Gradient-based active learning for Global Sensitivity Analysis

Guerlain Lambert[†,1,2], Céline Helbert[§,1], Claire Lauvernet[§,2]

[†] PhD student [§] PhD supervisor

PhD expected duration: Oct. 2023 – Sep. 2026

[1] Institut Camille Jordan, CNRS UMR 5208, École Centrale de Lyon, Écully, France
{guerlain.lambert,celine.helbert}@ec-lyon.fr

[2] INRAE, RiverLy, 69625 Villeurbanne, France
claire.lauvernet@inrae.fr

Abstract

Global Sensitivity Analysis (GSA) is widely used to understand and improve physical models in industry and environmental sciences by quantifying how uncertain inputs drive the variability of a quantity of interest. In the classical setting, one considers a computational model $Y = f(X)$ where $X = (X_1, \dots, X_d)$ are the inputs, and aims at ranking the components of X according to their influence on the output variability. Variance-based sensitivity measures, in particular Sobol' indices, are popular because they provide an interpretable variance attribution; for instance, the first-order Sobol' index of X_i can be written as $S_i = \frac{\mathrm{Var}(\mathbb{E}[Y|X_i])}{\mathrm{Var}(Y)}$, while the total-effect index $S_i^{\top} = 1 - S_{\sim i}$ accounts for all effects involving X_i, including interaction terms. Such indices are standard tools for GSA [2]. Despite their interpretability, accurate Sobol' index estimation often requires numerous calls to the expensive computer code f, which is prohibitive when a single run may take several dozen hours. A standard approach is to replace f by a surrogate model $\hat{f}$ (e.g., a Gaussian-process, GP) and then to compute Sobol' indices from $\hat{f}$ using Monte Carlo.

Nevertheless, constructing a reliable surrogate still requires carefully chosen observations of f. This motivates adaptive designs of experiments (DoE) that allocate expensive runs where they most improve the learning objective. Recent work [1, 3] suggests using derivative-based global sensitivity measures (DGSM) to drive such sequential designs. DGSM rely on gradient information and indicate which directions in the input space are locally influential. Under a GP prior with standard kernels (RBF, Matérn), gradients of the GP posterior mean and covariance can be computed analytically, enabling acquisition functions tailored to sensitivity learning. Figure 1 shows, on a 2D toy function, the spatial distribution of active learning points selected by Sobol-based random sampling (left) and by two DGSM-driven acquisition functions, highlighting how gradient-based criteria exploit local sensitivity information.

In this talk, we present an active learning strategy adapted to DGSM in the context of GP metamodeling [3], focusing on acquisition functions designed to reduce uncertainty on sensitivity-related quantities. Figure 2 reports the RMSE of DGSM estimates versus active learning steps on scalar-valued toy functions, demonstrating the effectiveness of the proposed approach.

We then discuss how to incorporate input dependencies into these acquisition criteria, an important requirement in realistic applications where uncertain inputs may not be independent. Finally, we examine how to extend the framework to complex functional inputs such as time series, where the effective dimension is high and adaptive designs must exploit temporal structure.

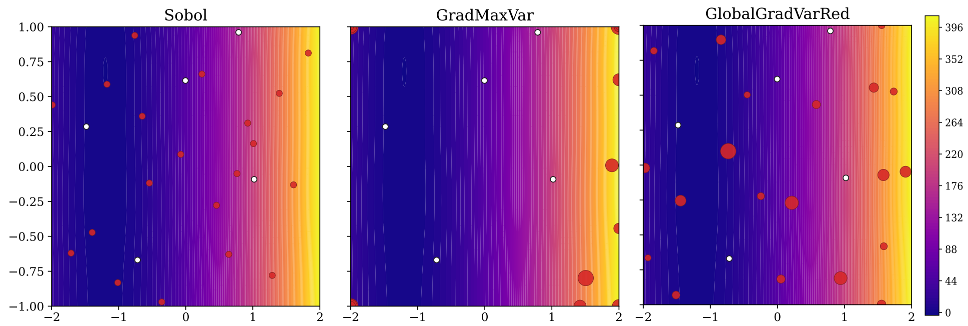

Figure 1: Spatial distribution of active learning points (red) on a two-dimensional toy function for Sobol' random sampling (left) and for DGSM-driven acquisition functions: `LocalGradVarRed` (middle) and `GlobalGradVarRed` (right). White points are initial DoE.

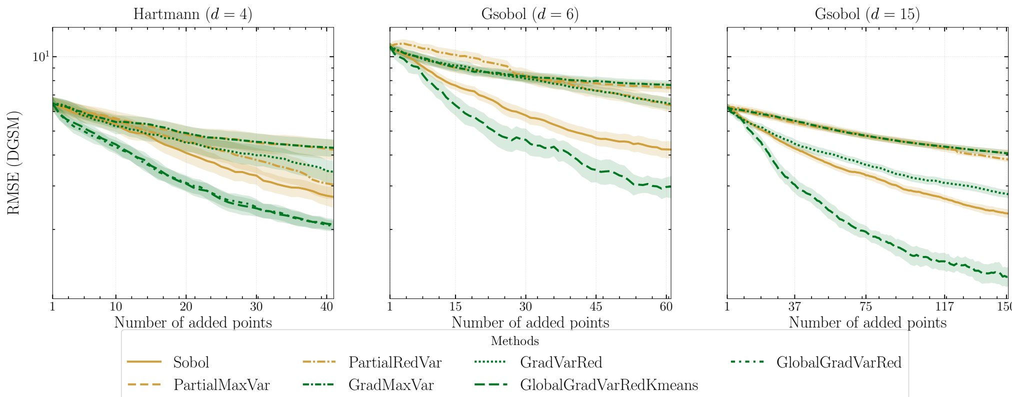

Figure 2: RMSE of DGSM and Sobol total indices estimates versus active learning steps on three toy functions.

The developed methodologies are illustrated on toy cases and are applied to a 3D spatio-temporal model of groundwater pollutant management.

Short biography (PhD student)

I am a third-year PhD student at Centrale Lyon and INRAE Lyon, supervised by Céline Helbert and Claire Lauvernet. My thesis develops active learning methods for complex input data, targeting applications in sophisticated environmental models. The work is part of Water4All's AQUIGROW project, which seeks to boost groundwater service resilience amid rising drought risk. The thesis is also supported by the CIROQUO consortium.

References

[1] S. Belakaria, B. Letham, J. Doppa, B. Engelhardt, S. Ermon, and E. Bakshy. Active learning for derivative-based global sensitivity analysis with gaussian processes. In *Advances in Neural Information Processing Systems*, 2024.

[2] B. Iooss and P. Lemaître. *A Review on Global Sensitivity Analysis Methods*, pages 101–122. Springer US, Boston, MA, 2015.

[3] G. Lambert, C. Helbert, and C. Lauvernet. Gradient-based active learning with gaussian processes for global sensitivity analysis. 2025.