Nat. Hazards Earth Syst. Sci., 13, 1161–1168, 2013
www.nat-hazards-earth-syst-sci.net/13/1161/2013/
doi:10.5194/nhess-13-1161-2013

Natural Hazards and Earth System Sciences
Open Access

Improving the calibration of the best member method using quantile regression to forecast extreme temperatures

A. Gogonel[1,2], **J. Collet**[1], **and A. Bar-Hen**[2]

[1]EDF R&D Division, OSIRIS Department, 1, avenue du Général de Gaulle, 92141 Clamart cedex, France
[2]MAP5, UFR de Mathématiques et Informatique, Université Paris Descartes, 45 rue des Saints-Pères, 75270 Paris cedex 06, France

Correspondence to: J. Collet (jerome.collet@edf.fr)

Received: 10 July 2012 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: –
Revised: 13 March 2013 – Accepted: 9 April 2013 – Published: 3 May 2013

Abstract. Temperature influences both the demand and supply of electricity and is therefore a potential cause of blackouts. Like any electricity provider, Electricité de France (EDF) has strong incentives to model the uncertainty in future temperatures using ensemble prediction systems (EPSs). However, the probabilistic representations of the future temperatures provided by EPSs are not reliable enough for electricity generation management. This lack of reliability becomes crucial for extreme temperatures, as these extreme temperatures can result in blackouts. A proven method to solve this problem is the best member method (BMM). This method improves the representation as a whole, but there is still room for improvement in the tails of the distribution. The idea of the BMM is to model the probability distribution of the difference between the forecast and realization. We improve the error modeling in BMM using quantile regression, which is more efficient than the usual two-stage ordinary least squares (OLS) regression. To achieve further improvement, the probability that a given forecast is the best one can be modeled using exogenous variables.

1 Introduction

The uncertainty of future temperatures is a major risk factor for an electric utility company such as Electricité de France (EDF). The demand for heating increases when the temperature is lower than 18 °C, and the demand for cooling increases when the temperature exceeds 18 °C. Moreover, high temperatures also create cooling problems for thermal plants.

To fulfill the risk management needs of the company, the ensemble prediction systems (EPSs) provided by weather forecasting institutes such as European Centre for Medium-Range Weather Forecasts (ECMWF) (ECMWF, 2002, 2006) provide an indispensable source of information. Ensemble forecasting is a numerical prediction method that is used to generate a representative sample of the possible future states of a dynamical system. Ensemble forecasting is a form of Monte Carlo analysis: multiple numerical predictions are computed using slightly different initial conditions, all of which are plausible given past and current observations. The available observations are combined to obtain estimates of the future temperatures as well as their uncertainties (Whitaker and Loughe, 1998) and predictive densities.

However, the probabilistic representations of future temperatures provided by EPSs suffer some lack of reliability, especially for extreme probabilities (for example 1 % quantile forecast is inaccurate), not only because the number of ensemble members is limited by computing resources, but also because EPS forecasts can be biased and typically do not display enough variability, thus leading to an underestimation of the uncertainty (Buizza et al., 2005). As a risk manager, and also because of its size and market power, EDF must use the most reliable information available (Diebold et al., 1998; Gneiting et al., 2007), so lack of reliability is prohibitive.

EDF faces also a regulatory constraint: the French technical system operator imposes that the probability of employing exceptional means (e.g., load shedding) to meet the demand for electricity must be lower than 1 % for each week (RTE, 2004), so EDF has to manage carefully risk at this level. Now, in France, most of the demand variability

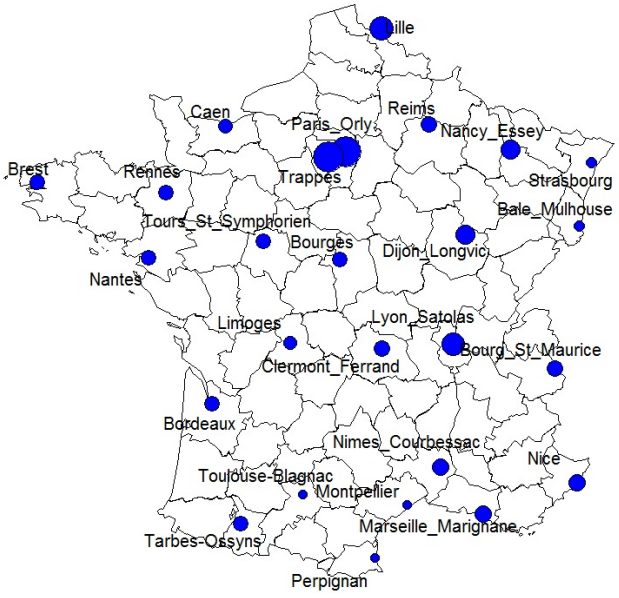

Fig. 1. Weather stations used to define global temperature for France: names, locations and circles with surfaces proportional to the weight of each station.

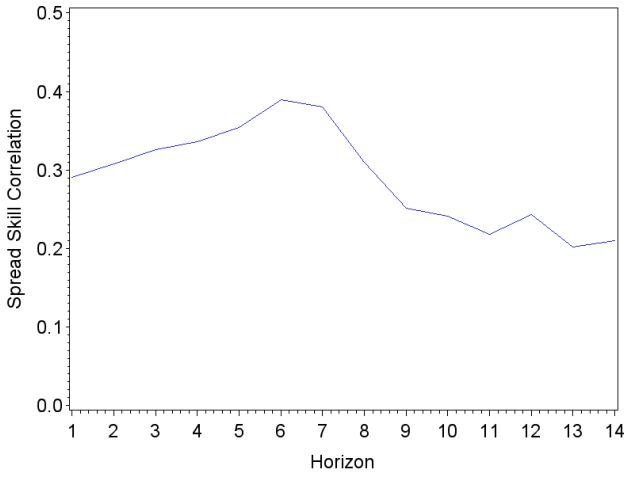

Fig. 2. Correlation between ensemble prediction system (EPS) interquartile range and forecast error, using EPS mean as a deterministic forecast.

is a consequence of temperature variability, with opposite effects in winter and summer: in winter demand increases when temperature decreases; the reverse is true in summer. Furthermore, variability in demand largely exceeds the variability in the supply. So, when the supply–demand balance is at the 1 % quantile, the temperature is nearly, but not exactly, at the 1 % quantile in winter, and at the 99 % quantile in summer. Needless to say, the EDF decision-making process takes account of other quantiles and quantities, but the regulatory constraint is the 1 % quantile.

A more technical limitation of EPSs is a lack of smoothness: the predictive cumulative distribution they provide is a step function, which can lead to problems in generation management tools.

Many methods have been developed to obtain a smooth and unbiased representation of the risk arising from temperature variations (Hagedorn, 2010; Wilks, 2011). However, as we are interested in extremes, it excludes methods assuming Gaussianity, such as the non-homogeneous Gaussian regression developed in Gneiting et al. (2005) and Unger et al. (2009). We then have to use one of the ensemble "dressing" methods: Bayesian model averaging (Raftery et al., 2004), Bayesian processor of output (Krzysztofowicz, 2004) or best member method (Roulston and Smith, 2002a). The best member method is the simplest amongst these three methods, and Gogonel-Cucu et al. (2011a) show that it gives better results on ECMWF ensemble forecasts than Bayesian model averaging. Furthermore, we suspect that using ranks, as proposed in Fortin et al. (2006), could be very useful to improve the representation of the tails, so it will be the basis of our modeling. In this paper, the key point is that the fitting of the dispersion of the "dressing" is improved when quantile regression is used in place of a two-stage ordinary least squares (OLS) regression.

In this paper, we first briefly review the use of the best member method (BMM) and then demonstrate the use of quantile regression to improve the error modeling required for the BMM. In the final section, we suggest several further improvements of the method.

2 An example application of the best member method

In this section, we briefly describe a simple application of the BMM to ECMWF forecasts, with a "dressing" depending on the rank of the forecast. We study all possible horizons, with each horizon treated independently of the others.

2.1 The data

The temperatures (in degrees Celsius) that we analyze are spatial and temporal averages. The time extent of the average is the day, and the spatial extent is France: we average the temperatures measured in 26 cities in France; the vector of 26 weights is chosen in order to estimate the electricity load in France (Dordonnat et al., 2008). The map in Fig. 1 shows the names, locations of the cities, as well as circles with surfaces proportional to the weight of each city. Because we use temperatures measured at specific observation sites, a statistical adaptation stage is included in the forecasting process. This statistical adaptation is performed by Météo-France.

The forecasts that we use are EPSs provided by ECMWF. As shown in Fig. 2, the ensemble spread is actually a proxy for the forecast error. Furthermore, the skill itself is seasonal, as demonstrated in Fig. 3.

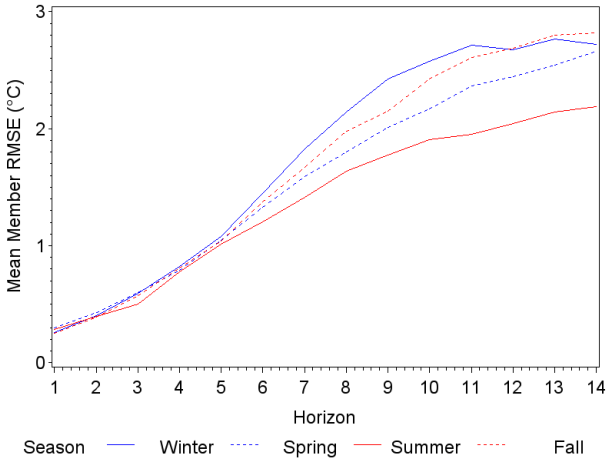

Fig. 3. Mean forecast error, using ensemble prediction system mean as a deterministic forecast, for all horizons and seasons.

The data consist of two different arrays. The first array is 3-dimensional and contains the forecasts.

- The dates of the forecasts range from 27 March 2007 to 30 April 2011 (for a total of 1473 dates).
- The horizons of the forecasts provided by ECMWF range from 1 to 14 days in the future.
- The member is identified by a number between 0 and 50. For a given date, a member corresponds to a given initial state.

The second array contains the realizations of the temperature variable. This array has 1 dimension corresponding to the date, with approximately the same extent as the forecasts.

The member 0 is different from others, as its initial conditions are unperturbed. Despite this, we consider the member number to be uninformative, and the members of the EPS are assumed to be exchangeable according to the definition provided in Bernardo (1996) (in contrast to multi-model ensembles; Gneiting et al., 2005).

2.2 The criteria

On can find many verification criteria for forecasts (see for example Jolliffe and Stephenson (2012)). As our goal is risk management, we will focus on criteria measuring the quality of uncertainty representation. That is why we use the probability integral transform (PIT) (Diebold et al., 1998; Hamill and Thomas, 2001). For each forecast horizon h and each date, we have a set of temperature simulations from BMM, and a realization. The PIT is defined as the proportion of the simulated values that are lower than the realization. If the forecast is reliable, then the PIT is uniformly distributed on $[0, 1]$. Formally, if we note y_t the sequence of realized temperatures, and F_t the sequence of density forecasts (if the F_t are continuous), we have

$$y_t \sim F_t \Rightarrow F_t(y_t) \sim U[0, 1].$$

The statistic $F_t(y_t)$ is called the PIT. So, measuring the reliability of a sequence of density forecasts comes down to comparing the distribution of the PIT to the uniform distribution. This condition can be checked in a straightforward manner using the empirical distribution function (EDF).

We also use continuous ranked probability score (CRPS), because this criterion is well known. The CRPS measures the difference between the forecast and observed cumulative distribution functions (CDFs). In the case of ensemble forecasts, when the size of the sample is sufficiently large, the forecasted CDF is approximated by the EDF of the ensemble. This definition has two consequences: The smaller the CRPS, the better, and CRPS has the same units as the observations.The CRPS can be decomposed in two components: the reliability component tests whether the forecast system has the correct statistical properties, whereas the potential CRPS measures the residual uncertainty of the forecast.

2.3 The best member method applied to ECMWF data

2.3.1 Principle

The BMM was first proposed by Roulston and Smith (2002a) and subsequently improved by Wang and Bishop (2005), Fortin et al. (2006) and in Gogonel-Cucu et al. (2011a,b). The last three references use the rank of a member in the sorted ensemble. As we state later in this section, this discrete variable is too detailed, so it is useful to aggregate some of its values.

The BMM is a statistical post-processing method, so using this method implies choosing a training period. In all cited works, and in our work, the training period is fixed and large (some months). We used cross-validation and stated that there is likely no effect of the training period on the performances of the method.

For each date and each horizon, the "best member" is the member closest to the realization (with the smallest absolute difference). The principle of the method is to model the probability distribution of the difference between the "best member" and the realization. The probabilistic forecast is then a convolution of the error distribution and the discrete distribution of the EPS members. Because all members of the EPS are exchangeable, all of the error models and weights are the same in the first step. An alternative idea, proposed by Fortin et al. (2006), is to use the rank of a member in the sorted ensemble. In this case, we first rank the forecasts for each date, according to the forecast value. The error model and weights are then functions of the rank. In order to write down some equations, we will use notations inspired from Fortin et al. (2006). For each date $t \in [1, T]$, we have the following:

- $(x_{t,(k)})_{k=1,\cdots,51}$ the sorted forecasts (using usual notations of order statistics) and

– y_t the realized temperature.

Then the best member is the x_t^* that minimizes $|x_{t,(k)} - y_t|$:

$$x_t^* = \arg\min_{x_{t,(k)}} |x_{t,(k)} - y_t|.$$

Similarly, the rank of the best member is

$$k_t^* = \arg\min_{k} |x_{t,(k)} - y_t|,$$

and the error of the best member is

$$\epsilon_t^* = y_t - x_t^*.$$

For each rank k, the weight assigned to the rank is

$$p_k = \left(\sum \mathbf{1}\{k_t^* = k\}\right)/T.$$

Because we prefer to use a parametric framework, the error model consists of the following:

1. a model for the mean μ_t of the error of the best member,
2. a model for the variance σ_t^2 of the error of the best member, and
3. a parametric distribution for ϵ_t^*.

Since the errors are unbounded, and approximately symmetric, we will use a normal distribution.

Many variables can be used to model the mean and variance of the error:

– variables summarizing features of the current EPS, such as the mean, square of the mean (to take account for the effect of extreme temperatures) and spread (here defined as the interquartile range) of the ensemble;

– variables to take account for a smooth influence of the date (the date t itself, $\cos(2\pi t/365)$ and $\sin(2\pi t/365)$);

– the rank of the best member k_t^*, considered as a discrete variable.

Furthermore, we may assume that the various discrete and continuous variables interact with one another: for example, the mean temperature of the EPS may have a different slope for each rank in the model.

An important point is that it is impossible to estimate all of the parameters simultaneously for all horizons as the variance of the residuals is approximately 3 times larger for horizon 14 than for horizon 1. All models are built separately for each horizon.

2.3.2 Rank aggregation

In this type of modeling, the rank of the member is a useful variable but is too detailed, at least in the case of ECMWF ensembles, as shown in Fig. 4.

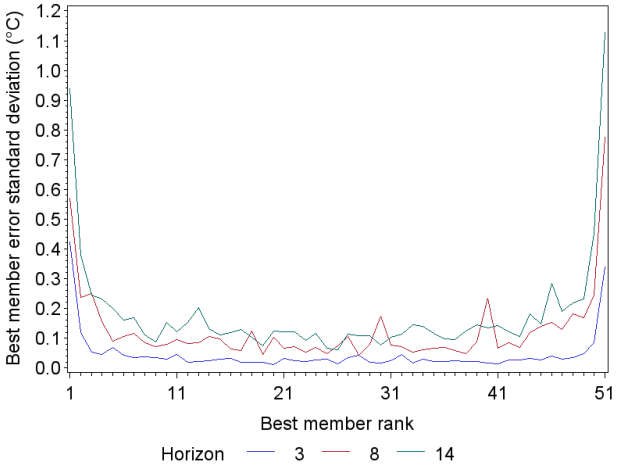

Fig. 4. Best member error (ϵ_t^*) standard deviation, with respect to best member rank (k_t^*), for 3 horizons: 3, 8 and 14 days.

All of the members with central ranks have nearly the same standard deviation in their errors; however, the errors of the members with extreme ranks behave very differently. We therefore propose to aggregate all of the central ranks. We define a new variable `edge` that is equal to 0 when the rank is central and equal to the rank otherwise. We must specify the definition of a "central rank", which will be made thereafter.

We apply automated variable selection method, such as the "stepwise" selection method (SAS, 2006), which provides some information on significant variables. So, for each possible definition of "central ranks", it is possible to select the significant variables, estimate their coefficient, and finally compute Akaike information criterion (AIC) or adjusted R^2. Following these criteria, the optimal choice is to consider the ranks $\{1, 2, 50, 51\}$ to be non-central and the remaining ranks to be central. Therefore, the value set of the variable `edge` is $\{0, 1, 2, 50, 51\}$. Finally, we obtain the following information.

– The rank never appears in the model; the relevant information carried by the rank is provided by the variable `edge`.

– The mean temperature is significant in modeling the mean of the error.

– The spread is significant in modeling the error variance, as are the mean temperature and its square.

– Nothing is gained by taking account of the date.

The following models were finally selected:

$$\begin{aligned} \epsilon_t^* &= a(\texttt{edge}_t) + b(\texttt{edge}_t)\cdot m_t + \eta_t \\ (\epsilon_t^* - \hat{\epsilon}_t^*)_t^2 &= \alpha(\texttt{edge}_t) + \beta(\texttt{edge}_t)\cdot m_t + \\ &\quad \gamma(\texttt{edge}_t)\cdot m_t^2 + \delta(\texttt{edge}_t)\cdot \texttt{IQR}_t + \xi_t, \end{aligned}$$

where

– ϵ_t^* is the error of the best member,

– m_t (respectively $\texttt{IQR}_t$) is the mean (respectively interquartile range) of the ensemble forecast,

– $\texttt{edge}_t$ is equal to the rank k_t^* if $k_t^* \in \{1, 2, 50, 51\}$, to 0 elsewhere, and

– η_t and η_t are the disturbances.

So, we may assign values to the parameters of the best member error model:

$$\mu_t = \hat{\epsilon_t^*}$$
$$\sigma_t^2 = \widehat{(\epsilon_t^* - \hat{\epsilon_t^*})^2}.$$

To account for possible evolution in the data, we used cross-validation, with a partitioning into two equal parts: we first estimated the BMM model on the period from 27 March 2007 to 12 April 2009, and used it on the period from 13 April 2009 to 30 April 2011, and we exchanged the roles in a second stage.

The BMM yields a substantial improvement in the reliability of the temperature density forecast, compared to raw EPS.

In Fig. 5, we plot the difference between the EDF of the PITs and the function $F_U(x) = x$ for horizon 3 for three temperature density forecasts: raw EPS, EPS post-processed using BMM, and EPS post-processed using BMM, with error modeling using quantile regression (which will be described in Sect. 3). We see the improvement in the reliability using the BMM is substantial for this horizon. Note we must plot the *difference* between the PIT and its theoretical distribution to observe the improvement; otherwise, we would see only two very close lines.

In order to compare raw EPS and BMM for all horizons, we need a numerical criterion: we will use continuous ranked probability score (CRPS) (Hersbach, 2000), and more specifically the reliability component of CRPS. We only performed statistical tuning, and we observe in Table 1 that the potential CRPS does not vary substantially. On the contrary, the reliability component is substantially decreased, except for horizon 5.

Furthermore, this reliability improvement substantially affects some practically relevant quantities. For example, the variance of the temperature 1 day ahead can be used as a rough risk indicator. Using the raw EPS, we obtain (on average) a value of 0.2 for this indicator, while we obtain a value of 0.44 using the BMM. For horizon 14, the values are 2.6 and 2.9. Note that this example does not prove that BMM is better than raw EPS; this proof is provided by comparison of the reliability component of CRPS. What it proves is the usefulness of the reliability improvement provided by BMM.

Despite the substantial reliability improvement using the BMM, the tails are still not well represented, as shown in Fig. 5.

Table 1. The continuous ranked probability score (CRPS) and its decomposition, for raw ensemble prediction system (column "Raw") and usual best member method (column "BMM"), for all horizons (column "h."). "BMM Score" is computed assuming that a perfect forecast has reliability component equal to 0. Temperature unit is degrees Celsius.

	Potential CRPS		Reliability component		
h.	Raw	BMM	Raw	BMM	BMM Score
1	0.24	0.25	2.6×10^{-2}	4.4×10^{-3}	83 %
2	0.31	0.32	1.6×10^{-2}	4.3×10^{-3}	73 %
3	0.38	0.39	6.9×10^{-3}	3.9×10^{-3}	43 %
4	0.49	0.50	4.7×10^{-3}	3.2×10^{-3}	32 %
5	0.63	0.64	3.5×10^{-3}	4.0×10^{-3}	−16 %
6	0.78	0.80	3.5×10^{-3}	2.2×10^{-3}	39 %
7	0.96	0.98	4.6×10^{-3}	2.2×10^{-3}	51 %
8	1.10	1.12	5.4×10^{-3}	2.4×10^{-3}	56 %
9	1.24	1.26	4.7×10^{-3}	2.4×10^{-3}	48 %
10	1.35	1.38	4.8×10^{-3}	2.7×10^{-3}	44 %
11	1.43	1.47	4.0×10^{-3}	2.3×10^{-3}	41 %
12	1.49	1.52	4.2×10^{-3}	2.2×10^{-3}	47 %
13	1.54	1.56	4.2×10^{-3}	3.0×10^{-3}	29 %
14	1.56	1.59	3.6×10^{-3}	3.1×10^{-3}	15 %

3 Improving the tail representation using quantile regression

3.1 Criteria to measure tail representation improvement

As we aim to model the tails, it is important to account for the *relative* errors in the *extreme* quantiles. Indeed, if we estimate the blackout risk is 2 %, while it is actually 1 %, we would plan much more emergency power supply than necessary. Therefore, we should neither employ the CRPS measure nor consider the PIT graph globally.

If accounting for the relative errors were the only issue, then we could employ measures based on likelihood ratio, e.g., the ignorance score (Roulston and Smith, 2002b) and its decompositions (Weijs et al., 2010; Tödter, 2011), or even statistical tests of the goodness-of-fit (Jager and Wellner, 2005).

However, in this study, we know the range of probabilities that are of interest: around the 1 % quantile and around the 99 % quantile. Therefore, to measure the dissimilarity between the PIT distribution and the uniform distribution, we use a χ^2 distance with the classes $[0; 0.01]$, $[0.01; 0.02]$ and $[0.02; 0.05]$ for the lower tail and the symmetric classes for the upper tail. This strategy will be used to assess all of the improvements.

More formally, we will compute

$$D = \sum_i \frac{\left((\hat{F}_{\mathrm{PIT}}(M_i) - \hat{F}_{\mathrm{PIT}}(m_i)) - (M_i - m_i)\right)^2}{M_i - m_i},$$

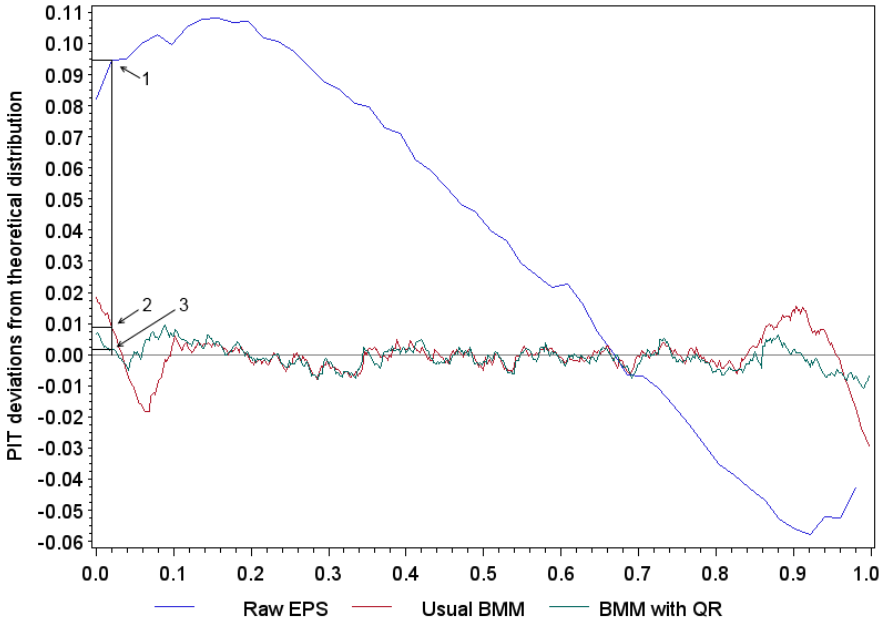

Fig. 5. Comparison of probability integral transform deviations from theoretical distribution for horizon 3. The point labeled "1", with coordinates $(0.02, 0.095)$, shows that, for raw ensemble prediction system ("raw EPS"), the realized temperature is lower than 2 % of the simulated values on $2 + 9.5 = 11.5$ % (rather than 2 %) of the dates. For best member method, denoted by "Usual BMM" (respectively best member method with error modeling using quantile regression, denoted by "BMM with QR"), point labeled "2" (respectively "3") realized temperature is lower than 2 % of the simulated values on 2.9 % (respectively 2.2 %) of the dates.

where

- $m_i = (0; 0.01; 0.02; 0.95; 0.98; 0.99)$ is the sequence of the classes lower bounds, and
- $M_i = (0.01; 0.02; 0.05; 0.98; 0.99, 1)$ is the sequence of the classes upper bounds.

3.2 Grounds for using quantile regression

We are interested in the tail representation, so we have to focus on the error of the extreme members. The dispersion of distribution of these errors is much larger (approximately 5 times) than the errors of central members. So, a first possibility to improve the quality of tail representation is to improve the estimation of the dispersion.

The error distribution of the extreme members is highly asymmetric, with some very large values. That is why using mean and standard deviation to locate and scale a normal distribution has no theoretical ground. We therefore propose to model two different quantiles ($Q_{1/3}$ and $Q_{2/3}$) of the error in a single stage, using quantile regression (Koenker and Bassett, 1978), with the same regressors as in OLS regression. Then, we choose the mean and variance of the normal distribution such that its quantiles $Q_{1/3}$ and $Q_{2/3}$ are equal to the empirical quantiles. The key point is that this estimation does not rely on an assumed distribution.

3.3 Implementation

We modeled the quantiles $Q(1/3)$ and $Q(2/3)$ of the error of best member ϵ_t^*, using the variable $\texttt{edge}$ together with ensemble mean and spread. For both quantiles, the chosen model is

$$Q(i/3)_t = \alpha(\texttt{edge}_t) + \beta(\texttt{edge}_t) \cdot m_t + \gamma(\texttt{edge}_t) \cdot m^2 + \delta(\texttt{edge}_t) \cdot \texttt{IQR}_t,$$

where

- $i = 1, 2$ is the index of the tercile,
- m_t (respectively $\texttt{IQR}_t$) is the mean (respectively interquartile range) of the ensemble forecast, and
- $\texttt{edge}_t$ is equal to the rank $k_t^* *$ if $k_t^* * \in \{1, 2, 50, 51\}$, to 0 elsewhere.

As we aim to simulate the errors using the normal distribution, the normal parameters are those required to obtain the estimated third and first terciles. We have

$$\mu_t = \frac{Q(2/3)_t + Q(1/3)_t}{2}$$
$$\sigma_t = \frac{Q(2/3)_t - Q(1/3)_t}{2 \times \phi^{-1}(2/3)},$$

where ϕ is the cumulative normal distribution function and $\phi^{-1}(2/3) = 0.431$. The rest of the model remains the same as in Sect. 2.3.

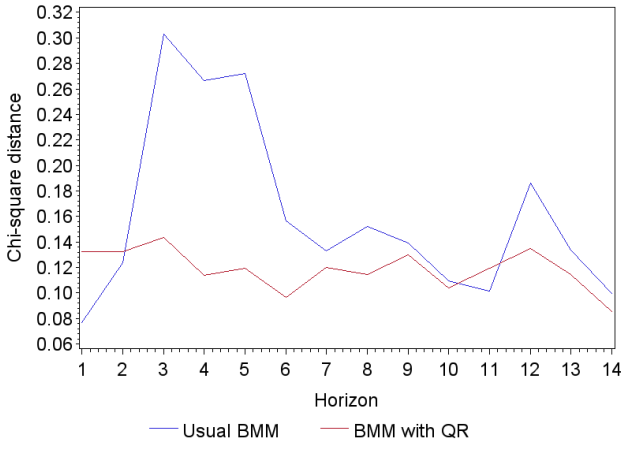

Fig. 6. Plot of χ^2 distances between theoretical and empirical distributions of Probability Integral Transform, with respect to horizon, for Best Member Method (BMM) and BMM with error modeling using quantile regression (BMM with QR).

The results are as follows. First, we consider the third PIT plot of Fig. 5: this third plot is a bit closer to the horizontal axis than the second one, which represents the usual BMM. We also need to compare BMM and BMM with error modeling using quantile regression for all horizons, so we use χ^2 distances, plotted in Fig. 6. The improvement is substantial, and there is a slight degradation of the forecast for only 3 of the horizons (out of 14), implying that the tails are represented much more accurately.

4 Conclusions

In this study, we used ECMWF temperature ensemble forecasts in order to improve the tails of temperature density forecasts, which are an important input in electrical blackout risk management. Any improvement enables a more accurate estimate of the required power supplies, resulting in substantial cost reductions.

Because the ECMWF temperature ensemble forecasts are under-dispersive, we modified the BMM improvement proposed by Fortin et al. (2006). All of the central ranks behave similarly, and we can therefore aggregate them, resulting in a far more parsimonious model. This strategy is most likely applicable to other under-dispersive ensemble forecasts as well.

Another improvement introduced in this paper is the use of quantile regression to determine the location and scale of the error distribution. This method is widely applicable, whenever the location and scale of a distribution must be determined for simulation purposes.

Moderate biases remain in the tail representations of the temperature density forecasts. The following strategies may provide further improvement in future work.

– The distribution of the errors of the extreme members is highly asymmetric, with some very large values, and it suggests using the extreme value distribution to model it.

– In our modeling, the probability that a member of a given rank is the best member does not depend on any exogenous variables. Testing and possibly rejecting this independence assumption may help to improve the tail estimation.

Acknowledgements. Authors are pleased to thank the reviewers for their comments and suggestions, which have been essential for improving the manuscript, its clarity and readability.

Edited by: A. Mugnai
Reviewed by: three anonymous referees

References

Bernardo, J.-M.: The concept of exchangeability and its applications, Far East J. Math. Sci., Special Volume, 111–121, 1996.

Buizza, R., Pellerin, G., Toth, Z., Zhu, Y., and Wei, M.: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems, Mon. Weather Rev., 133, 1076–1097, doi:10.1175/MWR2905.1, 2005.

Diebold, F. X., Gunther, T. A., and Tay, A. S.: Evaluating Density Forecasts with Applications to Financial Risk Management, Int. Economic Rev., 39, 863–883, 1998.

Dordonnat, V., Koopman, S.-J., Ooms, M., Dessertaine, A., and Collet, J.: An hourly periodic state space model for modelling French national electricity load, Int. J. Forecast., 24, 566–587, doi:10.1016/j.ijforecast.2008.08.010, 2008.

ECMWF: The Ensemble Prediction System, Tech. rep., ECMWF, 2002.

ECMWF: Part V: The Ensemble Prediction Systems, Tech. rep., ECMWF, 2006.

Fortin, V., Favre, A.-C., and Saïd, M.: Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member, Q. J. Roy. Meteorol. Soc., 132, 1349–1369, doi:10.1256/qj.05.167, 2006.

Gneiting, T., Raftery, A. E., Westveld III, A. H., and Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, Mon. Weather Rev., 133, 1098–1118, doi:10.1175/MWR2904.1, 2005.

Gneiting, T., Balabdaoui, F., and Raftery, A. E.: Probabilistic forecasts, calibration and sharpness, J. Roy. Stat. Soc. Ser. B, 69, 243–268, doi:10.1111/j.1467-9868.2007.00587.x, 2007.

Gogonel-Cucu, A., Collet, J., and Bar-Hen, A.: Implementation of two Statistic Methods of Ensemble Prediction Systems for Electric System Management, Case Studies in Business, Industry and Government Statistics (CSBIGS), accepted, 2011a.

Gogonel-Cucu, A., Collet, J., and Bar-Hen, A.: Implementation of Ensemble Prediction Systems Post-Processing Methods, for Electric System Management, in: 58th Congress of the International Statistical Institute, 2011b.

Hagedorn, R.: Post-Processing of EPS Forecasts, available at: http://www.ecmwf.int/newsevents/training/meteorological_presentations/pdf/PR/Calibration.pdf (last access: 26 April 2013), 2010.

Hamill, T. M. and Thomas, M.: Interpretation of Rank Histograms for Verifying Ensemble Forecasts, 129, 550–560, doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2, 2001.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather Forecast., 15, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Jager, L. and Wellner, J. A.: A new goodness-of-fit test: the reversed Berk-Jones statistic, Tech. Rep. TR443, Department of Statistics, University of Washington, available at: http://www.stat.washington.edu/www/research/reports/2004/tr443.pdf (last access: 26 April 2013), 2005.

Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification: A Practitioner's Guide in Atmospheric Science, Wiley, 2012.

Koenker, R. and Bassett, G.: Regression Quantiles, Econometrica, 46, 33–50, 1978.

Krzysztofowicz, R.: Bayesian Processor of Output: a new technique for probabilistic weather forecasting, in: 17th Conference on Probability and Statistics in the Atmospheric Sciences, available at: https://ams.confex.com/ams/pdfpapers/69608.pdf (last access: 26 April 2013), 2004.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian Model Averaging to Calibrate Forecast Ensembles, Mon. Weather Rev., 133, 1155–1174, doi:10.1175/MWR2906.1, 2004.

Roulston, M. and Smith, L.: Combining dynamical and statistical ensembles, Tellus, 55A, 16–30, doi:10.1034/j.1600-0870.2003.201378.x, 2002a.

Roulston, M. and Smith, L.: Evaluating probabilistic forecasts using information theory, Mon. Weather Rev., 130, 1653–1660, doi:10.1175/1520-0493(2002)130<1653:EPFUIT>2.0.CO;2, 2002b.

RTE: Mémento de la sûreté du système électrique, édition 2004, Tech. rep., Réseau de Transport d'Électricité, available at: http://www.rte-france.com/uploads/media/pdf_zip/publications-annuelles/memento_surete_2004_complet__.pdf (last access: 26 April 2013), 2004.

SAS: The GLMSelect Procedure: Stepwise Selection(STEPWISE), available at: http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_glmselect_a0000000241.htm (last access: 26 April 2013), 2006.

Tödter, J.: New Aspects of Information Theory in Probabilistic Forecast Verification, Ph.D. thesis, MS thesis, Institute of Theoretical Physics, Goethe-University Frankfurt, 138 pp., 2011.

Unger, D., van den Dool, H., O'Lenic, E., and Collins, D.: Ensemble regression, Mon. Weather Rev., 137, 2365–2379, doi:10.1175/2008MWR2605.1, 2009.

Wang, X. and Bishop, C.: Improvement of ensemble reliability with a new dressing kernel., Q. J. R. Meteorol. Soc., 131, 965–986, 2005.

Weijs, S., Van Nooijen, R., and Van De Giesen, N.: Kullback-Leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition, Mon. Weather Rev., 138, 3387–3399, doi:10.1175/2010MWR3229.1, 2010.

Whitaker, J. and Loughe, A.: The Relationship between Ensemble Spread and Ensemble Mean Skill, Mon. Weather Rev., 126, 3292–3302, doi:10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2, 1998.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, Elsevier, 2nd Edn., 2011.