



Stage de Master 2 (6 mois)

Post-traitement statistique des prévisions probabilistes et application en météorologie

English version below

Type de poste

Dans le cadre du projet ANR T-REX, "*Nouveaux challenges pour la prédiction des extrêmes et sa validation*" (Projet de Recherche Collaborative de janvier 2021 à décembre 2024) une bourse de stage M2 est ouverte au recrutement. Le financement prévu est de 6 mois selon indemnisation légale. Le commencement du stage est envisagé début Mars 2021.

A l'issue du stage, une poursuite en thèse sera discutée (financement d'une bourse doctorale de 36 mois par le projet ANR T-REX).

Domaine de recherche

machine learning, prévision probabiliste, prévision météorologique, statistique des extrêmes

Lieu

Rattachement au LmB en collaboration avec le LSCE et le CNRM :

- Laboratoire de mathématiques de Besançon (LmB), Besançon, France
- Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Gif-sur-Yvette, France
- Centre National de la Recherche Météorologique (CNRM), Toulouse, France

Sujet de recherche

Les événements extrêmes peuvent avoir des conséquences socio-économiques majeures. Les crues ou les vagues de chaleur sont des exemples de risques environnementaux associés à des événements extrêmes. La théorie des valeurs extrêmes fournit un cadre théorique pour l'estimation du risque et une estimation mathématique rigoureuse des probabilités d'événements extrêmes.

Nous proposons de développer des outils et modèles statistiques adaptés aux applications complexes, notamment en météorologie, où la prévision des événements extrêmes constitue un enjeu important et un challenge difficile. Les prévisions des modèles numériques sont généralement bonnes pour les événements de tous les jours mais manquent de précision concernant les extrêmes. L'objectif du projet est de proposer et de valider de nouvelles méthodes de post-traitement statistique des modèles numériques permettant d'améliorer ces prédictions, notamment dans les extrêmes.

Les modèles numériques d'ensemble en météorologie fournissent un échantillon de prévisions interprété comme une loi prédictive. Le post-traitement statistique vise à corriger cette loi prédictive de manière à la rendre à la fois calibrée et précise (Gneiting et Raftery, 2007). Des méthodes non-paramétrique de régression quantile par forêts aléatoires peuvent être utilisés pour la calibration (Taillardat et al., 2016). Pour les prévisions des précipitations où les extrêmes sont prononcés, des modèles semi-paramétrique de type Extended Generalized Pareto Distribution ont été proposés pour modéliser la queue de la distribution prédictive (Taillardat et al., 2019). Les méthodes de deep learning



(Lm^B)



font également l'objet de travaux récents dans le cadre du post-traitement statistique (Grönquist et al., 2020).

Le premier objectif est de développer des modèles de type machine learning (forêts aléatoires, deep learning) permettant de modéliser à la fois le cœur et les extrêmes de la loi prédictive, afin d'améliorer les prévisions météorologiques. Le second objectif est la validation de ces méthodes, notamment à travers le développement de scores adaptés aux extrêmes.

Références

- T. Gneiting and A. Raftery (2007). *Strictly proper scoring rules, prediction, and estimation*. J. Am. Stat. Assoc. 102.
- M. Taillardat, O. Mestre, P. Naveau and M. Zamo (2016). *Calibrated Ensemble Forecasts using Quantile Regression Forests and Ensemble Model Output Statistics*. Monthly Weather Review 144.
- M. Taillardat, A.-L. Fougères, P. Naveau and O. Mestre (2019). *Forest-Based and Semiparametric Methods for the postprocessing of Rainfall Ensemble Forecasting*. Weather and Forecasting 34.
- P. Grönquist, C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T. Hoefler (2020). *Deep Learning for Post-Processing Ensemble Weather Forecasts*. *arXiv preprint arXiv:2005.08748*.

Profil et compétences du candidat

Master en Statistique ou Machine learning

Statistique (régression, régression quantile, statistique des extrêmes, apprentissage)

Machine Learning (forêts aléatoires, réseaux de neurones)

Logiciels (R, Python)

Rédaction scientifique et présentation orale (français et anglais)

Motivation pour la recherche et pour une thèse de doctorat

Encadrement / personnes à contacter

Clément Dombry (LmB), clement.dombry@univ-fcomte.fr

Philippe Naveau (LSCE), philippe.naveau@lsce.ipsl.fr

Maxime Taillardat (CNRM, Météo-France), maxime.taillardat@meteo.fr

La candidature comprendra obligatoirement un CV et une lettre de motivation.



Master Internship (6-month) Statistical post-processing of probabilistic forecast and application to weather forecast

Job offer

Within the ANR project T-REX "*new TRends in EXtremes, prediction and validation*" (running from January 2021 to December 2024), a grant for a 6-month Master internship is open. The funding is for a duration of 6 months, starting in March 2021, with salary according to CNRS rules.

Possibility for a PhD grant (funded by ANR project T-REX for 36 months) will be discussed at the end of the master internship.

Research fields

machine learning, probabilistic forecast, weather forecast, statistics of extremes

Location

The main location is the LmB with collaboration with LSCE and CNRM.

- Laboratoire de mathématiques de Besançon (LmB), Besançon, France
- Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Gif-sur-Yvette, France
- Centre National de la Recherche Météorologique (CNRM), Toulouse, France

Research subject

Extreme events can have major socio-economic consequences. Flooding or heat waves are examples of environmental hazards that can be driven by a few rare events. Extreme value theory provides a theoretical framework for risk assessment and mathematically justified estimates for small tail probabilities.

We will develop new methods to get effective tools and statistical models, suitably designed for applications, mostly in the fields of weather forecast where predicting extreme events is at the same time a crucial issue and a challenging task. In meteorology, the current numerical weather prediction (NWP) models are rather skillful for usual weather events, but often fail to provide accurate predictions on extreme weather events. One objective of the project is to derive new statistical post-processing methods that are tailored to the output of existing NWP models in order to improve the forecast of extreme weather.

Numerical Ensemble models for weather forecast provide a sample of forecasts interpreted as a predictive distribution. The statistical postprocessing aims at correcting this predictive distribution so as to make it calibrated and sharp at the same time (Gneiting et al., 2007). Non-parametric methods such as quantile regression forests can be used for calibration (Taillardat et al., 2016). For weather forecast of precipitation, that involve strong extremes, semi-parametric models based on the Extended Generalized Pareto Distribution have been proposed to improve the modeling of the predictive distribution tail (Taillardat et al., 2019). Deep learning methods have been considered



(Lm^B)



recently in the framework of statistical postprocessing (Grönquist et al., 2020) and should be investigated further.

The main goal is to develop machine learning models (random forest, deep learning) providing an accurate description of the bulk and the tail of the predictive distribution, so as to improve weather forecast. The second objective is the validation of those methods, with scoring methods adapted to extremes.

References

- T. Gneiting and A. Raftery (2007). *Strictly proper scoring rules, prediction, and estimation*. J. Am. Stat. Assoc. 102.
- M. Taillardat, O. Mestre, P. Naveau and M. Zamo (2016). *Calibrated Ensemble Forecasts using Quantile Regression Forests and Ensemble Model Output Statistics*. Monthly Weather Review 144.
- M. Taillardat, A.-L. Fougères, P. Naveau and O. Mestre (2019). *Forest-Based and Semiparametric Methods for the postprocessing of Rainfall Ensemble Forecasting*. Weather and Forecasting 34.
- P. Grönquist, C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T. Hoefler (2020). *Deep Learning for Post-Processing Ensemble Weather Forecasts*. *arXiv preprint arXiv:2005.08748*.

Background and skills required

Master in Statistics or Machine learning
Statistics (regression, quantile regression, statistics of extremes, statistical learning)
Machine Learning (Random Forest, neural networks, deep learning)
Statistical softwares (R, Python)
Scientific writing and oral presentation (English and French)
Motivation for research and for a PhD

Contacts

Clément Dombry (LmB), clement.dombry@univ-fcomte.fr
Philippe Naveau (LSCE), philippe.naveau@lsce.ipsl.fr
Maxime Taillardat (CNRM, Météo-France), maxime.taillardat@meteo.fr

The application should include a CV together with a motivation letter.