Available online at www.sciencedirect.com

ScienceDirect

Solar Energy 110 (2014) 22–38

SOLAR ENERGY

www.elsevier.com/locate/solener

Data-driven model for solar irradiation based on satellite observations

Ilias Bilionis *, Emil M. Constantinescu, Mihai Anitescu

Mathematics and Computer Science Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439, United States

Received 17 May 2014; received in revised form 28 August 2014; accepted 5 September 2014

Communicated by: Associate Editor Jan Kleissl

Abstract

We construct a data-driven model for solar irradiation based on satellite observations. The model yields probabilistic estimates of the irradiation field every thirty minutes starting from two consecutive satellite measurements. The probabilistic nature of the model captures prediction uncertainties and can therefore be used by solar energy producers to quantify the operation risks. The model is simple to implement and can make predictions in realtime with minimal computational resources. To deal with the high-dimensionality of the satellite data, we construct a reduced representation using factor analysis. Then, we model the dynamics of the reduced representation as a discrete (30-min interval) dynamical system. In order to convey information about the movement of the irradiation field, the dynamical system has a two-step delay. The dynamics are represented in a nonlinear, nonparametric way by a recursive Gaussian process. The predictions of the model are compared with observed satellite data as well as with a similar model that uses only ground observations at the prediction site. We conclude that using satellite data in an area including the prediction site significantly improves the prediction compared with models using only ground observation site data.

Keywords: Insolation; Irradiance; Bayesian; Recursive Gaussian process; Dimensionality reduction; Dynamical system

1. Introduction

Solar irradiation is the amount of power per square meter that reaches the Earth from the Sun. In solar energy applications, part of the solar irradiation can be converted to electricity. In contrast to conventional power sources such as coal or gas, solar irradiation is volatile and uncontrollable by the user. The two most important factors of the solar irradiation variability are the movement of the Sun and weather fluctuations. The former can be captured mathematically to great accuracy, because it is a deterministic effect. The latter is a chaotic effect, and hence the main cause for the difficulties associated with forecasting solar irradiation.

All the stages of a solar-power conversion project need to take into account the risks associated with solar irradiation. For the feasibility and design phases of the project, historical data can be employed to quantify these risks. The risks associated with the operation phase, however, require the ability to make short-term predictions (from 1 to 8 h ahead) of the solar irradiation.

The most widely used solar irradiation forecasting methodologies are those that rely only on pointwise ground measurements of the solar irradiation. The reason is that ground measurements are readily available at any solar energy production plant. Mathematically, these techniques fall into the category of single-value time series analysis. Time series analysis methods include the autoregressive integrated moving average (ARIMA) processes

* Corresponding author.
E-mail addresses: ibilion@purdue.edu (I. Bilionis), emconsta@mcs.anl.gov (E.M. Constantinescu), anitescu@mcs.anl.gov (M. Anitescu).

http://dx.doi.org/10.1016/j.solener.2014.09.009
0038-092X/

(Box et al., 2008) (see Boland, 2008 for a first-order autoregressive model (AR(1)) and (Kleissl, 2013, Ch. 15.2.2.) for ARIMA examples), and artificial neural networks (ANNs) (Mellit, 2008). The accuracy of the predictions of these models degrades rapidly as the forecasting window is increased. This result is expected because the weather fluctuations exhibit a nonlocal behavior; see (Kleissl, 2013, Ch. 15) for a comprehensive review.

More accurate forecasts can be achieved only if nonlocal data are taken into account. For short-term (0–30 min) forecasts, a promising approach is to use total sky imager technologies (Chow et al., 2011). One takes pictures of the sky from a particular site, extracts information about the clouds, constructs the cloud motion vectors (CMVs) and moves the image forward in time. Using geometrical arguments and semi-empirical models, one recovers the solar irradiation from the cloud information at that later time. The time frame for which this approach is useful depends on the velocity of the clouds.

Longer forecasts (hours to days) are feasible if satellite data are used. For forecasts ranging from 30 min to 6 h ahead, the data-driven technique introduced in Lorenz et al. (2004) may be used. As a first step, the semi-empirical heliostat method of Hammer et al. (2003) is used to extract cloud structure information from the satellite images. Then, as in the sky imager-based techniques, two consecutive images are compared in order to construct the CMVs. The cloud information is moved forward in time using the CMVs and goes through a final smoothing phase. Solar irradiation is recovered by again employing the heliostat method. Numerical Weather Prediction (NWP) may be used for forecasts of up to 6 days or longer. For example, in Kleissl (2013, Ch. 10) the sky-cover fraction of the U.S. National Digital Forecast Database is coupled with semi-empirical models to produce long-term forecasts of solar irradiation.

The main disadvantage of most these aforementioned techniques is that they are difficult to use, and, sometimes, unable to quantify the uncertainty in their predictions. Given the current evolution of decision systems for energy toward incorporating stochastic representations (Eto and Thomas, 2011), this may be a serious shortcoming. It is hard to see how to consistently add an uncertainty model to the heliostat approach. NWP models can in principle be modified to support an ensemble-based approach to uncertainty, but at a significant computational cost that requires a dedicated supercomputer (Constantinescu et al., 2011). Among the methods described, only the ARIMA-based methods can provide error bars for the predictions with a small or moderate effort. Yet, it is exactly these error bars that help quantify the potential risks and allow the stakeholders to properly price them.

These considerations have motivated us to develop a fully stochastic model that can quantify forecast uncertainties. In addition, aiming for a model that is convenient even for lean operations, we propose a satellite-based model that is considerably more robust than existing ones and can produce predictions in realtime with minimal computational resources.

Our model building philosophy and paper can be summarized as follows. As a first step, the Sun's movement effect on the satellite observed solar irradiation field (Section 2.1) is removed by dividing it with a clear sky model (Section 2.2) to get the clear sky index field. Our goal is to use consecutive observations of the clear sky index field in order to learn its dynamics. Because of its high-dimensional nature, we construct a reduced-dimensionality representation of it (Section 2.3). To learn the dynamics of this low-dimensional representation, we use a nonlinear, nonparametric technique known as recursive Gaussian process (Section 2.4). Having constructed the dynamics of the reduced space, forecasts can be performed for an arbitrary number of time steps ahead (Section 2.5). Our recursive Gaussian process is similar in concept to the ANN used in Mellit (2008). However, our model is Bayesian, a key feature that enables us to make not only best estimates but also probabilistic forecasts. We then present (Section 3) our numerical results for an 8-h-ahead forecast, and we compare them pointwise with those obtained by a recursive Gaussian process model based only on ground observations. We observe that using satellite data significantly reduces the forecast uncertainties and improves the forecast itself. We attribute this improvement to the nonlocal information carried by the satellite images and to the space–time correlation between the solar irradiation at the prediction site and at neighboring sites.

2. Methodology

Throughout this work, we denote the solar irradiation field by $I(\phi, \lambda, t)$, where ϕ, λ and t are the latitude, longitude, and time, respectively. The units of $I(\phi, \lambda, t)$ are power per square meter.

2.1. Observations of solar irradiation

The solar irradiation field can be observed almost instantaneously by processing satellite images. For simplicity, we restrict our attention to data coming from the continental United States (CONUS) scan of the GOES-13 satellite. The CONUS scan takes place almost every 30 min and has a resolution of 1 km. The solar irradiation field is constructed by processing the raw data collected by the Advanced Very High Resolution Radiometer (AVHRR) via the algorithms described in Laszlo et al. (2008). This product is reported as part of the Clouds from AVHRR Extended (CLAVR-x, https://cimss.ssec.wisc.edu/clavr/) data and is available in realtime from the University of Wisconsin (ftp://ftp.ssec.wisc.edu/clavr/goes_east/processed/. The CLAVR-x data are available on a grid of latitudes and longitudes described by ϕ_{ij} and λ_{ij} for $i = 1, \ldots, P_\phi,\ j = 1, \ldots, P_\lambda$, where P_ϕ and P_λ denote the number of pixels on each dimension. That is, at time t we observe a matrix $\mathbf{I}(t) = (I_{ij}(t))$ of size $P_\phi \times P_\lambda$: