

Uncertainty Quantification for Scientific Machine Learning Models

Supervisor Dr. Dimitrios Loukrezis (d.loukrezis@cwi.nl)

Start Upon agreement

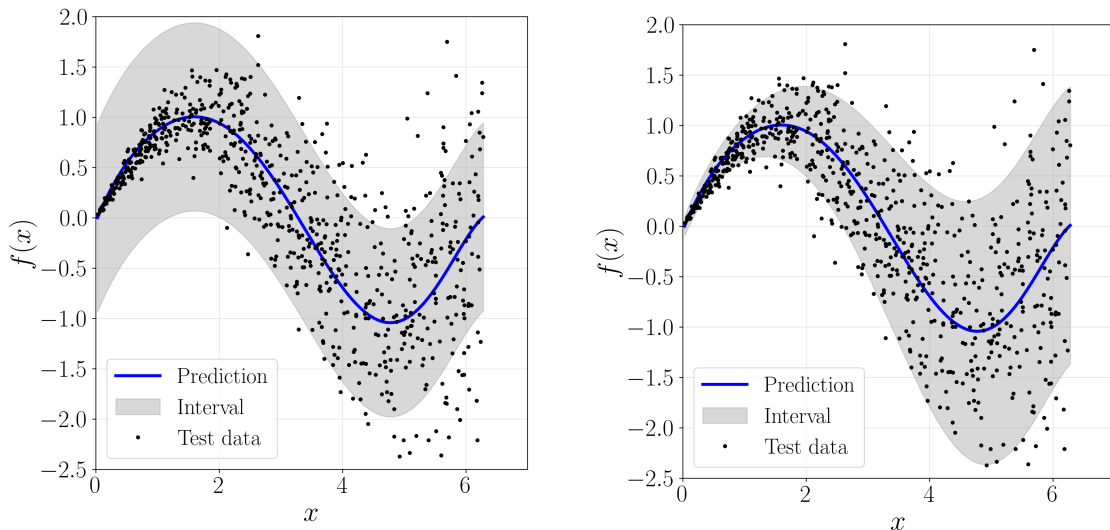
Keywords conformal prediction; scientific machine learning; surrogate modeling; uncertainty quantification

Problem setting

Numerical simulations allow scientists and engineers to investigate *in silico* phenomena that would otherwise be too difficult or expensive to observe and study. However, high-fidelity simulations can be time- and resource-demanding, requiring significant computing resources and long simulation times. This computational burden becomes especially problematic considering repeated analysis tasks, such as optimization or *uncertainty quantification (UQ)*. Practitioners often resort to *surrogate models* (also called *metamodels* or *emulators*) that approximate the input-output map of a costly simulator [1]. One common approach is to construct surrogate models from available datasets using machine learning (ML). More recently, *scientific ML (SciML)* surrogates that combine ML with physics-based modeling principles are getting traction [2]. Nonetheless, most data-driven surrogate models based on either pure ML or SciML produce point estimates that lack information regarding the uncertainty in their predictions. For surrogate-based estimates that may affect safety, cost, or scientific conclusions, information about *predictive uncertainty* is often as important as the prediction itself. This challenge is particularly relevant for data-scarce applications in physics and engineering.

Goal and tasks

This thesis will investigate *conformal prediction (CP)* methods [3] for providing reliable *predictive UQ* metrics to accompany the predictions of SciML surrogate models, see Figure 1. Methodology development will combine domain-aware algorithmic design with careful empirical evaluation, to produce predictive UQ tools that are both statistically principled and usable in data-scarce applications in physics and engineering.



(a) Constant conformal prediction intervals.

(b) Adaptive conformal prediction intervals.

Figure 1: 1D heteroscedastic regression problem. The conformal prediction intervals provide 90% coverage.

Exemplary objectives to be pursued include (but are not limited to):

- Implementation and comparison of different CP variants for representative surrogate models.
- Design of CP methods that incorporate physical knowledge or model structure [4].
- Uncertainty-aware surrogate modeling for dynamical physical systems [5].

Carefully designed numerical experiments will be conducted to determine the capabilities of the offered predictive UQ methods. Key evaluation criteria will be empirical coverage, predictive interval efficiency, robustness under data scarcity, computational cost, and ease of integration with existing SciML workflows. Studies concerning sensitivity to training data availability and comparisons to baseline UQ approaches will also be conducted.

Requirements: Familiarity with concepts and methods from Scientific Computing and Machine Learning. Interest in Uncertainty Quantification. Programming skills in Python.

Doing your thesis at the Scientific Computing group of CWI

The Scientific Computing group of CWI investigates and develops mathematical models to simulate and predict real-world phenomena with inherent uncertainties. Our work is targeted in particular at applications in science and engineering, for example related to climate and energy modeling. As an M.Sc. student in our group, you will benefit from our members' expertise in numerical simulations, reduced order modeling, uncertainty quantification, and scientific machine learning. You will also be able to attend talks and seminars organized by our group, for example, the seminar series on Machine Learning and Uncertainty Quantification in Scientific Computing. You can expect an organized and targeted supervision plan, while enjoying significant autonomy in your research project.

Remark: There exists the possibility for co-supervision with Dr. Stéphane Lanteri, head of the ATLANTIS team at Inria Centre at Université Côte d'Azur. In that case, electromagnetic field applications will be considered. A short-term stay at Inria will also be organized.

References

- [1] Reza Alizadeh, Janet K Allen, and Farrokh Mistree. Managing computational complexity using surrogate models: a critical review. *Research in Engineering Design*, 31(3):275–298, 2020.
- [2] Chuizheng Meng, Sam Griesemer, Defu Cao, Sungyong Seo, and Yan Liu. When physics meets machine learning: A survey of physics-informed machine learning. *Machine Learning for Computational Science and Engineering*, 1(1):20, 2025.
- [3] Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.
- [4] Vignesh Gopakumar, Ander Gray, Joel Oskarsson, Lorenzo Zanisi, Daniel Giles, Matt J Kusner, Stanislas Pamela, and Marc Peter Deisenroth. Uncertainty quantification of surrogate models using conformal prediction. *Machine Learning: Science and Technology*, 7(1):015025, 2026.
- [5] Margaux Zaffran, Olivier Féron, Yannig Goude, Julie Josse, and Aymeric Dieuleveut. Adaptive conformal predictions for time series. In *International Conference on Machine Learning*, pages 25834–25866. PMLR, 2022.