Can predictive AI algorithms be used for statistical inference?

Florian Hartig*[1]

[1]Theoretical Ecology, University of Regensburg – Allemagne

Résumé

The popularity of machine learning (ML), deep learning (DL), and artificial intelligence (AI) has grown rapidly in recent years. It is often assumed that the advances offered by ML and DL are mainly useful for predictive tasks. However, ML and DL algorithms can, at least in principle, also be used for statistical inference, including tasks such as effect size estimation or partitioning of explained variation. In my talk, I will discuss the promise and pitfalls of using AI for statistical inference in ecological research, using examples from different fields such as species distribution models, community ecology, or the analysis of bipartite networks. I will first review methods from the field of Explainable AI (xAI) that can be used to extract effects and variable importance from fitted models, and show how these map to traditional statistical indicators. As a next step, I will discuss how we can provide statistical guarantees such as p-values and confidence intervals for these xAI metrics. A key problem that arises in this context is understanding what inductive biases are introduced by the various ML algorithms and their hyperparameters, and how these affect the bias in xAI effect estimates. On the other hand, well-tuned ML and DL algorithms perform automatic model and complexity selection, and thus may offer better alternatives for dealing with structural uncertainty than statistical model selection tools. I will conclude that there is reason to be cautiously hopeful that ML algorithms may merge with other methods of statistical inference, opening up an exciting new option for ecological analyses.

Mots-Clés: Statistics, Machine Learning, AI, Deep Learning, xAI, Inference

*Intervenant