Downloaded from orbit.dtu.dk on: Nov 07, 2026

DTU Library

Discussion of “Prediction intervals for short-term wind farm generation forecasts” and “Combined nonparametric prediction intervals for wind power generation”

Pinson, Pierre; Tastu, Julija

Published in:
IEEE Transactions on Sustainable Energy

Link to article, DOI:
10.1109/TSTE.2014.2323851

Publication date:
2014

Link back to DTU Orbit

Citation (APA):
Pinson, P., & Tastu, J. (2014). Discussion of “Prediction intervals for short-term wind farm generation forecasts” and “Combined nonparametric prediction intervals for wind power generation”. *IEEE Transactions on Sustainable Energy*, *5*(3), 1019-1020. https://doi.org/10.1109/TSTE.2014.2323851

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Discussion of "Prediction intervals for short-term wind farm generation forecasts" and "Combined nonparametric prediction intervals for wind power generation"

Pierre Pinson, *Senior Member, IEEE*, Julija Tastu

IN A SERIES of recent work published in the *IEEE Transactions on Neural Networks and Learning Systems*, the *IEEE Transactions on Power Systems*, *Electric Power Systems Research* and here in the *IEEE Transactions on Sustainable Energy* (among others), Khosravi and co-authors propose and utilize a new score for the evaluation of interval forecasts, the so-called Coverage Width-based Criterion (CWC). This score has been used for the tuning (in-sample) and genuine evaluation (out-of-sample) of prediction intervals for various applications, e.g. electric load [1], electricity prices [2], general purpose prediction [3] and wind power generation [4], [5]. Indeed, two papers by the same authors appearing in the *IEEE Transactions on Sustainable Energy* employ that score, and use it to conclude on the comparative quality of alternative approaches to interval forecasting of wind power generation.

Probabilistic forecasting is to become a core aspect in modern power systems engineering, with increased penetration of renewable energy sources, and with their inherent variability and lack of predictability, e.g., wind and solar energy. Besides, load patterns are becoming more variable and less predictable due to changes in consumption patterns with the apparition of proactive prosumers. It will overall result in more uncertainty in market-clearing outcomes such as energy volumes and prices. Probabilistic forecasts in the form of quantiles, intervals, predictive densities or more generally trajectories, are optimal inputs to a wide range of decision-making problems defined in a stochastic or robust optimization framework. These forecasts are more difficult to evaluate than the more common single-valued (/deterministic) predictions, owing to their very nature. Scoring rules to be used for the evaluation of probabilistic forecasts are required to be proper [6]–[8]: propriety is the basic property of a score to insure that perfect forecasts should be given the best score value, say, the lowest one if the score is negatively oriented. If not the case, one could then hedge the score, by finding tricks that permit to get better score values without attempting to issue better forecasts. More generally, employing a score that is not proper makes that one can never be sure of the validity of the results from an empirical comparison or benchmarking of rival approaches. Research on the topic of proper evaluation of probabilistic forecasts, in the form of prediction intervals, can at least be traced back to the work of Winkler [9].

Pierre Pinson is with the Department of Electrical Engineering at the Technical University of Denmark, while Julija Tastu is with the Department of Mathematics and Computer Science at the same University (emails: ppin@dtu.dk, juvl@dtu.dk).

Unfortunately in the case of the aforementioned papers on interval forecasting of wind power generation (and other quantities), the CWC score employed is not proper, as will be illustrated below based on a simple example. As a consequence, it is difficult to appraise the quality of the results in these manuscripts. This is while there exists simple known scoring rules that could be readily used instead, for instance inspired by the original proposal of Winkler [9].

Let us first remind the reader about the definition of the CWC score. For a given lead time and nominal coverage rate $(1-\beta)$, it writes

$$\text{CWC} = \bar{\delta}\big(1 + \mathbf{1}\{\Delta b > 0\} \exp\left(\eta \Delta b\right)\big), \ \eta > 0, \qquad (1)$$

with $\mathbf{1}\{.\}$ an indicator function, returning 1 if the condition between brackets realizes, and to 0 otherwise. In parallel, $\Delta b = (1-\beta) - b$ is the difference between nominal $(1-\beta)$ and empirical (b) coverage rates (that is, a form of probabilistic bias), while $\bar{\delta}$ is the average width of the prediction intervals. η is a free parameter that can be set to any positive value. It is argued that based on the above definition, the CWC penalizes intervals that are not probabilistically reliable, while it rewards them for their sharpness (since sharp intervals are intuitively expected to be more informative). The CWC is negatively oriented: lower values indicate prediction intervals of higher quality.

We now introduce a simple example in order to show how the CWC is not proper and may give a better score value to intervals that should actually be deemed of lower quality. Consider a stochastic process $\{X_t, \ t = 1, \ldots, T\}$ defined as a sequence of T independent and identically distributed (i.i.d.) random variables X_t with probability density function (pdf) defined on a compact support, with

$$g(x) = 12 \left(x - \frac{1}{2}\right)^2, \quad x \in [0, 1]. \qquad (2)$$

We denote by G the cumulative distribution function (cdf) associated to g, given by

$$G(x) = 4 \left(x - \frac{1}{2}\right)^3 + \frac{1}{2}, \quad x \in [0, 1]. \qquad (3)$$

One can readily verifies that G is an increasing function, with $G(0) = 0$ and $G(1) = 1$.

For this stochastic process consisting of i.i.d. random variables, it straightforward to define the optimal interval forecasts directly based on the density in (2). For instance, for a nominal coverage rate of 0.9 (to cover observations 90% of the times), optimal central prediction intervals $\mathcal{I}_t^*$ for any time t are defined by the quantiles with nominal levels 0.05 and 0.95:

$$\mathcal{I}^* = \left[G^{-1}(0.05), G^{-1}(0.95)\right]. \quad (4)$$

And, based on the expression for G given in (3),

$$\mathcal{I}^* = [0.017, 0.983]. \quad (5)$$

These intervals are perfectly reliable by definition, and therefore the CWC value assessing their quality is equal to their average width, i.e., $\text{CWC}^* = 0.966$. Since the above prediction intervals are the perfect ones, no other intervals should be given a better score.

Now in order to hedge the score, simply consider generating prediction intervals in a binary manner, although acknowledging that the nominal coverage rate should be respected in practice. Following such a binary approach, intervals are defined as full intervals [0,1] 90% of the times, and as empty intervals (i.e., any single value in [0,1]) 10% of the times. This writes

$$\mathcal{I} = \begin{cases} [0,1], & \text{if } u_t \geq 0.1 \\ 0.5, & \text{otherwise} \end{cases}, \quad (6)$$

using 0.5 as an example value for the empty intervals, and where u_t is a realization at time t from a sequence of i.i.d. uniform random variables $U_t \sim \mathcal{U}[0,1]$. These intervals are clearly not sophisticated ones, and not informative at all. Since covering the actual observations of the process 90% of the times, by construction, their CWC score values is also given by their average width, that is, $\text{CWC} = 0.9$ (significantly lower than the value obtained for the perfect prediction intervals).

In the frame of an empirical investigation comparing the quality of alternative interval forecasting methods for the stochastic process $\{X_t,\ t = 1, \ldots, T\}$, using the CWC score would lead to the conclusion that the binary-type of intervals are better than the optimal ones. Due to the lack of propriety of the CWC score, this type of problem may appear in any type of empirical investigation, making that one can never conclude on the respective quality of the interval forecasts being evaluated.

REFERENCES

[1] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," *IEEE Transactions on Power Systems*, vol. 25, pp. 1496-1503, 2010.

[2] A. Khosravi, S. Nahavandi, and D. Creighton, "A neural-network-GARCH based method for construction of prediction intervals," *Electric Power Systems Research*, vol. 96, pp. 185–193, 2013.

[3] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of Neural Network-based prediction intervals and new advances," *IEEE Transactions on Neural Networks*, vol. 22, pp. 1341–1356, 2011.

[4] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction intervals for short-term wind farm generation forecasts," *IEEE Transactions on Sustainable Energy*, available online, 2013.

[5] A. Khosravi, and S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *IEEE Transactions on Sustainable Energy*, available online, 2013.

[6] J. Bröcker, and L. A. Smith, "Scoring probabilistic forecasts: on the importance of being proper," *Weather and Forecasting*, vol. 22, pp. 382–388, 2007.

[7] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society B*, vol. 69, pp. 243–268, 2007.

[8] T. Gneiting, and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation", *Journal of the American Statistical Association*, vol. 102, pp. 359-378, 2007.

[9] R. L. Winkler, "A decision-theoretic approach to interval estimation," *Journal of the American Statistical Association*, vol. 67, pp. 187191, 1972.