Available online at www.sciencedirect.com

ScienceDirect

Solar Energy 114 (2015) 314–326

SOLAR ENERGY

www.elsevier.com/locate/solener

Very short term irradiance forecasting using the lasso

Dazhi Yang [a,*,1], Zhen Ye [b], Li Hong Idris Lim [c], Zibo Dong [d]

[a] *Singapore Institute of Manufacturing Technology (SIMTech), Agency for Science, Technology and Research (A*STAR), 71 Nanyang Drive, Singapore 638075, Singapore*

[b] *Modules Division, REC Solar Pte Ltd., 20 Tuas South Avenue 14, Singapore 637312, Singapore*

[c] *Department of Electronic Systems, University of Glasgow (Singapore), 535 Clementi Road, Singapore 599489, Singapore*

[d] *Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583, Singapore*

Received 30 September 2014; received in revised form 14 January 2015; accepted 16 January 2015
Available online 27 February 2015

Communicated by: Associate Editor Jan Kleissl

Abstract

We find an application of the lasso (least absolute shrinkage and selection operator) in sub-5-min solar irradiance forecasting using a monitoring network. Lasso is a variable shrinkage and selection method for linear regression. In addition to the sum of squares error minimization, it considers the sum of ℓ_1-norms of the regression coefficients as penalty. This bias–variance trade-off very often leads to better predictions.

One second irradiance time series data are collected using a dense monitoring network in Oahu, Hawaii. As clouds propagate over the network, highly correlated lagged time series can be observed among station pairs. Lasso is used to automatically shrink and select the most appropriate lagged time series for regression. Since only lagged time series are used as predictors, the regression provides true out-of-sample forecasts. It is found that the proposed model outperforms univariate time series models and ordinary least squares regression significantly, especially when training data are few and predictors are many. Very short-term irradiance forecasting is useful in managing the variability within a central PV power plant.

Keywords: Lasso; Irradiance forecasting; Monitoring network; Parameter shrinkage

1. Introduction

Variability in solar irradiance reaching the ground is primarily caused by moving clouds. To accurately forecast the irradiance, cloud information must be directly or indirectly incorporated into the formulation. Due to the stochastic nature of the clouds, it is difficult to fully model their generation, propagation, and extinction using physical approaches. Statistical methods are therefore often used to extract cloud information from observations (e.g. Yang et al., 2015; Dong et al., 2014; Lonij et al., 2013).

We are particularly interested in very short term (sub-5-min) irradiance forecasting as the clouds are relatively persistent during a short time frame. Unlike the forecasts with longer horizons where the results are essential for electricity grid operations, very short term forecasts find their applications in large photovoltaics (PV) installations. Knowing the potential shading/unshading over a particular section of a PV system in advance may be advantageous to maximum power point tracking algorithms (Hohm and Ropp, 2000). Accurate sub-minute forecasts could also bring

* Corresponding author. Tel.: +65 9159 0888.
E-mail address: yangdazhi.nus@gmail.com (D. Yang).

[1] Previously at: Solar Energy Research Institute of Singapore (SERIS), National University of Singapore, Singapore.

http://dx.doi.org/10.1016/j.solener.2015.01.016
0038-092X/

possibilities to better control of ramp-absorbing ultracapacitors (Mahamadou et al., 2011; Teleke et al., 2010).

Inman et al. (2013) reviewed the state-of-the-art methods for very short term irradiance forecasting. The methods involve using either sky cameras (Nguyen and Kleissl, 2014; Yang et al., 2014c; Quesada-Ruiz et al., 2014) or a sensor network (Lipperheide et al., 2015; Bosch and Kleissl, 2013; Bosch et al., 2013). All of these listed references aim at explicitly deriving the cloud motion and thus forecast the irradiance. Beside many assumptions, such as linear cloud edge, that have to be made, various types of error will be embedded in different phases of such methods, especially during the conversion from cloud condition to ground-level irradiance. It is therefore worth investigating the alternative methods where cloud information is considered indirectly.

Along-wind and cross-wind correlations observed between two irradiance time series have been studied intensively in the literature (e.g. Arias-Castro et al., 2014; Hinkelman, 2013; Lonij et al., 2013; Perez et al., 2012). If along-wind correlation between a pair of stations can be observed, we can use regression-based methods for forecasting. However, several problems have to be addressed before we describe our method:

- The discrepancy between the direction of a station pair and the direction of wind may result in a smaller correlation. How do we incorporate the strength of cross-correlation between monitoring sites into the forecasting model?
- When the wind speed changes from day to day or even within a day, the choices of lagged time series also need to be constantly updated. How do we then automatically select the most appropriate spatio-temporal neighbors for forecasting?
- When the correlation is unobserved, do we need to switch the spatio-temporal forecasting algorithm to a purely temporal algorithm in an ad hoc manner?

With these questions, we consider the lasso (least absolute shrinkage and selection operator) regression (Efron et al., 2004; Tibshirani, 2011, 1996). Lasso is a variable shrinkage and selection method for linear regression. In our application, the predictors (regressors) are the time series collected at the neighboring stations at various time lags (autocorrelated time series may also be used); the responses (regressands) are the time series collected at the forecast station. Some advantages of the lasso over the ordinary least squares regression, ridge regression and subset selection methods are discussed in Section 2.

1.1. Data

Data from a dense grid of irradiance sensors located on Oahu Island, Hawaii, are used in this work. The network is installed by the National Renewable Energy Laboratory (NREL) in March 2010. It consists of 17 radiometers, as shown in Fig. 1. The sampling rate of these stations is 1 s. Previously, Hinkelman (2013) showed the possibility of observing highly correlated time series from this network; data from 13 days dominated by broken clouds were used in that study. We therefore use the data from the exact same days (Hinkelman, 2014) to study the predictive performance of such network configuration. The data are freely available at http://www.nrel.gov/midc/oahu_archive/.

Throughout the paper, the 1 s irradiance data will be averaged into various intervals to evaluate the forecasts with different forecast horizons. As high frequency data often have local maxima and minima caused by noise rather than cloud effects (Bosch and Kleissl, 2013), the smallest aggregation interval is 10 s. Prior to any forecasting, the global horizontal irradiance (GHI) time series from these 17 stations are first transformed into clearness index time series. Such transformation is commonly used in irradiance forecasting to stabilize the variance, i.e., to remove the diurnal trends in the GHI time series. We use the solar positioning algorithm developed by Reda and Andreas (2008) for extraterrestrial irradiance calculation. Finally, we include a zenith angle filter of <80°.

1.2. Error metrics

All the forecasting models in this paper are built using the clearness index time series and the errors are evaluated using the GHI transformed back from the forecast clearness index. Two error metrics are used in this paper, namely, the normalized mean absolute error (nMAE) and the forecast skill (FS). The nMAE is given by:

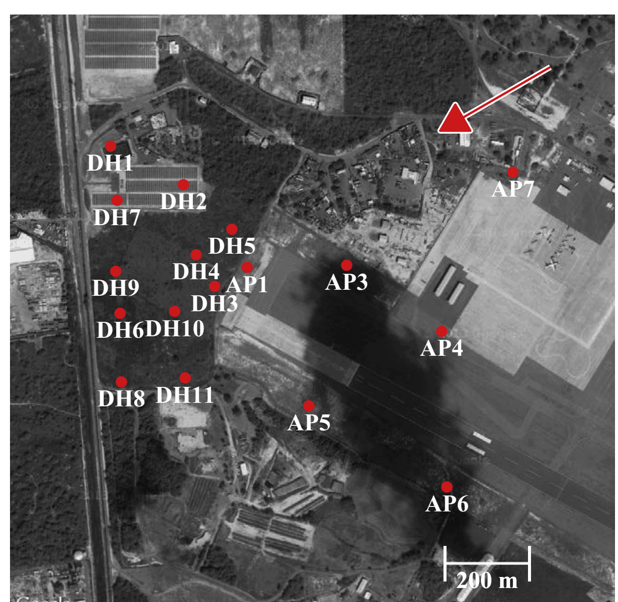

Fig. 1. Layout of the 17 stations of the NREL Oahu network. The scale of the map is shown in the bottom right corner. The arrow in the top right corner shows the prevailing trade winds direction (60° from north). The average wind speed during the periods of analyses is 10 m/s. See Arias-Castro et al. (2014) and Hinkelman (2013) for more details on the data.