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

Citation (APA):
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.

Proper makes that one can never be sure of the validity of the results from an empirical comparison or benchmarking of actions on Neural Networks and Learning Systems. 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

\[
CWC = \delta \{ 1 + 1 \{ \Delta b > 0 \} \exp(\eta \Delta b) \}, \quad \eta > 0, \quad (1)
\]

with \(1\{\cdot\}\) an indicator function, returning 1 if the condition between brackets realizes, and 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 \(\delta\) is the average width of the prediction intervals. \(\eta\) 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

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

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

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

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 $I^*_t$ for any time $t$ are defined by the quantiles with nominal levels 0.05 and 0.95:

$$I^* = [G^{-1}(0.05), G^{-1}(0.95)].$$

(4)

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

$$I^* = [0.017, 0.983].$$

(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

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

(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 $U_t \sim U[0, 1]$. 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


