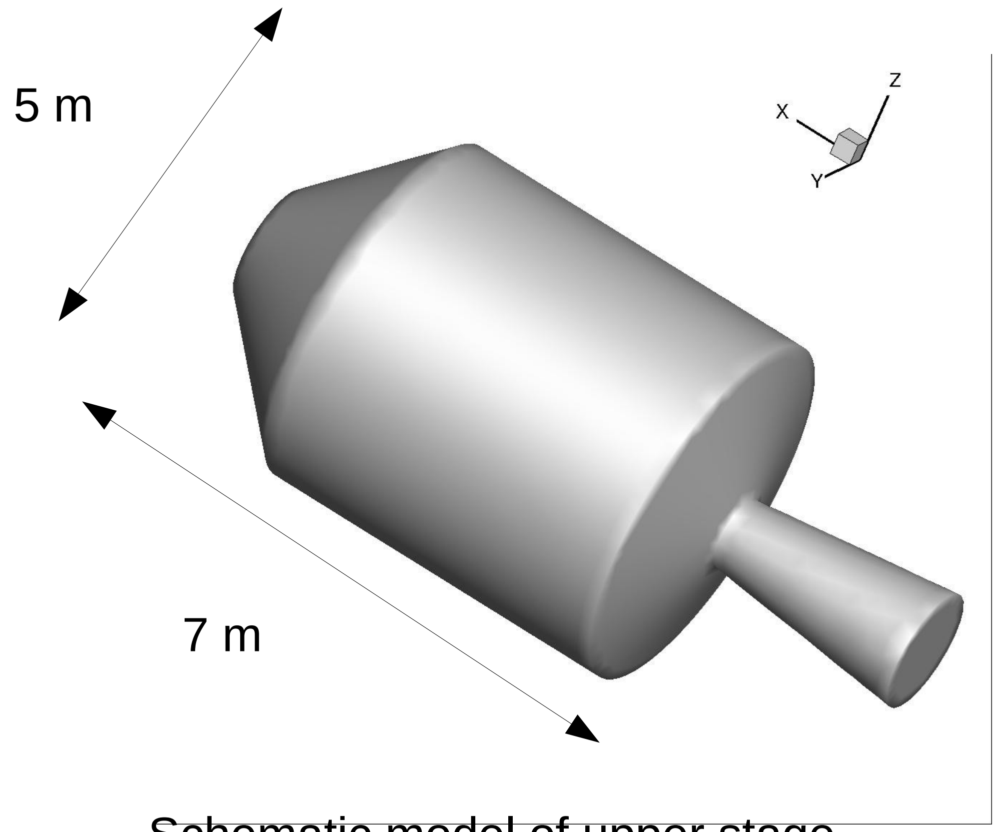

Initial conditions, object aerodynamics ?
Breakup criteria ?


Initial conditions, object aerodynamics ?
Breakup criteria ?
Debris shape, release conditions ?

Uncertainty modeling

22 uncertain parameters :

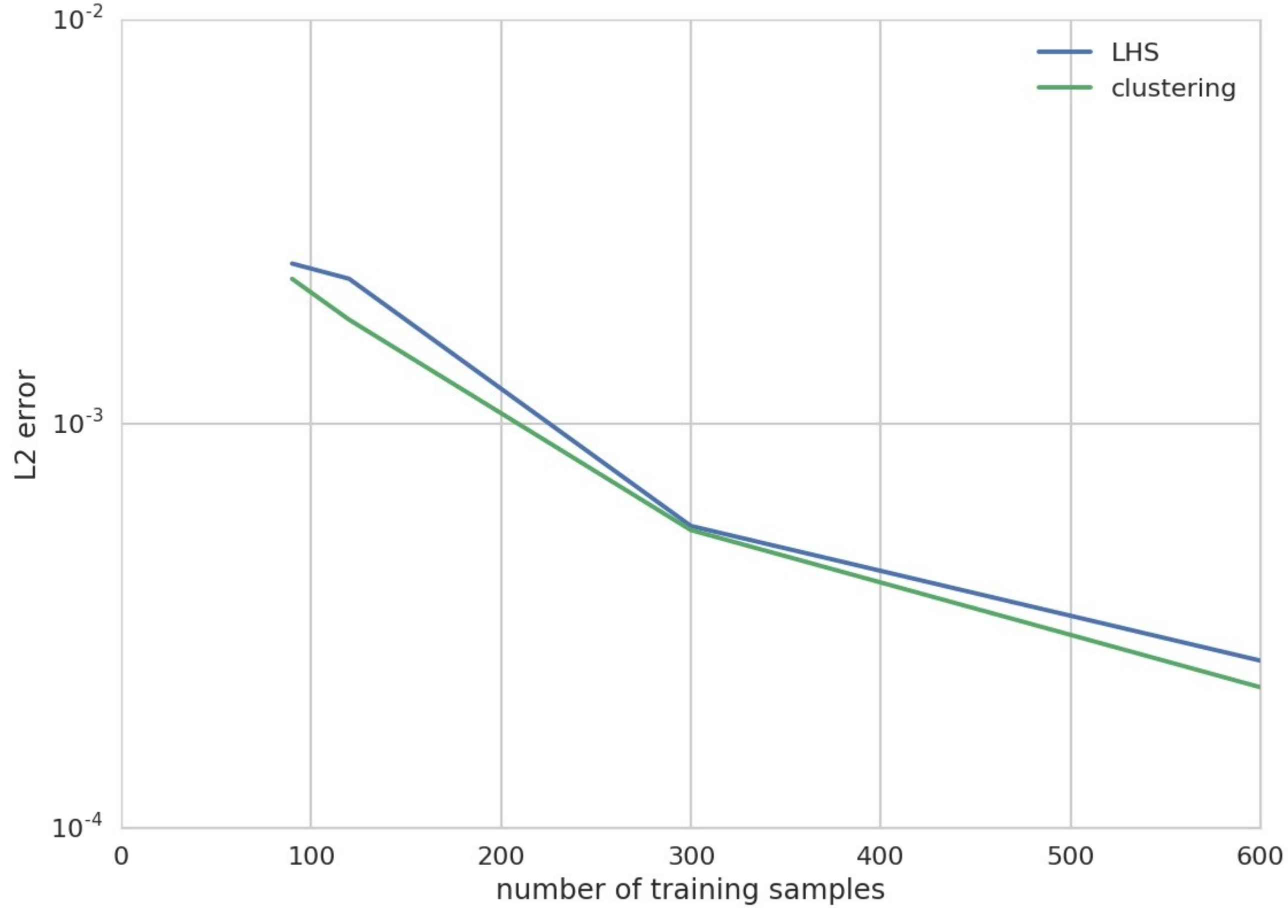
- Deorbiting boost conditions
- Initial orbit conditions
- Atmosphere uncertainties
- Material characteristics
- Breakup model parameters



Schematic model of upper stage

inventeurs du monde numérique

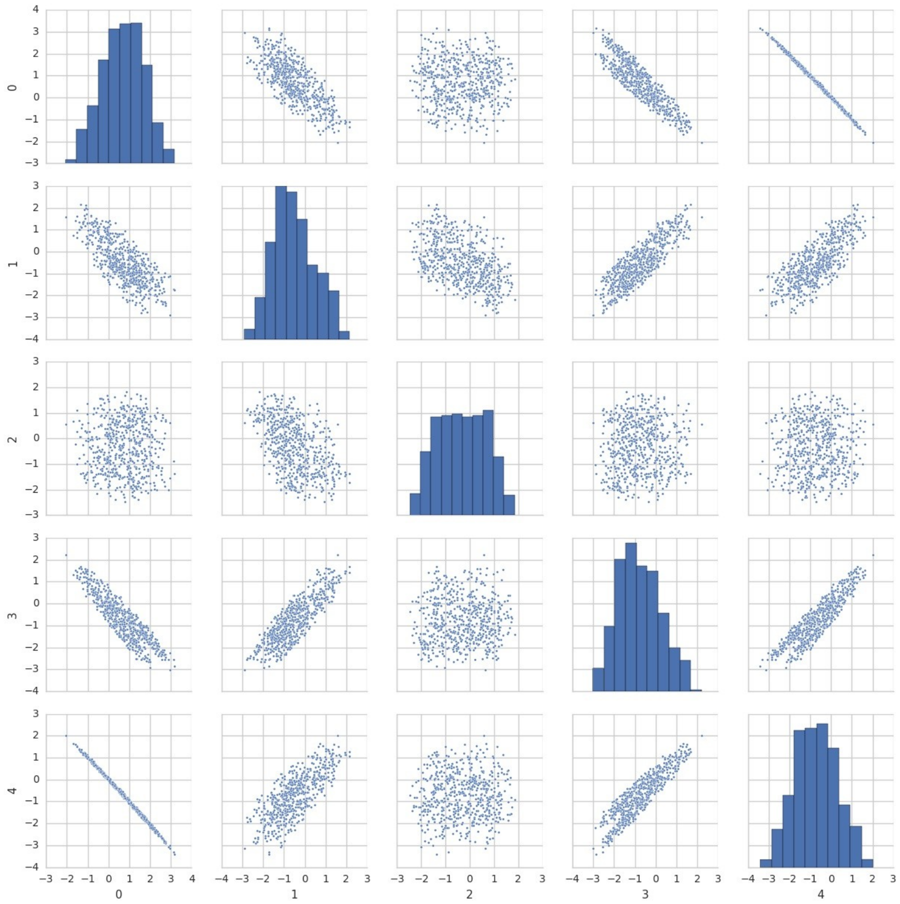
Results



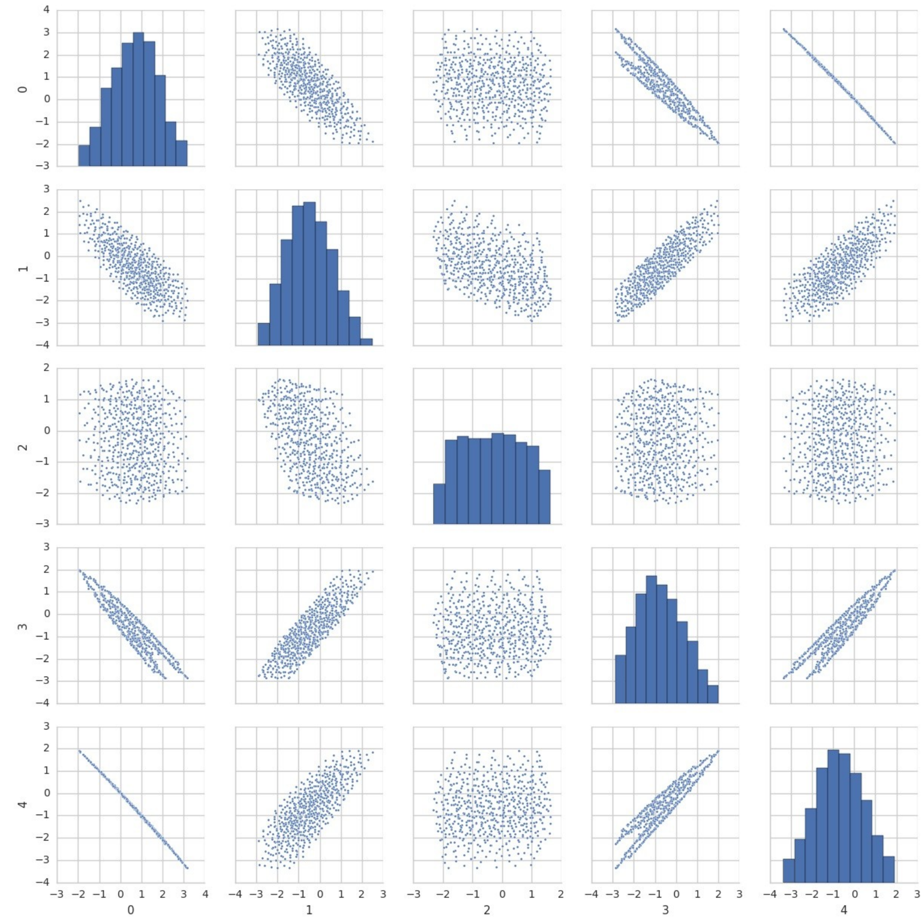
Very similar results :

- The mapping between intermediate variables and output very linear
- The output distributions do not present peaks where LHS typically accumulates points

Training set comparisons



Propagated LHS



Clustering

Conclusions

- Propagation of LHS training sets is not optimal for intermediate solvers
- We presented a clustering based approach to re-sample training sets
- The methods performs better :
    - When the intermediate variable distribution features high probability regions
    - When the intermediate solver is complex to learn
- The clustering approach with centroids does not seem to have good projection properties in lower dimensional spaces.
- This problem can be alleviated by picking a point at random per cluster