Modeling Forecast Uncertainty Using Fuzzy Clustering

Ashkan Zarnani[1] and Petr Musilek[1,2]

[1] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada
{azarnani,petr.musilek}@ualberta.ca
[2] Department of Computer Science, VSB – Technical University of Ostrava, 17. Listopadu 15, Ostrava, Czech Republic

Abstract. Numerical Weather Prediction (NWP) systems are state-of-the-art atmospheric models that can provide forecasts of various weather attributes. These forecasts are used in many applications as critical inputs for planning and decision making. However, NWP systems cannot supply any information about the uncertainty of the forecasts as their immediate outputs. In this paper, we investigate the application of Fuzzy C-means clustering as a powerful soft computing technique to discover classes of weather situations that follow similar forecast uncertainty patterns. These patterns are then utilized by distribution fitting methods to obtain Prediction Intervals (PIs) that can express the expected accuracy of the NWP system outputs. Three years of weather forecast records were used in a set of experiments to empirically evaluate the applicability of the proposed approach and the accuracy of the computed PIs. Results confirm that the PIs generated by the proposed post-processing procedure have a higher skill compared to baseline methods.

1 Introduction

Although the deterministic interactions of physical simulations in Numerical weather prediction (NWP) models yield the expected values of different weather attributes in the mid-range future, such forecasts are uncertain due to the inaccuracy of initial conditions, low spatial resolution, and various simplifying assumptions [12][13]. Yet, such uncertainty information is not available in the immediate outputs of the system.
In many applications, it is desirable that forecasts be accompanied by the corresponding uncertainties. Information about forecast uncertainty can have important role in the planning and decision making processes that utilize the forecasts [2][8]. For instance, the expected accuracy of NWP temperature and wind speed forecasts can have crucial impact on the optimized operational planning and management of power grids using Dynamic Thermal Rating (DTR) systems which is the motivation of this study [6] [10]. The uncertainty of a forecast is typically presented using prediction intervals (PIs) that are accompanied by a percentage expressing the level of confidence, or expected nominal coverage rate (e.g., T = [2°C, 14°C] conf = 95%) [2] [5].
A major common method to assess the uncertainty of weather forecasts is ensemble modeling. However, running multiple ensemble members to analyze the forecast accuracy can be very costly thus infeasible in many cases.

V. Snasel et al. (Eds.): SOCO Models in Industrial & Environmental Appl., AISC 188, pp. 287–296.
springerlink.com

As an alternative, statistical post-processing methods can be applied on a result of an individual forecast. It is a well-known fact that the extent of forecast uncertainty varies with its context: the weather situation [13]. For example, low pressure systems are known to be less predictable than the more stable high pressure systems.

Soft computing techniques are increasingly applied in problems with large amount of data and uncertainty [3][14]. Lange *et al.* [9] used clustering over a historical performance data set of wind speed predictions and demonstrated a relationship between the forecast uncertainty and different meteorological situations. However, this analysis was not practically employed as a method of obtaining PIs for wind speed forecasts.

A practical application of weather classification to obtain PIs was proposed by Pinson et al. [17][18]. The authors used two predicted values of wind speed and wind power to categorize the historical forecast situations into four manually defined classes, each with different error distribution. The distribution of a new forecast case was then expected to follow the distribution of these classes based on an expert-based fuzzy membership definition. However, this method suffers from a major shortcoming of the manual grouping of predictions.

In this contribution, we use unsupervised learning over the historical performance of the NWP model to learn the patterns of forecast accuracy. To discover groups of forecast records that follow a similar prediction error distribution, Fuzzy C-means clustering algorithms is applied on a data set of past prediction accuracy records. Such objective-driven discovery of forecast situations is expected to find better groups compared to the manual definition of weather situations [15]. In addition, fuzzy association of forecast records with the discovered weather situations appears to be a more natural choice.

The process of evaluating of PIs forecasts, and probabilistic forecasts in general, is more complex compared to point forecasts. To empirically test the proposed approach, we apply the developed PI models to a large, real-world data set. We also develop a comprehensive PI evaluation framework. It not only covers all major measures from the PI evaluation literature, but also brings new insights to the PI verification process, leading to more accurate judgments.

The rest of this paper is organized as follows. Section 2 reviews the basic concepts and definitions of prediction intervals and forecast uncertainty modeling. Section 3 presents the proposed fuzzy clustering approach to discover forecast uncertainty patterns. The verification measures are described and analyzed in section 4. Experimental setup and results are provided in section 5. The final section 6 outlines main conclusions and indicates possible directions for future work.

2 Forecast Uncertainty and Prediction Intervals

The relation between the forecast $\hat{y}_t$ and its observation y_t can be described as:

$$y_t = \hat{y}_t + e_t, \tag{1}$$

i.e., each observation can be decomposed to the predicted value $\hat{y}_t$ for time *t*, and an error term e_t for the specific forecast instance.

Based on a probabilistic forecast, the cumulative distribution function (cdf) F_{y_t} is explicitly available. The prediction interval I_t^α is defined as $(1-\alpha)$-confidence interval into which observation y_t is expected to fall with probability $1-\alpha$. Therefore, it can be described as a range satisfying [5][15][17]:

$$P(y_t \in I_t^\alpha) = P(y_t \in [L_t^\alpha, U_t^\alpha]) = 1-\alpha, \tag{2}$$

where L_t^α and U_t^α are, respectively, the lower and upper bound of prediction interval I_t^α defined by the corresponding distribution quantiles as:

$$L_t^\alpha = q_{y_t}^{\alpha_l=(\alpha/2)} = F_{y_t}^{-1}(\alpha/2),\ \ U_t^\alpha = q_{y_t}^{\alpha_u=(1-\alpha/2)} = F_{y_t}^{-1}(1-\alpha/2). \tag{3}$$

For instance, with $\alpha = 0.05$, the interval has a 95% confidence level bounded by quantiles $L_t^{0.05} = q_{y_t}^{0.025}$ and $U_t^{0.05} = q_{y_t}^{0.975}$, as $\alpha_l = 0.025$ and $\alpha_u = 0.975$.
Systematic characterization of forecast error can lead to modeling of forecast uncertainty for the target variable. This can be achieved by considering e_t in (1) as an instance of the random variable e, and associating F_t^e (or its estimate $\hat{F}_t^e$) as its cumulative distribution function, The corresponding estimated quantiles for the predictive distributions would hence be $\hat{L}_t^\alpha$ and $\hat{U}_t^\alpha$ [15][22][21]:

$$\hat{L}_t^\alpha = \hat{y}_t + \hat{q}_{e,t}^{(\alpha/2)},\ \ \hat{q}_{e,t}^{(\alpha/2)} = \hat{F}_t^{e^{-1}}(\alpha/2), \tag{4}$$

$$\hat{U}_t^\alpha = \hat{y}_t + \hat{q}_{e,t}^{(1-\alpha/2)},\ \ \hat{q}_{e,t}^{(1-\alpha/2)} = \hat{F}_t^{e^{-1}}(1-\alpha/2), \tag{5}$$

where $\hat{q}_{e,t}^{(\alpha)}$ is the estimated α quantile of "error" based on the estimated forecast error distribution $\hat{f}_t^e$. The distribution of y_t, and hence the desired quantiles, are not explicitly known. Therefore, to find the $\hat{I}_t^\alpha$ prediction interval of y_t, the quantiles of e (i.e., the error associated with the forecast) are estimated and added to the predicted value $\hat{y}_t$ to obtain the lower and upper bounds for the original variable [17]. Thus, by finding quantiles over the forecast error distribution, one can find the quantiles over the forecast value that is expected to enclose the target observation.

3 Fuzzy C-means for Prediction Interval Modeling

A fine grouping of forecast situations can lead to clusters of predictions with a similar error behavior [13]. Such groupings can be found by clustering all available cases using the relative influential prediction variables as the features. Subsequently, each cluster can be independently analyzed by the method described in the previous section. This way, rather than considering all past errors together as a single set, the characteristics of error distribution within each cluster determines the prediction interval of that particular cluster.

In this study, we apply two different clustering algorithms to find optimal groupings of the NWP past forecasts: K-means [20] and Fuzzy C-means (FCM) clustering. K-means is a simple yet powerful clustering algorithm that has been used in many applications [19]. Consider a dataset $D = \{\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_N\}$, where each data point $\mathrm{x}_j = \{x_j^1, x_j^2, \dots, x_j^d\}$ represents d influential features (such as predicted temperature, wind speed and wind direction, precipitation, location, elevation, etc.), and N is the total number of available forecast cases. The algorithm finds the set of k cluster centers $C = \{c_1, c_2, \dots, c_k\}$, and assigns a subset of points $D_i \in D$ to each cluster i. Each case j also has a forecast error e_j^y associated with the predictand y. Hence, each cluster has its own set of forecast errors for target y in the set E_i^y such that:

$$E_i^y = \left\{e_j^y | x_j \in D_i, j = 1..n^i\right\}, i = 1..k, \quad (6)$$

where n^i is the number of sample points in cluster i.

In the second stage of the process, a probability distribution ($\hat{F}_{i,t}^e$) is fitted over each set of errors $E_i^y, i = 1..k$ to represent the forecast error characteristics of each cluster. We consider three fitting schemes: Gaussian distribution, Kernel Density Smoothing (using a Gaussian kernel) and Empirical distribution.

For instance, based on the Gaussian fitting method, each cluster i of forecast errors has its own estimated probability distribution described by its mean $\hat{\mu}_e^i$ and standard deviation $\hat{\sigma}_e^i$. When a new forecast x_{new} is made, the cluster to which it belongs can be identified by the nearest cluster center and boundaries of the corresponding prediction interval can be estimated using $\hat{F}_{i,t}^e$ instead of $\hat{F}_t^e$ in equations (4) and (5) which would provide $\hat{L}_i^\alpha$ and $\hat{U}_i^\alpha$ for each cluster independently.

Using K-means, each forecast case is assigned into a single cluster only. In a more natural approach, the forecast cases could be associated with various situations up to different degrees. This can be achieved using Fuzzy C-means algorithm that finds cluster patters based on fuzzy membership assumption of points over clusters [20]. The objective function of the clustering process is [1]:

$$J = argmin_C \sum_{i=1}^{N} \sum_{j=1}^{k} u_{ij}^m \left\|x_i - c_j\right\|^2, \quad (7)$$

where u_{ij} ($\sum_{l=1}^{k} u_{il} = 1$) represents the degree of membership of the point x_i in cluster j, and $m > 1$ is the fuzzification. The objective function can be minimized using gradient descent in an iterative process where the membership matrix and cluster centers are updated as follows:

$$u_{ij} = 1 / \sum_{l=1}^{k} \left(\frac{\|x_i - c_j\|}{\|x_i - c_l\|}\right)^{2/m-1} \quad (8)$$

As a result, a forecast case can simultaneously belong to more than one forecast situations. Many situations, such as transitions between different types of weather, can be better captured using this approach. Similarly to K-means, these fuzzy patterns of historical forecasts can be used to model forecast errors by fitting appropriate distributions. However, E_i^y is now a fuzzy set defined by membership of each error sample to the i^{th} cluster (i.e. $u_{li,}, l = 1..N$). Subsequently, the fitting methods must consider these membership values as the vector of sample weights in the process. Thus, error samples that have higher levels of association with a cluster have more impact on the corresponding error distribution.

In addition, any new forecast case x_{new} is now associated with all k clusters, but with different degrees of membership, $u_{new,j}$, $j = 1..k$. Hence, we need to devise a method to combine the error characteristics of different clusters, based on the new samples membership values. For this purpose, we apply a weighted opinion pool to consolidate the forecast error characteristics among the clusters. Because $\sum_{j=1}^{k} u_{new,j} = 1$, the weighted sum of the computed quantiles in each cluster based on the new forecast's levels of membership provides an intuitive method to obtain the final upper and lower quantiles:

$$\hat{L}_{new}^{\alpha} = \sum_{j=1}^{k} u_{new,j} \cdot \hat{L}_j^{\alpha}, \tag{9}$$

where $\hat{L}_j^{\alpha}$ represents the lower quantile of the prediction interval in the j^{th} fuzzy cluster. The same method is used to compute the upper quantile.

4 Prediction Interval Verification

It is expected that, in a test setting, prediction interval forecasts will have empirical coverage of the observations as close as possible to their confidence level. This primary property of a PI forecaster M, called "reliability," is denoted Rel_M^{α}. [17]:

$$\bar{\xi}_M^{I^{\alpha}} = \frac{1}{T}\sum_{i=1}^{T} \xi_i^{I^{\alpha}}, \text{ where } \xi_i^{I} = \begin{cases} 1 & if\ \hat{L}_{\hat{y}_t}^{\alpha} \le y_t \le \hat{U}_{\hat{y}_t}^{\alpha} \\ 0 & \text{empirical}, \end{cases} \tag{10}$$

where T is the number of PIs in the evaluation data set, and ξ_i^I is an indicator of hit. ξ_i^I evaluates to one when the observation falls within the PI boundaries, otherwise it is set to zero, expressing a miss. Hence, Rel_M^{α} simply accounts for the difference between average hit of the forecasts (coverage rate) and the required nominal coverage defined for the PI.

A forecaster providing PIs with less vagueness, corresponding to the width of the PI, is clearly preferred. This leads to the second major measure of PI forecast quality called "sharpness" [11][17]:

$$Shp_M^{\alpha} = \overline{Width}_M^{\alpha} = \frac{1}{T}\sum_{i=1}^{T} Width_i^{\alpha} \tag{11}$$

where $Width_i^{\alpha} = \hat{U}_{\hat{y}_i}^{\alpha} - \hat{L}_{\hat{y}_i}^{\alpha}$ is the width of the i^{th} prediction interval. Another important quality aspect of a PI computation method is its ability to provide intervals of variable width, depending on the forecast situation. A method with high "resolution" (Res_T^{α}) is capable of distinguishing forecasts with different amounts of uncertainty, and assign wider (high uncertainty) or narrower (low uncertainty) intervals accordingly. The standard deviation of PI widths is a natural choice to measure the method's resolution [15]:

$$Res_T^{\alpha} = \left[\frac{1}{T-1}\sum_{j=1}^{T}\left(\hat{U}_j^{\alpha} - \hat{L}_j^{\alpha} + Shp_M^{\alpha}\right)^2\right]^{\frac{1}{2}} \tag{12}$$

Having access to a single scalar summary measure of forecast quality is always attractive and useful for objective comparison of various methods. The most common prediction interval skill score is the Winkler's score [7], widely used as a conclusive objective evaluation measure for PI forecasting methods [11], [15], [18]. A comprehensive study performed by Gneiting and Raftery [4] prove that this score is "strictly proper" and would hence give the maximum score to a forecast that is actually the true belief of the forecaster and cannot be "hedged".

Using the notations defined above and the overall miss rate $(1-\bar{\xi}_M^{I^\alpha})$, the total score gained by a PI forecasting method M over the T cases in the test set can be expressed as:

$$SScore_M = T\left(-\frac{\alpha}{2}\overline{Width}_M^\alpha - (1-\bar{\xi}_M^{I^\alpha})\bar{\delta}_M^\alpha\right) = -T\left(\frac{\alpha}{2}\overline{Width}_M^\alpha + \bar{\Delta}_M^\alpha\right), \quad (13)$$

where $\bar{\delta}_M^\alpha$ is the average distance of an observation from the PI boundaries among the missed cases, and $\bar{\Delta}_M^\alpha$ is the average of this distance among all test cases owing to the fact that Δ_i is equal to zero for hit cases and δ_i for misses.

Due to availability of limited number of test cases in each cluster, the $SScore_i$ measurements incur some uncertainty as well. The width component of this score is constant in each cluster. However, the $\bar{\Delta}_M^{\alpha,j}$ measure's uncertainty (where $j=1..K$) decreases when evaluated by more test cases or when its sample values are closer to each other cluster j. To analyze the uncertainty of $\widehat{\mathrm{SScore}}_\mathrm{M}$, the one-sided confidence interval of the $\bar{\Delta}_M^{\alpha,j}$ measure with a specific confidence level is used to compute the skill score. After using this upper limit for all clusters, a lower limit on the $SScore_M$ with the desired confidence level can be determined:

$$P\left(\bar{\Delta}_M^{\alpha,j} < \bar{\Delta}_M^{\alpha,j^\beta}\right) = \beta \Rightarrow P\left(SScore_M > SScore_M^\beta\right) = \beta \quad (14)$$

where β is the desired confidence level over the measure as a percentage. Because $\bar{\Delta}_M^{\alpha,j}$ is a mean statistic, the Central Limit Theorem [21] can be used and hence its sampling distribution is essentially Gaussian. This leads to the following relation to obtain the one-sided confidence interval:

$$\bar{\Delta}_M^{\alpha,j^\beta} = \bar{\Delta}_M^{\alpha,j} + t\left(\beta, |T_j| - 1\right)\frac{s_{\Delta_M^{\alpha,j}}}{\sqrt{|T_j|}} \quad (15)$$

where $\bar{\Delta}_M^{\alpha,j}$ is the measured sample mean over the available sample test set, and $s_{\Delta_M^{\alpha,j}}$ is the sample standard deviation of individual $\Delta_i^{\alpha,j}$ values in cluster j. Hence, we can find the lower limit of the true $\bar{\Delta}_M^{\alpha,j}$ measure.

5 Experimental Evaluation

A hindcast data set of hourly predictions has been coupled with the respective observations of weather stations from the National Center for Atmospheric Research (NCAR) data repository. The WRF v3 simulations were run in three nested grids with resolutions of 10.8 km, 3.6 km and 1.2 km. The data set covers three years (2007, 2008 and 2009) of forecasts for two stations in BC. This data set contains about 51,000 records of historical performance of forecasts. There are total of 35 features available in this data set. The observations are used to derive the forecast error for temperature forecasts, and the described PI computation methods are applied to obtain prediction intervals for the forecasted temperature.

To investigate the role of influential variables and to select the optimal feature set in PI forecasts, 14 different subsets of the 25 available features were defined. These feature sets are combinations of BF1 (10 basic weather attributes), BF2 (a more complex feature set including attributes at different geopotential levels) and PG (derived features that represent the temporal gradient of surface pressure). The feature sets with letters PC*x* were obtained using Principal Component Analysis to decrease the dimensionality of the data to *x*. The results are based on three-fold cross-validation in which two years of data are used to train the PI model and the third year is set aside just to evaluate the trained model and calculate the quality measures of the resulting interval forecasts.

To compare the various proposed methods with baseline methods, some basic approaches are considered. The first baseline method is the *climatological* approach that considers all past error samples together (i.e. K=1) and computes the PI based on the fitted distribution. The second baseline method applies a manual grouping of the forecast situations based on the forecast *month*. In the evaluated approaches, the number of clusters was set in the range of 2 to 100, and the fuzzification parameter (m) in FCM was set to 1.2. Table 1 lists the PI quality details of the best performing setups for each algorithm. The results show that clustering methods considerably improve the skill of the PIs compared to the baseline methods and that FCM has a better performance compared to K-means.

Table 1. Top 4 setups from C-means and K-means along with baseline methods and detailed measures for temperature PIs based on Sscore$^{0.95}$ in 3-fold (yearly) cross validations

Algorithm	K	Features	Fit	Sharp-ness	Cov-erage	Cover-age$^{0.95}$	Reso-lution	RMSE	SScore	SScore$^{0.95}$	Rank
FCM	45	BF2	Kernel	10.62	94.89	92.77	1.59	2.77	0.3220	0.3432	1
FCM	30	BF2PG	Kernel	10.91	94.93	93.26	1.65	2.86	0.3285	0.3452	2
FCM	50	BF2PG	Kernel	10.67	94.78	92.49	1.79	2.81	0.3231	0.3459	3
FCM	80	BF2PG	Kernel	10.25	94.58	91.53	1.74	2.71	0.3150	0.3460	4
K-means	50	BF2	Kernel	10.78	94.96	92.74	1.87	2.80	0.3254	0.3485	13
K-means	45	BF2	Kernel	10.86	94.89	92.78	1.87	2.83	0.3273	0.3492	15
K-means	40	BF2	Kernel	10.89	94.82	92.85	1.84	2.83	0.3303	0.3499	16
K-means	50	BF2PG	Kernel	10.94	94.87	92.60	2.20	2.87	0.3281	0.3506	18
Base-Month	12	Month	Kernel	12.21	95.12	94.10	1.91	3.12	0.3601	0.3704	943
Base-Clim.	1	-	Normal	12.17	94.78	94.49	0.00	3.11	0.3740	0.3774	1492

In Figure 1.a, a sample forecast error distribution is shown and the corresponding fitted kernel density distribution is also plotted. In the first stage of experiments the K-means algorithm was run with the different feature sets and fitting methods. Figures 1.b and 2.b show the box plots of the SScore$^{0.95}$ measure for these alternative choices. As can be seen, the Kernel fitting method and the BF2 feature sets can obtain PIs with higher skill. It must be noted here that when the measured *SScore* (and not its confidence bound) is used for evaluations, very large number of clusters (e.g. 200) would always achieve the best results. However, this is due to the fact that with such large values of *K* there would be very few test cases available to have a reliable measurement of the $\bar{\Delta}_M^{\alpha,j}$ statistic in individual clusters.

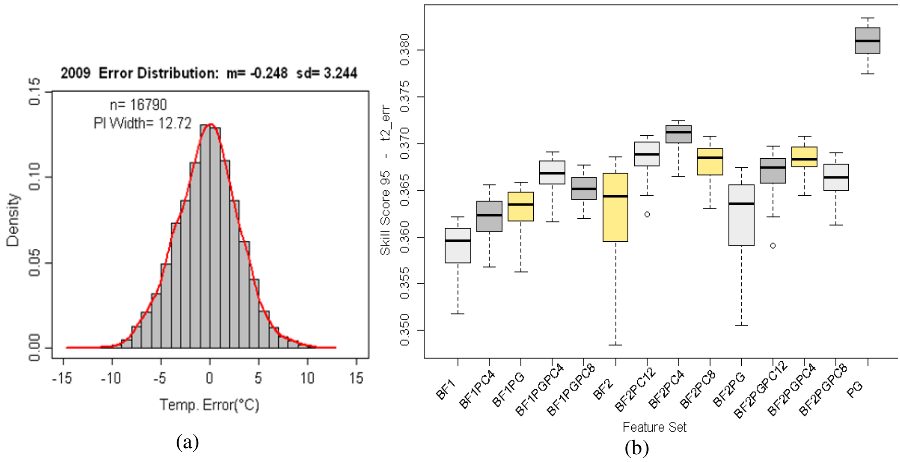

Fig. 1. (a) Forecast error distribution in 2009 and kernel fitted distribution (b) $SScore^{0.95}$ of the fourteen different feature sets using K-means in 3-fold (yearly) cross validations

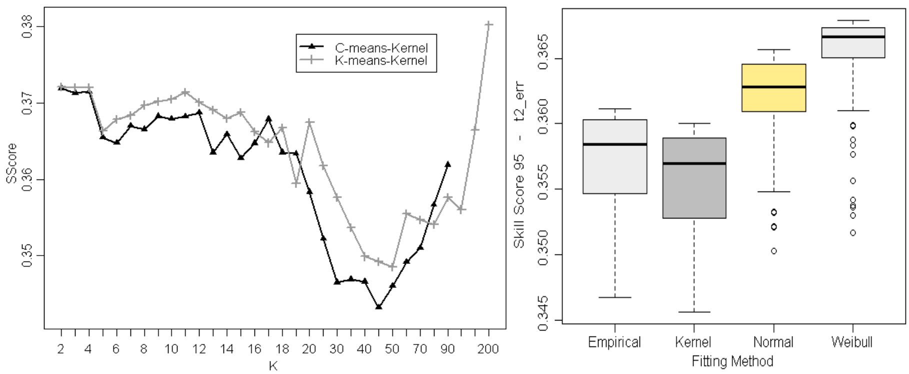

Fig. 2. (a) The trend of $SScore^{0.95}$ with increasing number of clusters (b) skill score of the four different fitting methods

Figure 2.a shows the trend of PI forecast skills for the best setups of K-means and FCM as the number of clusters is increasing. The curves also show the better performance of the Fuzzy C-means algorithm around *K=45*.

6 Conclusions

A new method is presented that can model forecast uncertainty from the historical performance of the NWP system and provide prediction intervals for new point forecasts. This is achieved using fuzzy clustering and density fitting methods over the prediction error records. The performance of this method was investigated through an experimental study employing an accurate evaluation framework. The availability of forecast uncertainty in the obtained PIs and their demonstrated higher skill compared to baseline methods suggests the effectiveness of this method. Due to the temporal nature of the weather attribute forecasts and their associated errors, application of time series analysis techniques in the PI forecasting methods can potentially improve the skill of the predictions in future work.

References

[1] Bezdek, J.C., Ehrlich, R., Full, W.: FCM: The fuzzy c-means clustering algorithm. Computers and Geosciences 10(2-3), 191–203 (1984)
[2] Chatfield, C.: Calculating Interval Forecasts. Journal of Business and Economics Statistics 11, 121–135 (1993)
[3] Corchado, E., Arroyo, A., Tricio, V.: Soft computing models to identify typical meteorological days. Logic Journal of the IGPL 19(2), 373–383 (2010)
[4] Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction and estimation. Journal of the American Statistical Association 102, 359–378 (2007)
[5] Hahn, G.J., Meeker, W.Q.: Statistical Intervals: A Guide for Practitioners. John Wiley, New York (1991)
[6] Hosek, J., Musilek, P., Lozowski, E., Pytlak, P.: Effect of time resolution of meteorological inputs on dynamic thermal rating calculations. IET GTD 5(9), 941–947 (2011)
[7] Jolliffe, I.T., Stephenson, D.B. (eds.): Forecast Verification: A Practitioner's Guide in Atmospheric Science. Wiley, Chichester (2003)
[8] Lange, M.: On the uncertainty of wind power predictions-Analysis of the fore-cast accuracy and statistical distribution of errors. Journal of Solar Energy Engineering 127, 177–184 (2005)
[9] Lange, M., Heinemann, D.: Relating the Uncertainty of Short-term Wind Speed Predictions to Meteorological Situations with Methods from Synoptic Climatology. In: Proceedings of the European Wind Energy Conference, EWEC, Madrid (2003)
[10] Khaki, M., Musilek, P., Heckenbergerova, J., Koval, D.: Electric Power System Cost/Loss Optimization Using Dynamic Thermal Rating and Linear Programming. In: EPEC 2010 (2010)
[11] Nielsen, H., Madsen, H., Nielsen, T.S.: Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. Wind Energy 9, 95–108 (2006)
[12] Orrell, D., Smith, L., Barkmeijer, J., Palmer, T.: Model Error in Weather Forecasting. Nonlinear Proc. Geophys. 8, 357–371 (2001)
[13] Palmer, T.N.: Predicting Uncertainty in Forecasts of Weather and Climate. Reports on Progress in Physics 63, 71–116 (2000)
[14] Pedrycz, W.: Knowledge-Based Clustering: From Data to Information Granules. Wiley, Hoboken (2005)

[15] Pinson, P.: Estimation of the uncertainty in wind power forecasting. PhD Dissertation, Ecole des Mines de Paris (2006)
[16] Pinson, P., Juban, J., Kariniotakis, G.: On the quality and value of probabilistic forecasts of wind generation. In: Proceedings of the PMAPS Conference, IEEE Conference on Probabilistic Methods Applied to Power Systems, Stockholm, Sweden (2006)
[17] Pinson, P., Kariniotakis, G.: Conditional prediction intervals of wind power generation. IEEE Transactions on Power Systems 25(4), 1845–1856 (2010)
[18] Pinson, P., Nielsen, H.A., Møller, J.K., Madsen, H., Kariniotakis, G.N.: Nonparametric probabilistic forecasts of wind power: required properties and evaluation. Wind Energy 10, 497–516 (2007)
[19] Vejmelka, M., Musilek, P., Palus, M., Pelikan, E.: K-means clustering for problems with periodic attributes. International Journal of Pattern Recognition and Artificial Intelligence 23(4), 721–743 (2009)
[20] Xu, R., Wunsch II, D.: Survey of clustering algorithms. IEEE Transactions on Neural Networks 16(3), 645–678 (2005)
[21] Wilks, D.S.: Statistical Methods in the Atmospheric Sciences, 2nd edn. Academic Press, New York (2006)
[22] Wonnacott, T.H., Wonnacott, R.J.: Introductory Statistics. Wiley, New York (1990)