EMS Annual Meeting Abstracts
Vol. 15, EMS2018-468, 2018

Verification of extreme events for ensemble forecasts using proper scoring rules and extreme value theory

Maxime Taillardat (1), Anne-Laure Fougères (2), and Philippe Naveau (3)

(1) Météo-France, Toulouse, France (maxime.taillardat@meteo.fr), (2) Univ. Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, F-69622 Villeurbanne, France, (3) Laboratoire de Sciences du Climat et de l'Environnement, IPSL-CNRS, Gif-sur-Yvette, France

Verification of ensemble forecasts for extreme events remains a challenging question. The general public as well as the media pay particular attention on extreme events and conclude about the global predictive performance of ensembles, which are often unskillful when they are needed.
Ashing classical verification tools to focus on such events can lead to unexpected behaviors, like destroying the properness of scores. To square up these effects, thresholded and weighted scoring rules have been developed. Most of them use derivates of the Continuous Ranked Probability Score (CRPS). However, some properties of the CRPS for extreme events generate undesirable effects on the quality of verification, and we show on a simple example how it could be tricky sometimes to assess verification, especially for forecast ranking. Using theoretical arguments and simulation examples, we illustrate some pitfalls of conventional verification tools for extremes. We also propose an original way to assess ensemble forecasts using extreme value theory and considering CRPS distributions, instead of CRPS or weighted CRPS averages classically.
This new tool is derived from the behavior of the CRPS distributions for extreme events. We show that this tool, subject to probabilistic calibration, is able to determine the skill of ensemble forecasts for all possible extreme events of interest, and so is a mean to assess exceedance calibration. This new index can also be considered as a symmetric index. Relying on it, we are able to say that the paradigm of maximizing the sharpness subject to calibration can be associated with the paradigm of maximizing the value for extreme events subject to a good overall performance. In this way, and as a future work, it would be convenient to study the specific properties of this CRPS-based tool and its potential paths and pitfalls.
We will finish by illustrating the use of our index on a six-hourly rainfall case study.