2018

Evaluation of Wind Energy Forecasts: the Undervalued Importance of Data Preparation

Gianni Goretti
Technological University Dublin, d15123939@mytudublin.ie

Aidan Duffy
Technological University Dublin, aidan.duffy@tudublin.ie

Recommended Citation

G. Goretti and A. Duffy (2018) "Evaluation of Wind Energy Forecasts: the Undervalued Importance of Data Preparation," 2018,*15th International Conference on the European Energy Market (EEM),* Lodz, 2018, pp. 1-5.

Evaluation of Wind Energy Forecasts: the Undervalued Importance of Data Preparation

Gianni Goretti
School of Civil and Structural Engineering
Dublin Institute of Technology
Dublin, Ireland
Email: gianni.goretti@dit.ie

Aidan Duffy
School of Civil and Structural Engineering
Dublin Institute of Technology
Dublin Ireland
Email: aidan.duffy@dit.ie

***Abstract*—The evaluation of wind energy forecasts is a key task for those involved in the wind power sector, and the accurate evaluation of forecasts is fundamental to make informed decisions both in business and research. To evaluate the accuracy of a forecast, observed values must be compared against forecast values over a test period. At times, however, the actual generation of a wind farm can be affected by factors that are outside the scope of the forecast model. Evaluating a forecast using a data set that includes such out-of-scope observations might give a biased or inconsistent assessment. In the data preparation phase, then, the evaluator should identify out-of-scope data and decide whether to include or remove these from the data set. In this paper, we carry out an empirical study based on data from an existing wind farm and a number of day-ahead forecasts in order to highlight the effects of including in- and out-of-scope data on forecast accuracies. The results show that the outcome of the evaluation varies significantly depending on the criteria adopted in the data selection.**

***Index Terms*—Data cleaning; data preprocessing; forecast evaluation; wind energy forecasting.**

I. Introduction

In electricity markets with high levels of wind energy penetration, wind energy forecasts are increasingly employed for unit commitment and energy trading purposes. Trading commitments are typically made one or two days in advance in these markets. Consequently, the wind power forecast horizon of interest typically ranges from six to 36 hours, this being referred to as 'short-term' wind power forecasting [1].

The evaluation of wind energy forecasts is a key task for those involved in the wind power sector. In academia and research, forecast developers need to evaluate forecasts to assess whether a new model represents an improvement from the state of the art and thus if it is worth developing. In industry, people using wind energy forecasts as part of their business (e.g. wind farm operators, transmission system operators, energy traders) need to choose the model that best suits their requirements and delivers best financial results [2]. The accurate and suitable evaluation of a forecast is therefore fundamental in order to make informed decisions both in business and research.

In the wind power sector, many actors base their evaluation of forecasts mainly - if not solely - on their accuracies [3], using statistical metrics, of which the most widely used are: Mean Error (or Bias); Mean Absolute Error (MAE); and Root Mean Square Error (RMSE). All these statistics are calculated starting from the definition of the *forecast* (or *prediction*) *error* at each time interval i:

$$e_i = y_i - \hat{y}_i$$

where y_i is the observed quantity and $\hat{y}_i$ is the predicted quantity at time interval i [4].

In more general terms, in order to evaluate the accuracy of a forecast, the observed values must be compared against forecast values over a certain period, hereinafter the *test period*. Observed values are the measurements of the actual generation of a wind farm - i.e. the amount of energy injected into the grid - over the test period. Forecast values are the predictions of electricity generation obtained from a forecast model for the wind farm over the test period.

At times, the actual generation of a wind farm can be affected by factors that are outside the scope of the forecast model. In other words, the forecast model has not been developed to predict the energy generation of the wind farm under those operating conditions and therefore forecast errors will arise which may be significant. A simple example of this may include the inclusion of turbine outage events in the energy generation data, whereas the forecast may deal only with fault-free behaviour. We will discuss this concept in more detail and provide examples in Section II.

Indeed, evaluating a forecast using a data set that includes such out-of-scope observations might give a biased or inconsistent assessment of the performance of the forecast model. Before carrying out any assessment, then, the forecast evaluator should systematically identify out-of-scope operational data and decide whether to include or remove these from the data set. We will refer to this task as *data selection*.

Data selection is performed in the preliminary phase of the forecast evaluation process, when the measurements of the actual generation are gathered and pre-processed. This preliminary phase is referred to as *data preparation* and includes three stages: data collection; data cleaning; and data

978-1-5386-1488-4/18/$31.00 ©2018 IEEE

selection. We will discuss the data preparation process in more detail in Section III.

In Section IV, we will then carry out an empirical study based on data from an existing wind farm and a number of day-ahead commercially-available forecasts in order to highlight the effects of including in- and out-of-scope data on forecast accuracies. Forecasts accuracies are evaluated over a four-month test period adopting three different data selection criteria.

In Section V, we will discuss how the adoption of different data selection criteria affects the evaluation results, and conclusions are then drawn in Section VI.

II. Motivation

Short-term wind power forecasts generally need two broad categories of input data [5]. One is the forecasts of site wind speeds and other relevant weather variables such as wind direction or atmospheric pressure. The other is information on the characteristics of the installed wind turbines (for example, power curves), the wind farm layout, historical observations of power generation and other relevant variables at the site of interest. It follows that the forecast electricity output from wind farms depends mainly on two factors:

- the forecast weather variables; and
- how the forecaster has modelled the energy conversion process.

The former model input data are obtained from Numerical Weather Prediction (NWP) models downscaled to the level of the wind farm being studied. The latter entails the modelling (explicit or implicit) of the power systems behaviour in response to a number of external and internal factors.

As already highlighted in Section I, there can be observations during the test period where the actual generation of the wind farm is affected by factors that are outside the scope of the forecast model. This may be the case when the factor affecting generation has not been anticipated by, or when its prediction goes beyond the interest of the forecaster. Examples include:

- forecasting power curtailment due to transmission constraints in the grid may not be of interest to the wind farm operator, if any depowering is financially compensated by the transmission system operator; or
- forecasting turbine shutdown due to component failure may be ignored due to the stochastic nature of such events.

The foregoing highlights the importance of precisely understanding what is being modelled by the forecaster. The scope of the wind power production system can be generally described as all the mechanical, ICT and electrical components of the wind farm, together with the weather acting on these components. However the system being modelled by the forecaster will always be more limited than this. For example, it may exclude some components (e.g. gearboxes may not be explicitly modelled), relevant weather variables (e.g. atmospheric pressure) or operating conditions (e.g. certain forced outages or blade icing).

Therefore, for clarity, we will refer to the following terms throughout the paper:

- the wind power production system including all operating components, weather variables and operating conditions will be called the *observed system*;
- the power system which is of interest to the modeller will be referred to as the *forecast system*.

III. Methodology

This section presents the three stages involved in the data preparation process undertaken prior to forecast evaluation.

The measurements of electricity generation and other relevant signals from the wind farm are gathered and aggregated to form the *raw data set*. All such measurements are obtained and collated during the *data collection* stage. Relevant signals may include wind speed, wind direction, air temperature, atmospheric pressure, as well as curtailment signals, and fault events.

The raw data set normally contains 'dirty' data; these are data which contain errors, inconsistencies, missing or incomplete information and which degrade the quality of the data set [6]. Since data of poor quality produce results of poor quality, it is necessary to pre-process raw data to guarantee their quality and integrity before undertaking any analysis. This second stage, which we will refer to as *data cleaning*, aims to identify and remove any measurements which are not reflective of the actual operation of the observed system [7]. The resulting data set is the *clean data set*.

The data cleaning process, for example, can involve the definition of rules on the validity of data; e.g., measured values of wind speed have to be non-negative, or no power output can occur if the wind speed is zero. Observations violating these rules are flagged as erroneous and removed from the data set.

The clean data set is reflective of the operation of the observed system rather than of the forecast system. As mentioned above, the scope of the observed system can differ from the scope of the forecast system. Indeed, there can be periods where the actual generation is affected by factors that were not considered in the forecast model. The forecast evaluator then, has to identify these periods and decide whether to remove or keep them in the data set in line with the objective of the study.

If, for example, the goal of the study is to evaluate the forecast strictly in line with the assumptions under which it was generated, all periods outside the forecast system's scope should be removed from the clean data set. Let us assume, for instance, that the forecast developer did not include grid curtailment in the model; then all the time intervals where the electricity generation is affected by grid curtailment should be removed from the data set. If, on the contrary, the goal is to

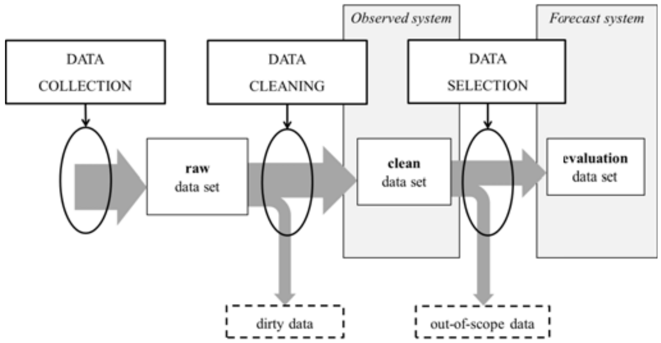

Fig. 1. The data preparation process.

evaluate the forecast in absolute terms, then all periods in the clean data set should be kept.

The data set resulting from this third stage, which we will refer to as *data selection*, is the *evaluation data set*.

Fig. 1 summarizes the data preparation process, showing the three stages and the data set resulting from each of them.

The statistical analyses for the evaluation of the forecast accuracy are eventually carried out on the evaluation data set.

IV. Empyrical Analysis

In this section, we carry out an empirical study based on data from an existing wind farm and a number of day-ahead forecasts. Data preparation is performed according to the methodology outlined in Section III. The accuracy of the forecasts is then evaluated over a four-month test period adopting three different data selection criteria.

A. Data Collection

Turbine-level measurements of electricity generation and wind speed are obtained from the Supervisory Control And Data Acquisition (SCADA) systems installed at the wind farm. These measurements are then aggregated at wind-farm level and averaged over 30-minute.

Information on grid curtailment at the wind farm is obtained from the curtailment logs; these are records of the curtailment signals from the transmission system operator, which include start and end times of power curtailment and the amount of energy being curtailed.

Information on fault events is obtained from the error logs for each turbine installed at the wind farm. These are event-based records, which include start and end times of the fault event, and a basic description of the type of fault occurred.

B. Data Cleaning

Once all the measurements of interest have been collected to form the raw data set, data cleaning is performed. The aim is to obtain a clean data set that is reflective of the actual operation of the observed system. The anomalies in the data are detected and treated following an iterative three-stage process consisting of: screening, diagnosis, and treatment [8].

The incidence of data removed from the raw data set due to data cleaning is equal to 0.26%, and the clean data set consists of 5841 time intervals.

C. Data Selection

Using the information gathered during the data collection phase, two factors affecting the operation of the wind farm are identified for isolation: power curtailment and fault events. Three separate evaluation scenarios are then defined, each one involving different data selection criteria and resulting in a different evaluation data set.

In the first scenario, it is assumed that the scope of the forecasts includes all the factors affecting the operation of the wind farm. In other words, the physical and operational scope of the forecast system exactly coincides with that of the observed system, so that the clean data set can be used for forecast evaluation purposes. In the results, this is referred to as the "*Clean*" scenario.

In the second scenario, it is assumed that curtailment falls outside the scope of the forecast system. In other words, the forecasts were not developed to predict the performance of the wind farm in the case of curtailment. As a consequence, periods affected by curtailment are removed from the clean data set. In the results this is referred to as the "*No Curtailment*" scenario.

The number of periods affected by curtailment is equal to the 20.3% of the clean data set, with the "*No Curtailment*" evaluation data set consisting of 4658 time intervals.

In the third scenario, it is assumed that fault events and curtailment are both factors outside the scope of the forecasts. In other words, the forecasts were neither designed to predict the performance of the wind farm in the case of fault events nor curtailment. As a consequence, the corresponding periods are removed from the clean data set. In the results this is referred to as the "*No Faults or Curtailment*" scenario.

The number of periods affected by fault events and curtailment is equal to the 44.7% of the clean data set, with the "*No Faults or Curtailment*" evaluation data set consisting of 3232 time intervals.

Fig. 2 shows a power vs. wind speed plot for the cleaned data set, categorized into each three scenarios described above.

D. Results

We consider five different day-ahead deterministic forecasts for a wind farm in Northern Europe; all five are commercially available forecasts that represent industry best practice. We evaluate these forecasts over a four-month test period for the three scenarios described in Section IV-C.

All summary statistics are normalized by the wind farm nominal power P_{nom} and are expressed as a percentage.

The box-and-whiskers plots in Fig. 3 show the distribution of forecast errors for each forecast in the three evaluation scenarios.

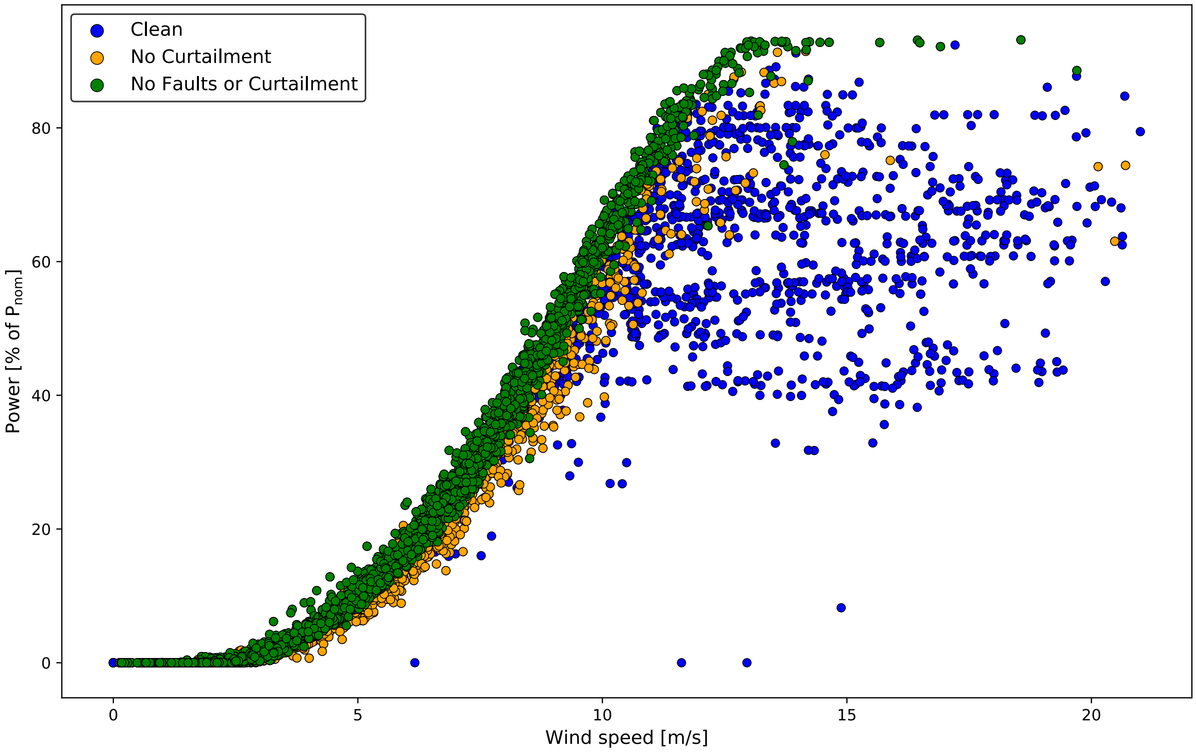

Fig. 2. Evaluation data sets for the three evaluation scenarios: Clean (blue, green and orange), No Curtailment (green and orange), No Faults or Curtailment (green).

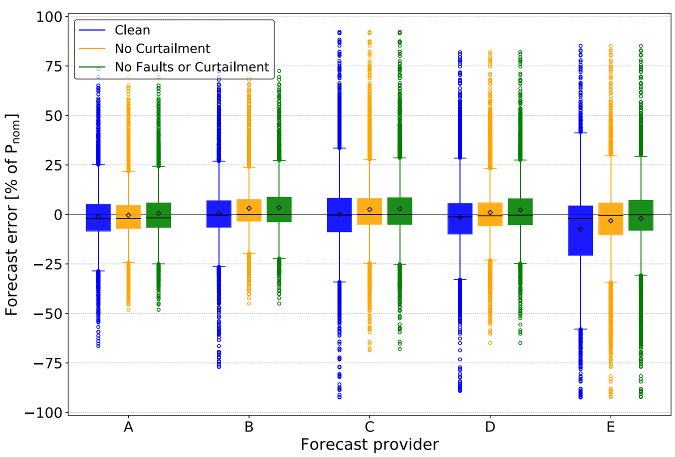

Fig. 3. Box-and-whiskers plots of forecast errors for the three evaluation scenarios; the solid line identifies the median of the distribution, while the diamond marker the mean value (i.e., the forecast bias).

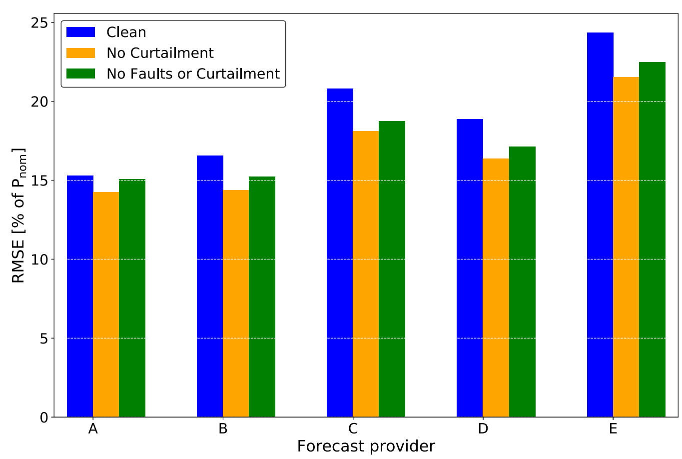

Fig. 4. Normalised RMSE for the three evaluation scenarios.

The RMSE of the forecasts was then calculated according to [4] for each evaluation scenario. The results are shown in Fig. 4.

V. Analysis and Discussion

It is evident from Fig. 2, 3 and 4 that data selection has a significant impact on the evaluation of forecast error. Fig. 2 clearly identifies power output observations under conditions of curtailment (blue). It is evident that these results have the effect of significantly reducing average power output and increasing forecast error unless properly accounted for in either the forecast model or the data selection process. It can also be seen in this figure that turbine faults (orange) result in observations which are lower than would be expected based on the wind farm's power curve and, therefore, increase forecast error unless properly accounted for in the model or data.

It can be seen in Fig. 3 that for all forecasts the mean error (diamond marker in the plots) increases consistently in value when moving from the "*Clean*" evaluation data set to the "*No Curtailment*" and further on to the "*No Faults or Curtailment*" one.

For the RMSE (Fig. 4), it is observed that the performance of all forecasts improves when moving from the "*Clean*" to the "*No Curtailment*" evaluation scenario, since the RMSE decreases in value. However, when moving from the "*No Curtailment*" to the "*No Faults or Curtailment*" evaluation scenarios, performances degrades slightly, with the RMSE increasing in value in all cases (but still remaining lower than for the "*Clean*" scenario).

These results seem to indicate that curtailment is a factor outside the scope of the forecast systems, while fault events are within the scope.

VI. Conclusion

The results of the empirical study in Section IV show how the outcome of forecast evaluation can vary depending on the evaluation data set used in the analysis, which in turn depends on the data selection criteria adopted.

To guarantee the integrity and consistency of their analysis, forecast evaluators should clearly identify the scope of the forecast system and include the data selection stage in the data preparation process.

Moreover, any work involving the evaluation of a forecast should clearly state the characteristics of the evaluation data set and the data selection criteria adopted. Indeed, one could claim a certain level of accuracy by evaluating the forecast on the data set that best suits the model.

Acknowledgment

The authors gratefully acknowledge financial support provided by DIT Fiosraigh. The authors would also like to thank Des Farren and Ciaran Gilbert, as well as the two anonymous reviewers for their valuable comments.

References

[1] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, The State-Of-The-Art in Short-Term Prediction of Wind Power A Literature Overview, *Tech. Report, ANEMOS.plus*, pp. 1109, 2011.

[2] R. J. Bessa et al., Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry, *Energies*, vol. 10, no. 9, pp. 148, 2017.

[3] C. Möhrlen, R. J. Bessa, M. Barthod, G. Goretti, and M. Siefert, Use of Forecast Uncertainties in the Power Sector: State-of-the-Art of Business Practices, *15th Wind Integr. Work.*, 2016.

[4] H. Madsen, P. Pinson, and G. Kariniotakis, Standardizing the performance evaluation of short-term wind power prediction models, *Wind Eng.*, vol. 29, no. 6, pp. 475489, 2005.

[5] J. F. Manwell, J. G. McGowan, and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application*. 2010.

[6] R. Y. Pipino, Leo L.; Lee, Yang W.; and Wang, Data quality assessment, *Commun. ACM*, vol. 45, no. 4, pp. 211218, 2002.

[7] E. Rahm and H. H. Do, Data Cleaning: Problems and Current Approaches, *IEEE Data Eng. Bull.*, vol. 23, no. 4, pp. 313, 2000.

[8] J. Van Den Broeck, S. A. Cunningham, R. Eeckels, and K. Herbst, Data cleaning: Detecting, diagnosing, and editing data abnormalities, *PLoS Med.*, vol. 2, no. 10, pp. 09660970, 2005.