Internship at Inria and EDF

Visual Sensitivity Analysis for Ensembles of Curves

Internship advisors:

Jean-Daniel Fekete, Inria, jean-daniel.fekete@inria.fr http://www.aviz.fr

Alejandro Ribes, EDF

Goals

EDF needs to forecast the behavior of complex hydraulic systems over time to take into account the evolution of the climate into its production and management of energy (e.g., managing dams and cooling nuclear plants). It uses a probabilistic framework with a parametric model taking p input parameters and generating n values for n time steps. EDF needs to study the behavior of the model around a target set of input parameters. For that, the engineers generate multiple simulations by perturbing the parameters around the target. This produces an ensemble of simulations (thousands to millions, see Figure 1) populated with members (i.e., one simulation with n time steps) to analyze the sensitivity of the simulation around the parameters.

The question we want to address is related to grouping the results into clusters of values varying around a median curve, corresponding to “modes” for the forecast.

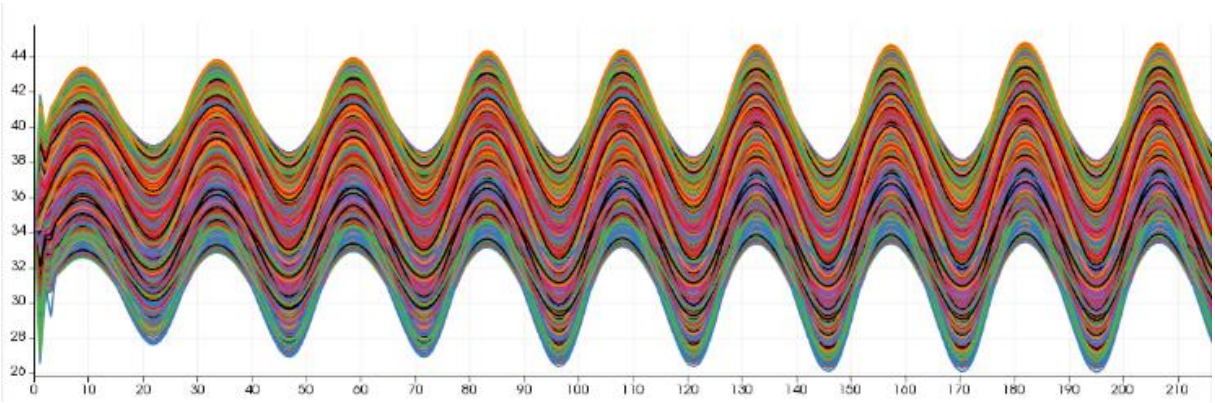

Figure 1: Raw visualization of curves coming from a multi-run hydraulics simulation: 1500 curves of water height evolving over time.

Internship

To study these modes, the engineers gather M simulations and project them using a multidimensional projection (PCA for start). They then cluster the projections from the projections (Figure 2-b) to get a group of curves that will be described through a median curve and a confidence interval. They can visualize each cluster using this median curve and CI as an envelope.

While this process is currently applied, it suffers from performance problems when the number of curves and the number of time-points increase. We want to allow interactive exploration of the parameter space from the envelopes: see which parameter set values produced the different clusters.

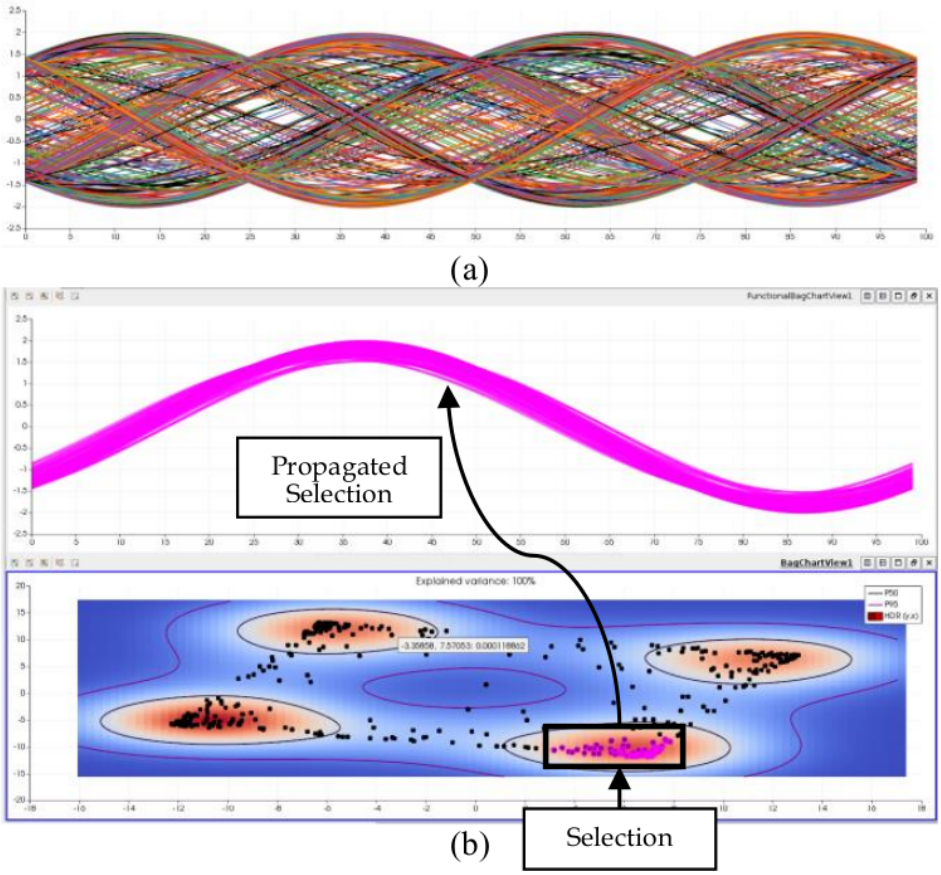

Figure 1: Analysis of curves from their projections

Standard PCA is too slow for data exploration at that scale. The internship will implement a progressive system using iterative computations. New algorithms have appeared recently to compute PCA iteratively quickly (see https://bryan-he.github.io/dawn-site/2017/07/14/accelerated-pca/). It should be coupled with visualization and interactions.

The research goal is to investigate new progressive methods to explore large amounts of data in interactive time.

Practical details

The internship will be located at Inria Saclay and EDF Saclay (10 min apart). Sometimes at EDF Chatou where the simulations are computed and studied.

Programming language is C++, eventually integrated with KitWare ParaView, but not by the intern.

We want to achieve an interactive applications using OpenGL or WebGL, designed for analysts.

There is possibility to continue with a PhD on that topic since visualization, clustering, and exploration of ensemble simulations are very important and limited with existing systems. Progressive algorithms and interface should allow addressing much larger problems.