Solar Energy 149 (2017) 294–301

Contents lists available at ScienceDirect

Solar Energy

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

Brief Note

Improving time series prediction of solar irradiance after sunrise: Comparison among three methods for time series prediction

Yoshito Hirata *, Kazuyuki Aihara

Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

ARTICLE INFO

Article history:
Received 3 February 2017
Received in revised form 3 April 2017
Accepted 10 April 2017
Available online 18 April 2017

Keywords:
Short-term prediction
Local measurements
Time series prediction
Infinite-dimensional delay coordinates (InDDeCs)

ABSTRACT

Short-term prediction for renewable energy outputs up to 6 h is important especially for preparing backup power plants such as thermal power plants and hydro power plants to keep the voltage and the frequency in a power grid constant. However, short-term prediction of solar irradiance for the morning is especially difficult because measurements of solar irradiance before sunrise are zeros and useless for the prediction. Here we propose to use recently derived infinite-dimensional delay coordinates for predicting solar irradiance after sunrise based on a time series before the sunrise. As a result, one can make time series prediction by taking into account the long history of previous changes of solar irradiance. We demonstrate that the examined short-term time series prediction has effectively predictive skills because its prediction errors are smaller by about 85% best compared with the 24 h clear sky index persistence.

1. Introduction

Introducing more renewable energy resources into a power grid is demanded socially to mitigate the global warming. To keep the stability of the power grid under such a circumstance, we need to predict the outputs of renewable energy resources. Although a global circulation model (Ohtake et al., 2015) is often used to predict photovoltaic outputs, it is known that time series prediction is better than prediction by a global circulation model when the prediction step is as short as up to 2 h ahead (Bacher et al., 2009). But, the time series prediction also has a problem that its prediction performance is bad immediately after sunrise because the solar irradiance is zero and constant before sunrise and useless for predicting its future values for the morning. This problem can be seen in a linear prediction (Boland, 2015) as well as the existing nonlinear predictions as we will see in the later part of this manuscript.

Here we propose to overcome the problem of time series prediction at the morning by infinite-dimensional delay coordinates (Hirata et al., 2015a) (InDDeCs). Our key idea is that InDDeCs can retain the long historical information for the temporal changes of solar irradiance so that we also can employ the information for the solar irradiance at previous days naturally to predict the solar irradiance for the next morning.

* Corresponding author.
E-mail address: yoshito@sat.t.u-tokyo.ac.jp (Y. Hirata).

http://dx.doi.org/10.1016/j.solener.2017.04.020

2. Methods

2.1. Dataset

The time series used is a set of solar irradiance measurements at 61 sites at the central region of Japan covered by the Chubu Electric Power Company. The measurements were taken by pyranometers (photodiode or thermopile depending on the availability on each site) between 1 November 2010 and 30 November 2016. The basic quality control was conducted such as the adjustments of pyranometers every two years up to 2013 and the cleanings for the surfaces of pyranometers every year after 2014. Although the original measurements were as fine as every 10 s, we took the temporal averages for every hour. Thus, the time series used later is a series of 61 dimensional hourly solar irradiances. A part of time series is shown in Fig. 1. Please note that there are some missing data, which are shown as the black horizontal lines. For missing data, we substituted -1 and avoided any interruptions for the calculations.

2.2. Methods of time series predictions

We briefly explain, in Section 2.2.1, the overview for the used methods of time series predictions, which are all based on k-nearest neighbor approaches with different implementations. We explain, in Sections 2.2.2 and 2.2.3, the detail of the used methods.

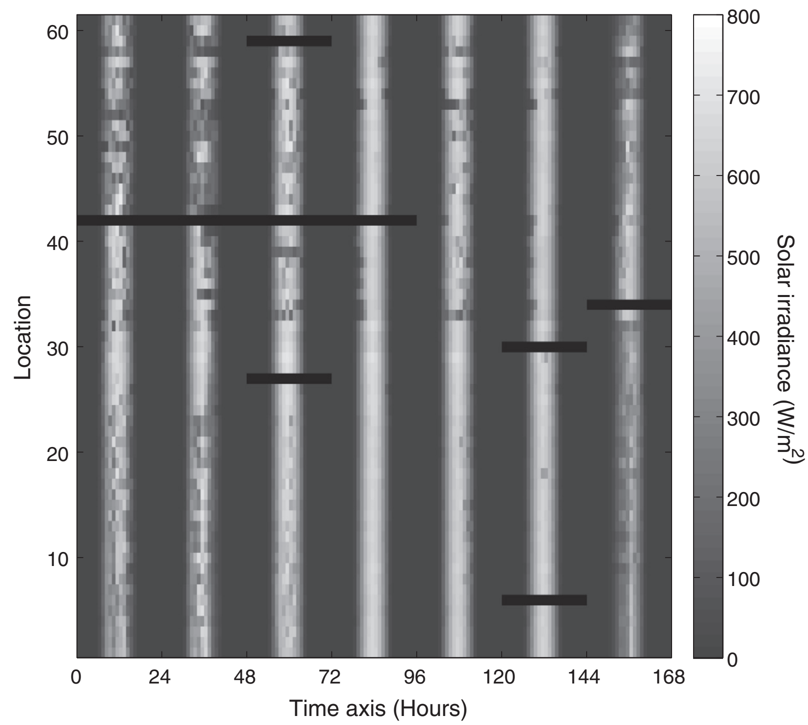

Fig. 1. Part of time series of solar irradiance used in this paper. The first 7 days in the dataset are shown here. In the vertical axis, we order 61 sites where we measured solar irradiances. The horizontal axis shows the time axis. The gray scale shows the strength of the solar irradiance, except for the darkest black parts which represent missing data points.

Thus, readers who are not interested in the detail may skip Sections 2.2.2 and 2.2.3.

2.2.1. Overview

We compared the prediction performance of InDDeCs (see Section 2.2.2) with the prediction using 24 h persistence defined by Mathiesen and Kleissl (2011) using the measured clear sky index, the extension (Hirata et al., 2014a, 2014b) of Kwasniok and Smith (2004), and barycentric coordinates (Mees, 1991; Hirata et al., 2015b) (see Section 2.2.3) using linear programming (Matoušek and Gärtner, 2007). We could not use the 1 h persistence described in Mathiesen and Kleissl (2011) because it does not produce meaningful prediction values for a few hours after the sunrise. In InDDeCs, we could virtually retain long historical data by a series of distances where the past components decay exponentially along the time (see Section 2.2.2). This kind of long history can be realized by a virtual infinite dimensional vector

$$\tilde{s}_{i,\lambda_j}(t) = (s_i(t), \lambda_j s_i(t-1), \lambda_j^2 s_i(t-2), \ldots), \quad (1)$$

for a given time series $\{s_i(t)\}$, namely a series of solar irradiances measured at 61 sites, and a decaying factor λ_j. This virtual vector can be accessed by a series of distances as discussed in Section 2.2.2. Although the night values are also included in $\{s_i(t)\}$, this InDDeCs approach can naturally overcome this problem because this approach can retain a long history virtually and can access the information related to the values in the previous days.

In the extension of Kwasniok and Smith (2004), we realized online prediction given multivariate time series (see Hirata et al. (2014a, 2014b)). In this extension of Kwasniok and Smith (2004), for each time, the 25 neighboring points $\{v(i)|i \in I_t\} = \{[v_1(i), v_2(i), \ldots, v_K(i)]|i \in I_t\}$, which are described in delay coordinates (Takens, 1981; Sauer et al., 1991) made of solar irradiance measurements, are selected by the "$L_{0.5}$-norm" defined by

$$\sum_k |v_k(i) - v_k(t)|^{0.5}, \quad (2)$$

where I_t means a set of time indexes for neighboring points of $v(t)$. In addition, we define K = (the number of sites) $\times$ (the maximum delay for the delay coordinates) = $61 \times 18 = 1098$, and k varies between 1 and K. Then, p steps ahead prediction for each p is given by

$$\frac{1}{25}\sum_{i \in I_t} v(i+p). \quad (3)$$

Here $v(i+p)$ is a point that a spatially neighboring point $v(i)$ goes after p steps. Later, the database is updated so that the algorithm can run online. In the method of barycentric coordinates using linear programming, we first found neighboring points for the current point in the past parts of the series using the above "$L_{0.5}$-norm" and obtained their weights so that the current point is represented as the weighted average of the neighboring points:

$$v(t) \sim \sum_{i \in I_t} \eta_i v(i), \quad (4)$$

where $0 \leqslant \eta_i \leqslant 1$ for each $i \in I_t$ and $\sum_{i \in I_t} \eta_i = 1$. Later, we calculated the weighted average for p steps ahead of the neighboring points

$$\sum_{i \in I_t} \eta_i v(i+p) \quad (5)$$

to generate the p steps ahead prediction (find the detail in Section 2.2.3). We predicted the spatial averages over 61 sites every 1 h up to 6 h ahead because the spatial averages are the pieces of information the operators need. For the InDDeCs as well as the extension of Kwasniok and Smith and barycentric coordinates, we applied the following two kinds of methods: in the first kind, we predicted the future spatial averages directly from the past spatial averages; in the second kind, we predicted the future spatial averages by first predicting each individual values at 61 sites and then