



Forecaster's dilemma: Extreme events and forecast evaluation

Sebastian Lerch (1), Thordis Thorarinsdottir (2), Francesco Ravazzolo (3), and Tilmann Gneiting (4)

(1) Institute for Applied Mathematics, Heidelberg University, Heidelberg, Germany, (2) Norwegian Computing Center, Oslo, Norway, (3) Norges Bank, Oslo, Norway, (4) Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

In discussions of the quality of forecasts in the media and public, attention often focuses on the predictive performance in the case of extreme events. Intuitively, accurate predictions on the subset of extreme events seem to suggest better predictive ability. However, it can be demonstrated that restricting conventional forecast verification methods to subsets of observations might have unexpected and undesired effects and may discredit even the most skillful forecasters. Hand-picking extreme events is incompatible with the theoretical assumptions of established forecast verification methods, thus confronting forecasters with what we refer to as the forecaster's dilemma. For probabilistic forecasts, weighted proper scoring rules provide suitable alternatives for forecast evaluation with an emphasis on extreme events. Using theoretical arguments, simulation experiments and a case study on probabilistic forecasts of wind speed over Germany, we illustrate the forecaster's dilemma and the use of weighted proper scoring rules.