Renewable and Sustainable Energy Reviews 96 (2018) 352–379

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

A review of uncertainty representations and metaverification of uncertainty assessment techniques for renewable energies

André Gensler[a,*], Bernhard Sick[a], Stephan Vogt[b]

[a] Intelligent Embedded Systems, University of Kassel, Wilhelmshöher Allee 71 - 73, 34121 Kassel, Germany
[b] Fraunhofer Institute for Energy Economics and Energy System Technology (IEE), Königstor 59, 34119 Kassel, Germany

ARTICLE INFO

Keywords:
Power forecasting
Probabilistic forecasting
Performance assessment
Forecast verification
Scoring rule decomposition
Ensemble methods

ABSTRACT

The performance evaluation of forecasting algorithms is an essential requirement for quality assessment and model comparison. In recent years, algorithms that issue predictive distributions rather than point forecasts have evolved, as they better represent the stochastic nature of the underlying numerical weather prediction and power conversion processes. Standard error measures used for the evaluation of point forecasts are not sufficient for the evaluation of probabilistic forecasts. In comparison to deterministic error measures, many probabilistic scoring rules lack intuition as they have to satisfy a number of requirements such as reliability and sharpness, whereas deterministic forecasts only need to be close to the actual observations. This article aims to empower practitioners and users of probabilistic forecasts to be able to choose appropriate uncertainty representations and scoring rules depending on the desired application and available data. A holistic view of the most popular forms of uncertainty representation from single forecasts and ensembles is given, followed by a presentation of the most popular scoring rules. We want to broaden the understanding for the working principles and relationship of different scoring rules and their decomposition for probabilistic forecasts of continuous variables by showing their differences. Therefore, we analyze the behavior of scoring rules, a process frequently referred to as metaverification, in detail on real-world multi-model ensemble forecasts in a number of case studies.

1. Introduction

Algorithms for power forecasting on short and mid-term horizons (e.g., intraday and day-ahead forecasts) are in nearly all cases based on numerical weather predictions (NWP). Weather forecasting is a stochastic process, meaning that though the current weather condition can be measured up to a certain degree, a future weather situation cannot be exactly predicted due to the chaotic behavior of the fluid-dynamics of the atmosphere. This uncertainty in the weather forecasting process affects the power forecasting process. Furthermore, the uncertainty is in many cases amplified, e.g., due to the nonlinearity of the wind turbine power curve. Therefore, while the quality of deterministic point forecasts is still improved (e.g., through model combination), the performance converges towards the intrinsic uncertainty of the underlying NWP generating processes. To overcome this problem, in recent years there has been a shift from the paradigm of creating point forecasts to creating distributional (or probabilistic) forecasts [1]. Probabilistic forecasting quantifies the amount and direction of the uncertainty of a prediction in a situation-adaptive way and it can be used to retain optimal decision-making performance under uncertain conditions. This is of particular interest in power forecasting for applications such as grid stability and power market operation as basis for better decision making. For instance, the amount of reserve capacity can be planned depending on the amount of (un-)certainty of a forecast [2,3]. Probabilistic forecasts can also be used in economic reward and cost functions [4,5]. While probabilistic forecasting for binary events has been widely applied, e.g., for weather event forecasting (such as the probability of the event of rainfall), there is an increasing demand for probabilistic forecasting of ordinal and continuous quantities in the past decade.

While the cause of the indisputable complication of the power grid operation are the intermittent power generation characteristics of RE power plants, the effects also influence other areas of RE, such as demand forecasting or electricity price forecasting. Though we mostly take (wind) power forecasting as an example in this article, the presented representations and scoring rules are also directly transferable to other applications within the RE area and beyond.

* Corresponding author.
E-mail addresses: gensler@uni-kassel.de (A. Gensler), bsick@uni-kassel.de (B. Sick), stephan.vogt@iwes.fraunhofer.de (S. Vogt).

https://doi.org/10.1016/j.rser.2018.07.042
Received 22 February 2018; Received in revised form 5 July 2018; Accepted 26 July 2018

Nomenclature

$\mathbf{x}$	Predictor D-dimensional), e.g., NWP forecast.
$\hat{y}$	Deterministic point forecast of power generation.
y	Predictand variable, e.g., power generation.
$\hat{p}(y)$	Predictive distribution (pdf) for power generation y.
o	Observed "true" power measurment.
$\hat{P}(y)$	Predictive distribution (cdf) for power generation y.
o_{inst}	Installed nominal capacity of the power plant.
τ	Quantile of predictive distribution.
N	No. of evaluated items with idx $n = 1, \ldots, N$.
$\hat{y}^{(\tau)}$	Point forecast for quantile τ with $\hat{y}^{(\tau)} = \hat{P}^{-1}(\tau)$.
J	No. of members in an ensemble with $j = 1, \ldots, J$.
L	No. of quantiles in distribution with idx $l = 1, \ldots, L$.

1.1. Probabilistic forecasting and evaluation

A variety of possibilities to create probabilistic forecasts of continuous quantities have emerged in recent years. These forecasting systems vary in form (e.g., continuously differentiable or stepwise constant probability distributions, intervals, or risk-indices) and the way they are computed (parametric or non-parametric uncertainty distributions, from single forecasts or ensembles). Depending on the form of uncertainty representation, different methods of performance assessment have emerged, which are in some cases specialized for a particular form of uncertainty representation. While these error scores may be optimal for a particular form of representation, they hinder the comparability between approaches with *different* representations.

Error scores for the performance assessment of probabilistic forecasts are frequently referred to as *scoring rules*. In comparison to deterministic error measures, many probabilistic scoring rules lack intuition. Where deterministic forecasts need to be close to observations, probabilistic forecasts have to correctly assess the conditional width of the probability distribution (commonly referred to as reliability) *and* ideally concentrate the probability mass close to the observations (they have to be sharp) depending on the amount of uncertainty [1] in the process. Intuition rises by understanding how different types of errors affect scoring rules. In order to compare the performance of probabilistic forecasting techniques, there has to be a clear definition of how the process of quality assessment is performed. The most general approach to evaluate probabilistic forecasts is using scoring rules which compare a predictive distribution to an actual observation.

In the following, we will give a short overview of deterministic and probabilistic forecast evaluation surveys. Point forecasting and the error assessment of deterministic forecasts are the basis for the creation and assessment of probabilistic forecasts. Some articles describe and summarize the assessment of forecasting errors of deterministic forecasts domain-independently, e.g., [6–9]. Other power forecasting surveys also include sections on deterministic forecasting error scores [10–13], however, they are partially inconsistent with each other and only mention a selection of error scores. In [14], the evaluation of deterministic forecasting is highlighted from a decision-theoretic perspective. A summary and a comparison of error scores is presented in [15]. In addition to deterministic errors, the uncertainty assessment of a forecast is an increasingly important aspect in power forecasting. Sections on probabilistic error scores for wind power forecasting are included in [16,17,1].

In industrial practice, uncertainty predictions have yet rarely been utilized. The work of [18] gives meaningful insights into the workings of probabilistic forecasts, however, there often is a lack of a deeper understanding of the information content of uncertainty forecasts for practical decision making. To meet this challenge, this work tries to define a uniform terminology and issue a call for standardization. It gives a deeper background to the uncertainty estimation in weather models and how this can be translated into an uncertainty of wind power. In addition, possible difficulties in the area of decision making are pointed out, and which errors can result from this.

1.2. Contributions and structure of this article

The main contributions of this article are a structured overview of existing forms of estimation and representation of forecasting uncertainty as well as an investigation of uncertainty assessment techniques for probabilistic forecasts including an investigation of the decomposition properties.

We give an overview of the most relevant techniques to estimate uncertainty and highlight the most common forms of uncertainty representation. Therein, we present a holistic view of the problem of probabilistic forecasting to enable a better comparability between different forms of uncertainty representations by converting them to density functions, which happen to be the most general form of uncertainty representation typically used for continuous probabilistic forecasts. Having a common form of representation, the assessment of their performance is easier.

In a number of case studies, the characteristics of the presented scoring rules are analyzed in detail. This process is defined as *meta-verification* in [19], which describes the evaluation of performance measures and lays out desirable properties of scoring rules such as the characteristic of being *proper* [20] and the robustness to *hedging* [21], both of which are further detailed in Section 6. From the insights of the case studies, advantages and limitations in the application of each error score are discussed.

The remainder of this article is structured as follows: Section 2 introduces general desired properties of uncertainty representation techniques, possible prediction spaces, and a form of representation that is suited to represent all common forms of uncertainty representation for probabilistic forecasts. Sections 3 and 4 give an overview on algorithms for creating uncertainty representations from single predictive models (a single NWP) and ensembles (multiple NWP), respectively. A rating of the strengths and weaknesses of the presented models is given in Section 5. Section 6 highlights ways to assess the quality of probabilistic forecasts using scoring rules for probabilistic forecasts. In Section 7, the properties of the presented uncertainty assessment techniques are investigated in a number of experiments. Our insights of the experiments are discussed in Section 8, practical examples of the use of probabilistic forecasts and scoring functions are given in Section 9. Finally, this article is summarized in Section 10.

2. Uncertainty representation techniques

While the area of deterministic forecasting aims at predicting a single value for each look-ahead time (point estimate), the area of probabilistic forecasting tries to additionally assess the uncertainty of a prediction. There are a number of techniques for computing, representing, and assessing uncertainty. Though uncertainty does not necessarily have to be represented as a probability [22], the representation of uncertainty in the form of a probability does have a number of advantages, which are described, e.g., in [23,24].

2.1. Representations of predictive distributions

The most universal representation of a continuous predictive distribution is in the form of a probability density function (pdf) $\hat{p}(y)$ which can be evaluated at an arbitrary value y (which typically is a power value in power forecasting applications) to get the probability density for this value. The corresponding cumulative density function (cdf) $\hat{P}(y)$ is computed in the form