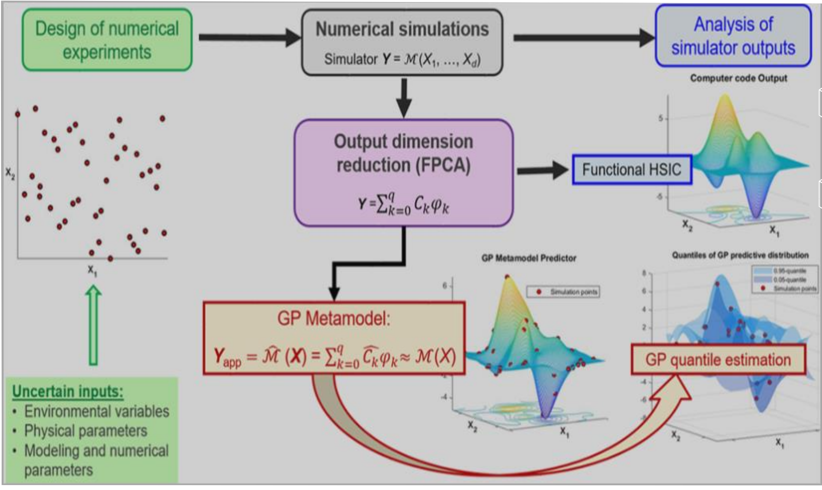

Machine Learning for functional outputs: comparison of dimension reduction methods for temporal and/or spatial variables

DER/SESI/LEMS

The internship will mainly take place at LEMS (Laboratoire d'Études et Modélisations des Systèmes), whose mission is to carry out detailed studies on nuclear systems. The laboratory's skills are multidisciplinary: thermal-hydraulics, severe accident physics, energetics, thermomechanics, statistics/uncertainty treatment.

The aim of the internship will be to study and develop machine learning approaches adapted to temporal and/or spatial outputs, based on dimension reduction techniques, and with the aim of enabling the propagation of input uncertainties in the numerical simulators used in the laboratory.

Nuclear accident analysis for risk assessment relies heavily on computer codes to simulate and predict physical phenomena, in order to estimate safety margins, for example. However, these codes are based on a large number of uncertain input parameters, leading to outputs that are also fraught with uncertainty. **Assessing the impact of uncertainties associated with input parameters on numerical simulator results is therefore essential for safety analysis.**

However, it is often difficult to propagate uncertainties using Monte Carlo approaches on the simulators used, due to the computation time required for each simulation. In practice, only a small number of simulations (several hundred to several thousand) are feasible and available. **A classic approach is to train a machine learning model (supervised statistical learning) on the available simulations.** This mathematical model is then used to carry out more demanding statistical studies (fine propagation of uncertainties, detailed sensitivity analysis, etc.).

In the context of this internship, we are particularly interested in **simulators providing functional variables as output**: temporal and/or spatial variables, enabling us to describe the phenomenon over time and/or space. **The challenge here is to adapt machine learning models to this type of variable.** A tried and tested strategy is to perform an initial output dimension reduction step before training machine learning models on each of the decomposition coefficients.

The aim of the internship will be to implement and compare different dimension reduction techniques, of varying complexity, such as functional principal component analysis, wavelet decomposition and auto-encoder compression.

The aim will also be to quantify and take into account the loss of information associated with this dimension reduction, and to assess its impact on the prediction of the final machine learning model.Translated with DeepL.com (free version)

- Desired schooling :

Master degree in applied mathematics, preferably in statistics or numerical analysis.

- Duration :

6 months

- Method/software(s):

Python/R

- Key words :

Metamodelization, machine-learning, dimension reduction

- Thesis opportunity :

No

- Contact :

BILLO Georis
georis.billo@cea.fr