Renewable Energy 101 (2017) 526–536

Contents lists available at ScienceDirect

Renewable Energy

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

Short-term probabilistic forecasts for Direct Normal Irradiance

Yinghao Chu, Carlos F.M. Coimbra*

Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering Center of Excellence in Renewable Resource Integration and Center for Energy Research University of California, 9500 Gilman Drive, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 24 June 2016
Received in revised form
3 September 2016
Accepted 8 September 2016

Keywords:
Solar forecasting
k Nearest neighbor
Probabilistic forecast
Direct Normal Irradiance
Ensemble predictions

ABSTRACT

A k-nearest neighbor (kNN) ensemble model has been developed to generate Probability Density Function (PDF) forecasts for intra-hour Direct Normal Irradiance (DNI). This probabilistic forecasting model, which uses diffuse irradiance measurements and cloud cover information as exogenous feature inputs, adaptively provides arbitrary PDF forecasts for different weather conditions. The proposed models have been quantitatively evaluated using data from different locations characterized by different climates (continental, coastal, and island). The performance of the forecasts is quantified using metrics such as Prediction Interval Coverage Probability (PICP), Prediction Interval Normalized Averaged Width (PINAW), Brier Skill Score (BSS), and the Continuous Ranked Probability Score (CRPS), and other standard error metrics. A persistence ensemble probabilistic forecasting model and a Gaussian probabilistic forecasting model are employed to benchmark the performance of the proposed kNN ensemble model. The results show that the proposed model significantly outperform both reference models in terms of all evaluation metrics for all locations when the forecast horizon is greater than 5-min. In addition, the proposed model shows superior performance in predicting DNI ramps.

1. Introduction

Global market penetration of centralized solar productions, particularly the Concentrated Solar Power (CSP) plants, has been growing rapidly due to the increasing demands for clean and carbon-free energy [1,2]. Direct Normal Irradiance (DNI), which is the sole energy source for CSP generations, is sensitive to the circumsolar cloud cover and therefore is highly variable at the ground level [3]. As a result, the variability of ground-level CSP productions imposes serious challenges to electrical transmission grids, which need to be balanced in real time but have limited storage capacity [4,5]. Quantitatively forecast of DNI provides important information for inverter control, plant management, unit commitment, and real-time dispatch operations [6,7]. Therefore, solar forecasting models are widely recognized as key components of a smart grid to mitigate the instabilities of centralized solar power generation [4,8–10].

Many effective solar forecasting models have been developed for different temporal horizons based on data-driven, physical, or hybrid methods [8,11–25]. Most of these available solar forecasting models generate deterministic point predictions without quantified uncertainty [26,27]. Point predictions are associated with inherent and irreducible forecasting errors because of the chaotic atmospheric processes, regardless of the mechanism of the model or the methods of data processing [28–30]:

$$I(t) = f(t) + \varepsilon(t), \tag{1}$$

where $I(t)$ represents the measured value at time t, $f(t)$ represents the optimal prediction, and $\varepsilon(t)$ represents the white noise. Therefore, probabilistic forecasts, which provide the Probability Density Function (PDF) of forecast variables, are recommended for real-world forecasting applications in the literature [26,28,29,31–33].

Probabilistic solar/solar power forecasts have been proposed in literature based on analog ensemble of Numerical Weather Prediction (NWP) models [34–38]. The analog ensemble is usually defined as a set of historical instances from a NWP model for a given location and forecast horizon. These historical instances have similar features as the current instance from the same NWP model. The actual observations of the historical instances are used to estimate the PDF of the future state for various weather conditions. For hourly forecast, these proposed analog ensemble models have shown superior performance over reference models, such as the

* Corresponding author.
E-mail address: ccoimbra@ucsd.edu (C.F.M. Coimbra).

http://dx.doi.org/10.1016/j.renene.2016.09.012
0960-1481/

Nomenclature

ANN	artificial neural network
BS	Brier score
BSS	Brier skill score
CDF	Cumulative density function
CRPS	Continuous ranked probability score
DIF	Diffuse irradiance
DNI	Direct Normal Irradiance
kNN	k-nearest neighbor
kNNEn	kNN ensemble model
kNNGD	kNN Gaussian model
MAE	Mean absolute error
MBE	Mean bias error
MRE	Missing rate error
NRBR	Normalized red to blue ratio
NWP	Numerical Weather Prediction
PDF	Probability density function
PeEn	Persistence ensemble model
PI	Prediction interval
PICP	Prediction interval Coverage probability
PINAW	Prediction interval normalized averaged width
RMSE	Root mean square error
B	Beam/Direct irradiance
clr	Clear-sky condition
FH	Forecast horizon
I	Irradiance
k	Clear-sky index
M	Number of ranks
N	Number of instances
P	Probability
p	Persistence
s	Forecast skill
t	Time instance
V	Irradiance variability

Persistence Ensemble (PeEn) model [37] or the quantile regression model [38], when validated using months of historical data collected from multiple locations, particularly during hours of low solar elevation [35–38]. However, the temporal and spatial resolution of NWP-based method are not appropriate for the intra-hour forecasts of DNI [3,4].

A few intra-hour forecasting models that provide Prediction Intervals (PIs) for DNI are available in the literature [24,33,39,40]. Nevertheless, these available models either provide empirical PIs without underlying PDFs [33,41] or construct PIs based on the assumption that forecast errors are Gaussian-distributed [26,28,29]. The PDF of DNI forecast errors may not follow a Gaussian and other common distributions. For example, Gaussian, Logistic, and Kernel functions are used to fit the distributions of the persistence DNI forecast errors in Fig. 1. The bandwidth of the distribution fittings are selected using the exhaustive method [33]. The persistence errors are obtained by evaluating a persistence model (discussed in Section 3.4) using the training data collected in Folsom and Oahu when solar elevation angle is greater than 10°. More details of the data will be explained in Section 2. In addition to visual inspections, the goodness of fit [42] for each PDF is assessed using the Kolmogorov-Smirnov test [43]. However, all of the applied PDFs are rejected using the 5% confidence level. In addition, time series of DNI usually have different behaviors under different weather conditions [3,22]. For example, the DNI variability is much higher under partially cloudy skies than under a clear skies [8]. Ideally, probabilistic forecasts for DNI should be adaptive to different weather conditions [26].

Therefore, in this work, a probabilistic forecasting model is developed based on the k-nearest neighbor (kNN) ensemble predictions to generate arbitrary PDFs of DNI for intra-hour forecast horizons: 5-, 10-, 15-, and 20-min. kNN searches and identifies *k* historical time instances whose weather features are closed to the weather features of current time instance [17,44]. With the identified historical time instances and corresponding DNI behaviors, the kNN generates unique PDF forecasts for different weather conditions. The proposed model is developed and evaluated using high-quality data collected in locations with different climates. The quantitative evaluation of the proposed model is performed based on the Prediction interval coverage probability (PICP), the Prediction interval normalized averaged width (PINAW), the Brier Skill Score (BSS), the Continuous ranked probability score (CRPS), and statistic consistency. A Persistence Ensemble (PeEn) model is employed as a reference model. Gaussian PDF forecasts are also computed using the same kNN ensemble predictions to assess the advantages of using arbitrary and adaptive PDF forecasts. Details of

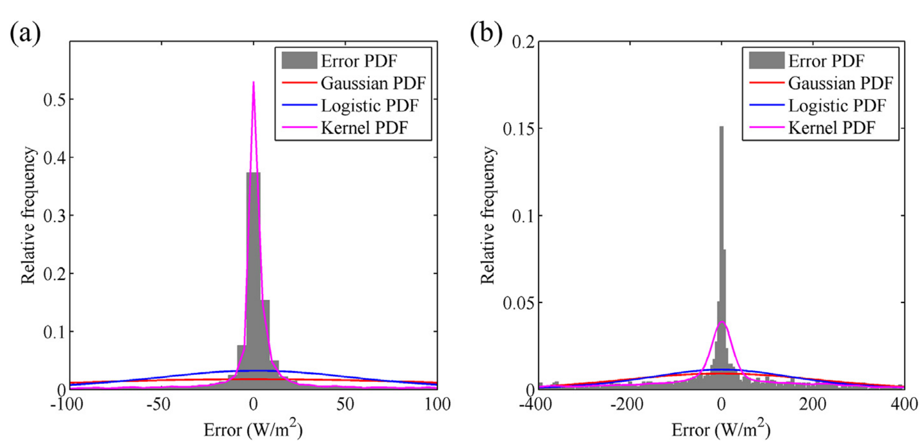

Fig. 1. Probability density functions generated based on the persistence forecast errors. The forecast errors are obtained by assessing the persistence model on the (a) Folsom and (b) Oahu training sets.