Computational Statistics and Data Analysis 75 (2014) 227–238

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components

Sándor Baran *

Faculty of Informatics, University of Debrecen, Kassai út 26, H-4028 Debrecen, Hungary
Institute of Applied Mathematics, University of Heidelberg, Im Neuenheimer Feld 294, D-69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 11 June 2013
Received in revised form 14 February 2014
Accepted 15 February 2014
Available online 22 February 2014

Keywords:
Continuous ranked probability score
Ensemble calibration
Truncated normal distribution

ABSTRACT

Bayesian model averaging (BMA) is a statistical method for post-processing forecast ensembles of atmospheric variables, obtained from multiple runs of numerical weather prediction models, in order to create calibrated predictive probability density functions (PDFs). The BMA predictive PDF of the future weather quantity is the mixture of the individual PDFs corresponding to the ensemble members and the weights and model parameters are estimated using forecast ensembles and validating observations from a given training period. A BMA model for calibrating wind speed forecasts is introduced using truncated normal distributions as conditional PDFs and the method is applied to the ALADIN-HUNEPS ensemble of the Hungarian Meteorological Service and to the University of Washington Mesoscale Ensemble. Three parameter estimation methods are proposed and each of the corresponding models outperforms the traditional gamma BMA model both in calibration and in accuracy of predictions.

© 2014 Published by Elsevier B.V.

1. Introduction

The most important aim of weather forecasting is to give a robust and reliable prediction of the future state of the atmosphere based on current observational data, prior forecasts and mathematical models describing the dynamical and physical behavior of the atmosphere. These models consist of sets of coupled hydro-thermodynamic non-linear partial differential equations of the atmosphere, ocean and land surface and only have numerical solutions. The difficulty with numerical weather prediction models is that since the atmosphere has a chaotic character the solutions strongly depend on the initial conditions and also on other uncertainties related to the numerical weather prediction process. Therefore, the results of such models are never fully accurate. A possible solution is to run the model with different initial conditions (since the lack of reliable set of the initial conditions is one of the most important sources of uncertainty) and produce an ensemble of forecasts. Using a forecast ensemble one can estimate the probability distribution of future weather variables which allows probabilistic weather forecasting (Gneiting and Raftery, 2005), where not only the future atmospheric states are predicted, but also the related uncertainty information. The ensemble prediction method was proposed by Leith (1974) and since its first operational implementation (Buizza et al., 1993; Toth and Kalnay, 1997) it has become a widely used technique all over the world. Recently users of meteorological forecasts more and more understand the merits of the method and its economic value as well. However, although e.g. the ensemble mean on average gives better forecasts of a meteorological quantity

* Correspondence to: Faculty of Informatics, University of Debrecen, Kassai út 26, H-4028 Debrecen, Hungary. Tel.: +36 52 512900 75212; fax: +36 52 512996.

E-mail address: baran.sandor@inf.unideb.hu.

http://dx.doi.org/10.1016/j.csda.2014.02.013
0167-9473/© 2014 Published by Elsevier B.V.

than any of the individual ensemble members, it is often the case that the ensemble is under-dispersive and in this way, uncalibrated (Buizza et al., 2005), so that calibration is needed to account for this deficiency.

The Bayesian model averaging (BMA) method for post-processing ensembles in order to calibrate them was introduced by Raftery et al. (2005). The basic idea of BMA is that for each member of the ensemble forecast there is a corresponding conditional probability density function (PDF) that can be interpreted as the conditional PDF of the future weather quantity provided the considered forecast is the best one. Then the BMA predictive PDF of the future weather quantity is the weighted sum of the individual PDFs corresponding to the ensemble members with weights based on the relative performance of the ensemble members during a given training period. The weight parameters and the parameters of the individual PDFs are estimated using linear regression and maximum likelihood (ML) method, where the maximum of the likelihood function is found by the EM algorithm. We remark that due to their flexibility mixture models play an essential role in data analysis (Böhning, 2014) and parameter estimation in mixture models is a typical application of the EM algorithm (see Dempster et al., 1977, McLachlan and Krishnan, 1997 or recently Lee and Scott, 2012, Chen and Lindsay, 2014). At the BMA calibration process one should also take into account whether the ensemble members can be distinguished clearly or some ensemble members are statistically exchangeable (see e.g. Fraley et al., 2010). In Raftery et al. (2005) the BMA method was successfully applied to obtain 48 h forecasts of surface temperature and sea level pressure in the North American Pacific Northwest based on the 5 members of the University of Washington Mesoscale Ensemble (Grimit and Mass, 2002). These weather quantities can be modeled by normal distributions, so the predictive PDF is a Gaussian mixture. Later, Sloughter et al. (2007) developed a discrete–continuous BMA model for precipitation forecasting, where the discrete part corresponds to the event of no precipitation, while the cubic root of the precipitation amount (if it is positive) is modeled by a gamma distribution. In Sloughter et al. (2010) the BMA method was used for wind speed forecasting and the component PDFs follow gamma distributions, while using von Mises distribution to model angular data, Bao et al. (2010) introduced a BMA scheme to predict surface wind direction. Finally, Sloughter et al. (2013) described a BMA model for wind vector forecasting, where the power transformed errors of wind vectors are modeled using a bivariate normal distribution.

The bivariate normal model for wind vectors is also used in the ensemble model output statistics (EMOS) method for post-processing ensemble forecasts (Schuhen et al., 2012). The EMOS, introduced by Gneiting et al. (2005) for calibrating ensemble forecasts following normal distribution (sea level pressure, temperature), produces a single normal PDF, where the mean and the variance depend on ensemble members. The method can be extended to truncated normal distribution (Thorarinsdottir and Gneiting, 2010), too, and in this way it can be used for calibrating wind speed data.

In the present paper we develop a BMA model for wind speed forecasting where the component PDFs, similar to the EMOS PDF of Thorarinsdottir and Gneiting (2010), follow a truncated normal distribution. The performance of the model is tested on the wind speed forecasts produced by the operational Limited Area Model Ensemble Prediction System (LAMEPS) of the Hungarian Meteorological Service (HMS) called ALADIN-HUNEPS (Hágel, 2010; Horányi et al., 2011) and on the forecasts of maximal wind speed of the eight-member University of Washington Mesoscale Ensemble (UWME, see e.g. Eckel and Mass, 2005). As a benchmark, in both case studies we investigate the goodness of fit of the gamma BMA model of Sloughter et al. (2010).

2. Data

2.1. ALADIN-HUNEPS ensemble

The ALADIN-HUNEPS system of the HMS covers a large part of Continental Europe with a horizontal resolution of 12 km and it is obtained by the dynamical downscaling (by the ALADIN limited area model) of the global ARPEGE based PEARP system of Météo France (Horányi et al., 2006; Descamps et al., 2009). The ensemble consists of 11 members, 10 initialized from perturbed initial conditions and one control member from the unperturbed analysis, implying that the ensemble contains groups of exchangeable forecasts. The data base contains 11 member ensembles of 42 h forecasts for 10 m wind speed (given in m/s) for 10 major cities in Hungary (Miskolc, Szombathely, Győr, Budapest, Debrecen, Nyíregyháza, Nagykanizsa, Pécs, Kecskemét, Szeged) produced by the ALADIN-HUNEPS system of the HMS, together with the corresponding validating observations for the period between October 1, 2010 and March 25, 2011 (176 days, or 1760 data points). The validating wind speed measurements are considered as instantaneous values (valid at a given time), however they are in fact mean values over the preceding 10 min. The model wind speed values are also considered as instantaneous, but they are representatives for a given model time step, which is 5 min in our case. This averaging (in both cases) removes some small scale noise and the gustiness of the wind, and the comparison of the modeled and observed values in this way is a common practice in meteorology. The validating observations were scrutinized by basic quality control algorithms including e.g. consistency check, and the assimilated observations at each ensemble member were quality controlled before their assimilation into the system. The forecasts are initialized at 18 UTC (8 pm local time when daylight saving time (DST) operates and 7 pm otherwise). The data set is fairly complete since there are only two days (18.10.2010 and 15.02.2011) where three ensemble members are missing for all sites and one day (20.11.2010) when no forecasts are available.

Fig. 1(a) shows the verification rank histogram of the raw ensemble. This is the histogram of ranks of validating observations with respect to the corresponding ensemble forecasts computed from the ranks at all stations and over the whole verification period (see e.g. Wilks, 2011, Section 7.7.2). This histogram is far from the desired uniform distribution as in many cases the ensemble members either underestimate or overestimate the validating observations (the ensemble

Download English Version:

https://daneshyari.com/en/article/416376

Download Persian Version:

https://daneshyari.com/article/416376

Daneshyari.com