Applied Energy 176 (2016) 80–91

Contents lists available at ScienceDirect

Applied Energy

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

Generation and evaluation of space–time trajectories of photovoltaic power

Faranak Golestaneh [a,*,1], Hoay Beng Gooi [a], Pierre Pinson [b]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[b] Department of Electrical Engineering at the Technical University of Denmark, Denmark

HIGHLIGHTS

- Spatio-temporal dependency in PV generation is modelled and analysed.
- Joint predictive distributions based on marginal densities are modelled.
- New scoring rules for PV generation space–time trajectories are introduced.
- Discrimination ability of relevant scoring rules for PV trajectories is evaluated.
- The value of space–time trajectories over probabilistic forecasts is investigated.

ARTICLE INFO

Article history:
Received 20 October 2015
Received in revised form 29 April 2016
Accepted 2 May 2016

Keywords:
Stochastic dependence
Multivariate distribution
Photovoltaic energy
Space–time correlation

ABSTRACT

In the probabilistic energy forecasting literature, emphasis is mainly placed on deriving marginal predictive densities for which each random variable is dealt with individually. Such marginals description is sufficient for power systems related operational problems if and only if optimal decisions are to be made for each lead-time and each location independently of each other. However, many of these operational processes are temporally and spatially coupled, while uncertainty in photovoltaic (PV) generation is strongly dependent in time and in space. This issue is addressed here by analysing and capturing spatio-temporal dependencies in PV generation. Multivariate predictive distributions are modelled and space–time trajectories describing the potential evolution of forecast errors through successive lead-times and locations are generated. Discrimination ability of the relevant scoring rules on performance assessment of space–time trajectories of PV generation is also studied. Finally, the advantage of taking into account space–time correlations over probabilistic and point forecasts is investigated. The empirical investigation is based on the solar PV dataset of the Global Energy Forecasting Competition (GEFCom) 2014.

1. Introduction

With the increase in the penetration of intermittent generation, a crucial requirement for power systems operation and planning is to enhance forecasting approaches such that they can inform about prediction uncertainties.

Over the past decade, researchers have intensively investigated solar irradiance and Photovoltaic (PV) power point forecasting for independent sites and forecast horizons. Time series analysis, AR, ARMA, ARIMA, artificial neural networks, support vector machine are among the mostly used methods in this area [1–4]. Point forecast methods work mainly based on minimum least square schemes and can only inform about conditional expectation of a random variable. Therefore, recently a significant share of practices in the energy forecasting area is concentrated on probabilistic forecasts. These approaches aim at equipping decision makers with appropriate information about stochastic behaviour of the random variables as well as uncertainties attached to the forecasts [5]. There are handful practices on probabilistic forecasts of PV generation available in the literature [6–9].

If probabilistic forecasts are properly employed, they can serve as a decision-aiding tool to alleviate challenges attached with stochastic generation. However, despite of the benefits of probabilistic forecasts over point forecasts, they fail to capture development of forecast errors through successive lead-times, interdependent generation in contiguous locations or negatively

* Corresponding author.
E-mail address: faranak001@e.ntu.edu.sg (F. Golestaneh).
[1] The authors are listed in alphabetical order.

http://dx.doi.org/10.1016/j.apenergy.2016.05.025

correlated generation levels in diverse geographic areas [10]. The reason is that they treat random variables for each lead-time and each location individually and separately while PV generations are stochastic processes with spatially spread and time interdependent infeeds. Therefore, in multi-stage decision making problems such as unit-commitment or optimal power flow, it is an integral requirement to estimate aggregated uncertainties in the system and model space–time stochasticity of intermittent resources [11,12].

Following complex meteorological mechanisms like cloud passages, PV generations act like a random variable distributed over time and space. Therefore, it is highly plausible that by leveraging spatio-temporal correlations, improved forecast accuracy can be gained.

Just recently, few practices have studied space–time correlations of PV power (or solar irradiance) and tried to benefit from them in point forecasting. Gueymard and Wilcox [13] have presented a general investigation on long-term variability of solar resources in united states. Yang et al. have proposed a statistical approach to obtain temporal and spatial stationarity for solar irradiance time series logged in few numbers of monitoring stations in Singapore. To do so, solar irradiance time series at individual sites are detrended to get temporal stationarity. Spatial stationarity also is obtained by coordinate transformation. Using time forward kriging and with respect to the persistence method, 25% improvement in RMSE is reported.

In [14], spatial cross correlations between all the PV sites under consideration are reported to be more than 0.85. In [15], almost the same results are shown where the cross correlations between each pairs of 12 PV zones with distance up to 1500 km in Sweden are found to be more than 0.8. These findings have supported the idea of using data from neighbouring sites as additional explanatory variables for PV generation forecasting at the target location [16].

Zagouras et al. have looked into space–time correlations of solar irradiance to devise forecast models for seven locations in California with one, two and three hours forecast horizons [16]. To do so, firstly, the regions which present high correlations between satellite-derived data and ground data are determined. Then, the data for those areas are employed as exogenous variables to predict global solar irradiance at the point of interest. To determine the most optimal time lags for local and exogenous variables, genetic algorithm is used.

In [17], to benefit from spatial–temporal dependencies of PV power, a vector autoregression based method is proposed. Using past observations from the neighbouring locations and for forecast horizons less than 4 h, up to 10% improvements in RMSE values are achieved. In [18] also to enhance predictability of PV power, measurements from the adjacent PV sites are used as exogenous variables.

As reviewed above, the foundation of the few present studies on the spatio-temporal PV power forecasting is on deploying measurements or meteorological data from the neighbouring locations in forecasting process of the site of interest (spatial analyses). Moreover, as the developed methods use lagged data, they are categorised as temporal investigations [19]. Here, though, spatio-temporal correlations of PV power are leveraged in a different way. In contrast with reviewed works in which dependencies are founded on the base of point forecasts, we model the dependencies based on probabilistic forecasts. The goal in this study is to provide more informative forecasts for probabilistic decision making.

Inspired by the recent multivariate analyses of the wind power, the cornerstone of this practice is to use marginal distributions given by probabilistic forecasts as infeed and couple them using copulas to form a multivariate distribution [20].

In [21], an approach is proposed to capture and model the time dependent structure of wind power generation for successive hours using marginal densities. With an Gaussian copula assumption, the covariance matrix carries information about temporal dependence. Due to non-stationary characteristics of wind power, an adaptive and recursive method is proposed to track dependency and estimate the covariance matrix for each time. The developed idea is then tested on a multi-MW wind farm with historical data available for a course of two years.

While a range of quantitative assessment frameworks for univariate quantities and independently generated marginal densities exists [22], only few frameworks for the case of multivariate quantity evaluation can be found in the literature. Energy score is the most commonly used scoring rule to evaluate multivariate densities which are described by a finite number of samples. However, the score does not discriminate the misspecified dependency structure between components of a multivariate quantity [23,24]. A variogram-based score is a recently proposed scoring rule and it is claimed to be more sensitive to misspecified mean, variance and correlations [24].

An event-based scoring rule is proposed in [23] as an diagnostic approach to assess the correspondence of trajectories generated on a multivariate base and related measurements. Frameworks are provided to evaluate time-dependent trajectories in predicting gradient and long-lasting events.

In this paper, the aim is to investigate and analyse spatio-temporal dependency of PV generations. Performance of multivariate Normal distribution with both recursive and empirical covariance matrices is evaluated. Quantile regression [25] is used here to obtain marginal densities independently. These marginal distributions then are employed as infeeds for dependency investigations. Discriminating capability of the relevant scoring rules on performance assessment of space–time trajectories of PV generation has been studied. The scores which originally have been proposed for the case of time-dependency are modified and used for space–time dependency. Three events are proposed for the case of PV generations where PV power measurements for each time of day are compared with maximum expected power for the same time. In order to track time dependency of PV generations for successive lead-times while taking into account seasonally variations of sunrise and sunset time, observations and marginal distributions are transferred to a time grid and dependency modelling is carried out on this grid. Eventually, the generating trajectories are transformed back to the original space. The quality of generated trajectories from multivariate distributions are compared with those drawn from predictive densities, normal distributions centred on point predictions and a generalisation of climatology forecasts. Analyses and verification are performed using a dataset including more than two years worth of data with hourly resolution and three neighbouring PV sites.

2. Experimental data description

As a basis for PV generation space–time dependency investigation, time series of PV generation for three contiguous zones are used. The installation specifications of these zones are given in Table 1. Technical specifications of the PV panels are given in [26–28].

To predict points and marginal densities of PV generations for three PV zones described above, 12 independent variables as the output of Numerical Weather Prediction (NWP) provided by European Centre for Medium-Range Weather Forecasts (ECMWF) are used as the explanatory variables. The period for which both NWP and PV measurements are available is from April 2012 until the end of June 2014. Therefore, in total around 800 days worth of data per zone with hourly resolution are used.

The NWP variables employed in this study are total column ice water, surface pressure, relative humidity at 1000 mbar, total