Statistical post-processing of ensemble forecasts for various weather quantitie

Final report, project NN 125679

Sándor Baran and Patrícia Szokol

The aim of the project was the development of new statistical post-processing methods for joint calibration of ensemble forecasts for various weather quantities, which models can also take into account the spatial dependence among the observation sites. We also intended to deal with the problem of parallelization of the existing post-processing techniques. Besides this, we planned the extension of similarity- and clustering based semi-local approaches for estimation of model parameters from wind speed to other weather quantities. Most of our goals have been achieved and some new research directions were also introduced. Moreover, we intensified the collaboration with the companion research team in Germany. The results of the project were published in 14 research papers with more than a half appearing in leading journals of the related fields (Scimago ranks: D1: 3; Q1: 5; Q2: 2; Q3: 3; Q4: 1). As of today, these works, together with the developed R package, have already received 86 independent citations. Besides publishing our research results, we also tried to increase the international visibility of our joint Hungarian-German research group by introducing our project in an article of the Research and Innovation section of *ERCIM News*.

1. Development and testing of multivariate post-processing methods

As a joint project for the whole international research group, we performed a detailed simulation-based comparison of multivariate ensemble post-processing methods [1]. We focused on generally applicable two-step approaches where ensemble predictions are first post-processed separately in each margin, and multivariate dependencies are restored via copula functions in a second step. The comparisons were based on simulation studies tailored to mimic challenges occurring in practical applications and allowed to readily interpret the effects of different types of misspecifications in the mean, variance and covariance structure of the ensemble forecasts on the performance of the post-processing methods.

As a continuation of this work, together with Sebastian Lerch and Stephan Hemri from the international research group, we investigated the same problem using global temperature, wind speed and precipitation accumulation ensemble forecasts of the Europena Centre for Medium-Range Weather Forecasts (ECMWF) and corresponding observations for a 12-year period between 2002 and 2014 with forecast horizons of 24,48,...,240 hours. We considered temporal dependencies by treating together all forecasts for a given observation station initialized at a given time point resulting in 10-dimensional forecast trajectories, and compared the predictive performance of 9 state-of-the-art two-step multivariate post-processing approaches [2].

The planned extension of the time series approach of Annette Möller and Jürgen Groß [3] from the German part of the team was finally performed by them to the heteroscedastic situation [4]; however, no results were achieved in the non-Gaussian case. Unfortunately, the planned non-Gaussian and multivariate extension of the Markovian ensemble model output statistics (EMOS) method [5] also appeared to be too difficult to handle.

2. Development of R packages

Together with Sebastian Lerch we extended the ensembleMOS package of R with truncated normal and log-normal EMOS models for calibration of wind speed and generalized extreme value and

censored shifted gamma EMOS methods for calibration of precipitation accumulation ensemble forecasts. The actual version 0.8.2 of the package is available at the CRAN repository [6].
The extension of the ensAR R package at least to the heteroscedastic case is under progress and performed by Annette Möller and Jürgen Groß.

3. Investigation of semi-local parameter estimation methods

We studied the effect of statistical calibration on ECMWF dual-resolution ensemble forecasts of 2-m temperature. This work was running parallel with experiments made at the ECMWF on mixing high- and low resolution ensemble forecasts having a fixed total computation cost, in order to determine the optimal combination having the best predictive performance. Four scenarios of various combinations of three different resolutions between 18 km and 45 km were tested. In the case of the raw ensemble forecasts for each scenario one could find an optimal combination having the best predictive performance. Our studies proved that clustering-based semi-local EMOS post-processing usually does not change the ranking of the different combinations in terms of forecast skill; however, it strongly reduces the differences among the equal cost configurations of single and dual-resolution ensembles [7].

Further, we tested different post-processing techniques for calibrating 9-member ensemble forecast of 2-m temperature for Santiago de Chile, obtained by the Weather Research and Forecasting (WRF) model using different planetary boundary layer and land surface model parametrization. In particular, EMOS and Bayesian model averaging (BMA) techniques were implemented, and since the observations were characterized by large altitude differences, the estimation of model parameters was adapted to the actual conditions at hand. Compared with the raw ensemble, all tested post-processing approaches significantly improved the calibration of probabilistic and accuracy of point forecasts, and the EMOS method using parameter estimation based on expert clustering of stations (according to their altitudes) showed the best forecast skill [8]. Later, we also investigated different parametric and non-parametric post-processing techniques for calibrating WRF ensemble forecasts of surface wind speed for locations around the cities of Valparaíso and Santiago de Chile. Statistical calibration was performed with the help of EMOS and quantile regression forest (QRF) methods both with regional and semi-local approaches to model estimation. Compared with the classical EMOS technique based only on the ensemble forecasts, QRF allows the inclusion of other features as predictors in the post-processing model. Among the competing calibration methods QRF using a semi-local approach and considering some specific weather variables from WRF simulations selected by their importance exhibited the best forecast skill [9].

Research directions beyond the planned activities

4. Development of parametric post-processing models

Together with Stephan Hemri we developed a new BMA approach to calibration of hydrological ensemble forecasts, which models Box-Cox transformed water levels with a mixture of doubly truncated normal distributions. To estimate model parameters we developed an efficient Expectation-Maximization algorithm for optimizing the likelihood function of the training data. The model was tested on multi-model ensemble forecasts of water level at gauge Kaub at river Rhine and its forecast skill was compared both with the raw ensemble and the state-of-the-art doubly truncated normal EMOS approach. The results of this case study showed that compared with the raw ensemble, post-processing always improved the calibration of probabilistic and accuracy of point forecasts. With rolling window training periods the BMA model significantly outperformed the reference EMOS approach. However, the use of more sophisticated analog-based selection of training data drastically decreased the gap in forecast skill leaving only a small advantage of BMA compared with EMOS [10].

We also studied the effect of statistical calibration on the probabilistic skill of ensemble forecasts of two different heat indices (discomfort index (DI) and indoor wet-bulb globe temperature (WBGTid), both calculated from the corresponding forecasts of temperature and dew point temperature. Two different methodological approaches to calibration were compared. In the first case, we started with joint post-processing of the temperature and dew point forecasts and then created calibrated samples of DI and WBGTid using samples from the obtained bivariate predictive distributions. This approach was compared with direct post-processing of the heat index ensemble forecasts. For this purpose, a novel EMOS model based on a generalized extreme value (GEV) distribution was proposed. The predictive performance of both methods was tested on ECMWF operational temperature and dew point ensemble forecasts and the corresponding forecasts of DI and WBGTid. For short lead times, both approaches significantly improved the forecast skill. Among the competing post-processing methods, direct calibration of heat indices exhibited the best predictive performance, very closely followed by the more general approach based on joint calibration of temperature and dew point temperature. Additionally, a machine learning approach was tested and showed comparable performance for the case when one was interested only in forecasting heat index warning level categories. This research was in-line with the the aims of the ECMWF TIGGE/S2S CHALLENGE on the study of different user oriented variables [11] (see https://www.ecmwf.int/sites/default/files/medialibrary/2018-11/TIGGE-S2S-WS_Challenge.pdf)

Furthermore, together with Sebastian Lerch we tested the calibration of total cloud cover (TCC) ensemble forecasts using various machine learning techniques, such as multilayer perceptron (MLP) neural networks, gradient boosting machines (GBM) and random forest (RF) methods. Based on ECMWF global TCC ensemble forecasts for 2002 – 2014, we compared these approaches with the proportional odds logistic regression (POLR) and multiclass logistic regression (MLR) models, as well as the raw TCC ensemble forecasts. We further assessed whether improvements in forecast skill can be obtained by incorporating ensemble forecasts of precipitation as additional predictor. Compared to the raw ensemble, all calibration methods resulted in a significant improvement in forecast skill. RF models provided the smallest increase in predictive performance, while MLP, POLR and GBM approaches performed best. The use of precipitation forecast data leaded to further improvements in forecast skill and except for very short lead times the extended MLP model showed the best overall performance [12].

We also developed a new EMOS model for calibrating wind speed ensemble forecasts, where the predictive distribution is a GEV distribution left truncated at zero (TGEV). The truncation corrects the disadvantage of the GEV distribution based EMOS models of occasionally predicting negative wind speed values, without affecting its favorable properties. The new model was tested on four data sets of wind speed ensemble forecasts provided by three different ensemble prediction systems, covering various geographical domains and time periods. The forecast skill of the TGEV EMOS model was compared with the predictive performance of the truncated normal, log-normal and GEV methods and the raw and climatological forecasts as well. Our case studies verified the advantageous properties of the novel TGEV EMOS approach [13].

Working again with Sebastian Lerch, we investigated the calibration of solar irradiance ensemble predictions. We proposed EMOS post-processing models for different measured types of solar irradiance, using corresponding ensemble predictions as input. To account for the specific discrete-continuous nature of solar irradiance due to the positive probability of observing zero irradiance during night-time, a censored logistic and a censored normal forecast distribution is used, motivated by similar models for post-processing ensemble forecasts of precipitation accumulation. The post-processing models were evaluated in two case studies covering distinct solar irradiance variables (direct, diffuse and global irradiance), NWP models (ICON-EPS of the German Meteorological Service, AROME-EPS of the Hungarian Meteorological Service), temporal resolutions (from 30 min to 6 h) and geographic regions (Hungary and Germany), for lead times of up to 48 and 120 hours, respectively. Periodic models were utilized to better capture seasonal variation in solar irradiance, and different temporal compositions of training datasets for model estimation were investigated. We found that post-processing consistently and significantly improved the forecast performance of the ensemble predictions for lead times up to at least 48 hours and was well able to correct the systematic lack of calibration [14].

Together with Ágnes Baran and in collaboration with forecasters of the Hungarian Meteorological Service (HMS), we investigated statistical calibration of short-range forecasts of wind speed measured at hub height (100m), which weather quantity plays an important role in wind energy production. Using AROME-EPS wind speed forecasts of the HMS and observations at three wind farms in Hungary, we compared the forecast skill of various EMOS and BMA models and developed an artificial neural network (ANN) based distributional regression network (DRN) approach to estimation of the parameters of the predictive distribution. In contrast to the existing methods using ANNs, no other predictors, such as pressure or wind direction, were involved as input features; estimation was based merely on some functionals of the raw ensemble. Compared with the raw AROME-EPS forecasts, the proposed method resulted in a 12% improvement in the predictive performance [15].

Fully documented R and Python codes of our wind speed and solar irradiance models for the AROME-EPS forecasts were provided to the HMS in order to prepare these methods for operational use. At the moment, our truncated normal and log-normal distribution-based EMOS and DRN models of wind speed, and censored logistic and censored normal distribution-based EMOS models of solar irradiance are under pre-operational testing, for an overview see [16] (in Hungarian).

5. Related theoretical topics

We investigated problems connected to optimal designs of experiments, which are closely related to the problem of optimal allocation of the observation stations. First we considered the problem of optimal designs for the prediction of complex Ornstein-Uhlenbeck processes and applied our result in the prediction of the Chandler wobble, the small deviation in the Earth's axis of rotation [17]. Furthermore, we proposed a novel approach identifying relationship and providing the exact relations between estimation and prediction for regression problems with correlated errors [18].

Finally, we proved a dichotomy result for strictly increasing bisymmetric maps, which can be applied in the field of utility functions, hence may play role in producing cost efficient weather forecasts [19].

References

1. Lerch, S., Baran, S., Möller, A., Groß, J., Schefzik, R., Hemri, S., Graeter, M., Simulation-based comparison of multivariate ensemble post-processing methods. *Nonlinear Process. Geophys.* **27** (2020), no. 2, 349–371.
2. Lakatos, M., Lerch, S., Hemri, S., Baran, S., Comparison of multivariate post-processing methods using global ECMWF ensemble forecasts. *Q. J. R. Meteorol. Soc.* **149** (2023), 856–877.
3. Möller, A., Groß, J., Probabilistic temperature forecasting based on an ensemble AR modification. *Q. J. R. Meteorol. Soc.* **142** (2016), 1385–1394.
4. Möller, A., Groß, J., Probabilistic temperature forecasting with a heteroscedastic autoregressive ensemble postprocessing model. *Q. J. R. Meteorol. Soc.* **146** (2020), 211–

224.
5. Möller, A., Thorarinsdottir, T. L., Lenkoski, A. and Gneiting, T., Spatially adaptive, Bayesian estimation for probabilistic temperature forecasts. arXiv:1507.05066.
6. Yuen, R. A., Baran, S., Fraley, C., Gneiting, T., Lerch, S., Scheuerer, M., Thorarinsdottir, T. L., R *package ensembleMOS, Version 0.8.2: Ensemble Model Output Statistics* (2018). Available at: https://cran.r-project.org/package=ensembleMOS
7. Baran, S., Leutbecher, M., Szabó, M., Ben Bouallègue, Z., Statistical post-processing of dual-resolution ensemble forecasts. *Q. J. R. Meteorol. Soc.* **145** (2019), 1705–1720.
8. Díaz, M., Nicolis, O., Marín, J. C., Baran, S., Statistical post-processing of ensemble forecasts of temperature in Santiago de Chile. *Meteorol. Appl.* **27** (2020), paper e1818, doi:10.1002/met.1818.
9. Díaz, M., Nicolis, O., Marín, J. C., Baran, S., Post-processing methods for calibrating the wind speed forecasts in central regions of Chile. *Ann. Math. Inform.* **53** (2021), 93–108.
10. Baran, S., Hemri, S., El Ayari, M., Statistical post-processing of water level forecasts using Bayesian model averaging with doubly-truncated normal components. *Water Resour. Res.* **55** (2019), 3997–4013.
11. Baran, S., Baran, Á., Pappenberger, F., Ben Bouallègue, Z., Statistical post-processing of heat index ensemble forecasts: is there a royal road? *Q. J. R. Meteorol. Soc.* **146** (2020), no. 732, 3416–3434.
12. Baran, Á, Lerch, S., El Ayari, M., Baran, S., Machine learning for total cloud cover prediction. *Neural. Comput. Appl.* **33** (2021), 2605–2620.
13. Baran, S., Szokol, P., Szabó, M., Truncated generalized extreme value distribution based EMOS model for calibration of wind speed ensemble forecasts. *Environmetrics* **32** (2021), paper e2678, doi:10.1002/env.2678.
14. Schulz, B., El Ayari, M., Lerch, S., Baran, S., Post-processing numerical weather prediction ensembles for probabilistic solar irradiance forecasting. *Sol. Energy* **220** (2021), 1016–1031.
15. Baran, S., Baran, Á., Calibration of wind speed ensemble forecasts for power generation. *Időjárás* **125** (2021), no. 4, 609–624.
16. Szépszó, G., Baran, Á., Baran, S., Jávorné Radnóczi, K., Kornyik, M., Tajti, D., Sugárzásra és magassági szélre vonatkozó rövidtávú előrejelzések operatív statisztikai utófeldolgozása. *Légkör* **68** (2023), 118–125.
17. Sikolya, K., Baran, S., On the optimal designs for the prediction of complex Ornstein-Uhlenbeck processes. *Comm. Statist. Theory Methods* **49** (2020), no. 20, 4859–4870.
18. Stehlík, M., Kišelák, J., Bukina, E., Lu, Y., Baran, S., Fredholm integral relation between ompound estimation and prediction (FIRCEP). *Stoch. Anal. Appl.* **38** (2020), no. 3, 427–459.
19. Burai, P., Kiss, G., Szokol, P., A dichotomy result for strictly increasing bisymmetric maps. *J. Math. Anal. Appl.* **526** (2023), no. 2, paper 127269, doi:10.1016/j.jmaa.2023.127269.