

## STOCHASTIC MACHINE LEARNING: APPLICATIONS FOR MATERIALS DESIGN

Dinis O. Abranches

CICECO – Aveiro Institute of Materials, Department of Chemistry, University of Aveiro, Portugal

### ABSTRACT

Gaussian processes (GPs) are a class of stochastic, non-parametric machine learning models. Owing to their characteristics, GPs can be trained with small and scarce datasets. They also provide uncertainty estimates for their own predictions. As such, GPs can accelerate and optimize the discovery and design of novel materials, particularly in data-scarce scientific domains where experimental trial-and-error approaches are commonplace.

Active learning strategies are a natural extension of GPs, leveraging uncertainty quantification to select the most informative data points for subsequent evaluation. This approach enables substantial reductions in the number of experiments or simulations required to achieve accurate models of complex physicochemical systems.

This work illustrates the usage of GPs in a variety of case studies, from solvent composition optimization to molecular design. Emphasis was placed on how GPs provide not only predictive means but also quantified uncertainties, making them ideal surrogates for guiding efficient experimental or computational campaigns.

Overall, this work highlights how stochastic machine learning frameworks, anchored by Gaussian Processes and active learning, offer a principled, data-efficient route to materials discovery. By combining uncertainty-aware modelling with iterative data acquisition, these approaches can significantly reduce cost and time while improving predictive performance in materials informatics and chemical engineering.

**Keywords:** artificial intelligence, chemical engineering, green solvents, phase diagrams.