Statistical Postprocessing of Numerical Weather Predictions

Thorsten Simon
(Universität Innsbruck)

23.01.2019, 16.00 (s.t.)
Institut für Statistik, Ludwigstr. 33, Seminarraum

The talk opens with a short review of numerical weather prediction (NWP), which includes a technical and historical overview. In particular, it highlights that NWP output reveal systematic errors due to uncertainties in the initial conditions, imperfect model physics, and spatial and temporal discretization. Theses systematic errors can be corrected by statistical postprocessing (also called calibration) the output of NWP.

The main part presents a recent study on postprocessing NWP: A method to predict lightning is developed for the region of the European Eastern Alps. Cloud-to-ground (CG) flashes—detected by the ground-based ALDIS network—are counted on the $18 \times 18\,km^2$ grid of the 51-member NWP ensemble of the European Centre of Medium-Range Weather Forecasts (ECMWF). These counts serve as the target quantity in count data regression models for the occurrence of lightning events and flash counts of CG. The probability of lightning occurrence is modelled by a Bernoulli distribution. The flash counts are modelled with a hurdle approach where the Bernoulli distribution is combined with a zero-truncated negative binomial. In the statistical models the parameters of the distributions are described by additive predictors, which are assembled using potentially nonlinear functions of NWP covariates. Measures of location and spread of 100 direct and derived NWP covariates provide a pool of candidates for the nonlinear terms. A combination of stability selection and gradient boosting identifies the nine (three) most influential terms for the parameters of the Bernoulli (zero-truncated negative binomial) distribution, most of which turn out to be associated with either convective available potential energy (CAPE) or convective precipitation. Markov chain Monte Carlo (MCMC) sampling estimates the final model to provide credible inference of effects, scores and predictions. The selection of terms and MCMC sampling are applied for data of the year 2016, and out-of-sample performance is evaluated for 2017. The occurrence model outperforms a reference climatology—based on seven years of data—up to a forecast horizon of 5 days. The flash count model is calibrated and also outperforms climatology for exceedance probabilities, quantiles, and full predictive distributions.

The talk concludes with an outlook on future research in the field of statistical postprocessing of NWP output. A special focus is set on multivariate approaches, i.e., not only the marginal distributions of a target quantity (such as temperature or precipitation) have to be predicted for multiple time steps or locations, but also the covariance between these time steps or locations. For this purpose, we propose to fit a single multivariate Gaussian model where all parameters (means, variances, and correlations) can be expressed by additive models.