

Massive weather (re)forecast data: Challenges and opportunities for statisticians

Tilmann Gneiting

University of Heidelberg, Heidelberg, Germany t.gneiting@uni-heidelberg.de

In state-of-the-art weather forecasting, ensembles of numerical weather prediction (NWP) model are run in real time to provide probabilistic forecasts of future states of the atmosphere. As NWP ensembles tend to be biased and uncalibrated, statistical post-processing techniques, such as Bayesian model averaging and nonhomogeneous regression, yield substantial improvement in the predictive performance. However, statistical post-processing relies on the availability of training cases from the past, which is hindered by frequent changes to the NWP models. An important recent development in this context is the advent of reforecast efforts, where today's operational NWP models are run retrospectively on historic initial conditions, to massively enlarge training sets. Given limited computational resources and the huge data sets involved, the design of reforecast experiments raises critical questions. How many members should a reforecast ensemble comprise? How many years should it go back in time? At what frequency should the reforecasts be run - every day, every third day, every week, or perhaps every two weeks only? I will describe challenges and opportunities for statisticians that arise in this big data setting from both applied and theoretical perspectives.