Solar Energy 171 (2018) 876–883

Contents lists available at ScienceDirect

Solar Energy

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

Kriging for NSRDB PSM version 3 satellite-derived solar irradiance

Dazhi Yang

*Singapore Institute of Manufacturing Technology, Agency for Science, Technology and Research (A*STAR), Singapore*

ARTICLE INFO

Keywords:
Solar resource
Kriging
Physical solar model
Satellite-derived irradiance

ABSTRACT

The National Solar Radiation Database (NSRDB) has provided solar resource data for over 25 years. Its most recent update, namely, physical solar model (PSM) version 3, generates half-hourly, gridded, satellite-derived irradiance data with a spatial resolution of 4 km × 4 km, covering most of America. The total volume of the data is over 40 TB. Since the main method to access the data is using API with a daily limit of 2000 requests, it would require 1000 days to obtain one year of PSM data from approximately 2 million pixels. Furthermore, such a big dataset is difficult to store and manipulate. In this regard, this paper empirically investigates the accuracies of various kriging methods, so that a suitable, dimension-reduced (in space) dataset can be opted during various spatio-temporal analyses, such as forecasting or monitoring network design.

1. Introduction

In the recent decades, solar resource assessment has shifted from using mainly empirical modeling and data collected at ground-based stations to using physical modeling and satellite-derived data. The National Solar Radiation Database (NSRDB) is a well-known database for solar resource assessment applications, with good spatio-temporal coverage and data quality. Throughout the 25 years of existence of the NSRDB, numerous updates have been made. Most noticeably, the database has pioneered the area of resource assessment, and has switched from a station-specific database to a database hosting gridded satellite-derived products. Its most recent update, namely, physical solar model (PSM) version 3 (Habte et al., 2017; Sengupta et al., 2015; Sengupta et al., 2014), generates 30-min satellite-derived irradiance over a regular grid with a spatial resolution of 4 km × 4 km. The geographical coverage of PSM v3 extends to (25°W, 60°N) in the northeast direction and (175°W, 20°S) in the southwest direction, covering most of America (both continents), whereas the temporal coverage is 1998–2016.

Besides resource assessment, satellite-derived irradiance can be used in a variety of ways. Whereas some applications require data from a single pixel, e.g., sizing of a photovoltaic energy system, others require data over an area, e.g., forecasting or monitoring network design (Yang et al., 2018). Due to its extensive spatio-temporal coverage, the PSM data is enormous in size, over 40 TB. To that end, the first challenge faced by researchers who wish to perform spatio-temporal analyses on the PSM data is accessing the data (see below). Furthermore, the PSM data is spatially rich, and such high spatial granularity is not always desired. For example, in a spatio-temporal statistical forecasting context, the many highly correlated time series from neighboring pixels lead to a $p > n$ regression problem that introduces instability to lasso-based predictor selection (Yang et al., 2015). Therefore, a dimension-reduced dataset is often desired.

A primary goal for dimension reduction is minimizing information loss. The tradeoff between the reduced data size and its information content thus needs to be carefully studied. In the present case, the dimension reduction is in the form of space, i.e., sampling fewer points within an area, hence, spatial prediction accuracy naturally becomes the most important evaluation criterion. Spatial prediction through interpolation and extrapolation has been studied extensively in spatial statistics (Cressie, 2015). In the field of solar engineering, there is also a rich literature (e.g., Lorenzo et al., 2017, Rodríguez-Amigo et al., 2017). In both spatial statistics and solar engineering, kriging—the optimal prediction—has received most attention, and is constantly being validated as the best spatial prediction method. Kriging is thus used here to predict the half-hourly PSM irradiance through various dimension-reduced datasets.

Kriging is often used for gap-filling tasks in various scientific domains, such as remote sensing or meteorology, e.g., predicting aerosol optical depth or ozone concentration at unobserved locations. These works are typified by the many contributions from Noel Cressie and his colleagues (e.g., Cressie et al., 2010, Noel and Gardar, 2008). The reader is referred to the book by Cressie and Wikle (2015), and the references therein, for a complete guide on kriging and its applications. In this paper, kriging is used for a different task, namely, uncertainty quantification.

The merit of this paper goes to the fact that it uses empirical evidence to challenge the common misperception on equivalence between

E-mail address: yangdazhi.nus@gmail.com.

https://doi.org/10.1016/j.solener.2018.06.055
Received 12 March 2018; Received in revised form 22 May 2018; Accepted 13 June 2018

data amount and information content in satellite-derived irradiance data. In various case studies below using lattices with lower resolutions, the kriged irradiance reaches the same accuracy as the raw data. This implies that the original dataset contains redundant information, and can be subsetted (in the form of space) in many applications. Some guidelines of subsetting the PSM data, spatially, are given towards the end.

The remaining part of the paper is arranged as follows. Section 2 describes the PSM v3 data and the kriging experiment setup. Section 3 briefly reviews the kriging methodology in spatial statistics. Unlike in other application papers where only the prediction equations and variogram model function forms are given, I follow Cressie (2015) closely, and provide a summary on modeling, prediction, and parameter estimation of some commonly used kriging variants. Since satellite-derived irradiance is often biased, its accuracy must be validated, and that is performed in Section 4. Section 5 presents the main results of this paper. Nevertheless, those results presented in Section 5 are apparent errors (see below); the linkage between the apparent error and "true" error is drawn in Section 6. Conclusions and recommendations follow at the end.

2. Experiment setup

The PSM v3 data can be downloaded in three ways: (1) via the NSRDB viewer, an online tool, (2) via API, and (3) via Globus, a research data management service. Due to a download-file-size limit, the manual method is convenient only if data at a single location, or over a very small area, is needed. For the API approach, once an API key is obtained—almost instantly after requesting—the user could download data by specifying the text-string request parameters, such as *attributes* (e.g., "dni,dhi,ghi"), *names* (e.g., "2016"), or *interval* (e.g., "60"). Since the downloading can be automated by computer programs,[1] the API approach is used this paper. That said, the API approach is restricted to use with only a single location, for a single year at a time. In this regard, the third approach is suitable for downloading very large archives.

2.1. Data

Since the PSM data is difficult to be handled as a whole, a subset is considered in this paper. More specifically, the 30-min data from the year 2016 from 4100 regularly-gridded locations in California is used. The grid has a spatial resolution of $0.1° \times 0.1°$, or approximately 10 km × 10 km. Whereas this data is used to fit the kriging models, data from a separate set of 50 randomly chosen, non-coincident (with the fitting lattice) locations is used for validation. The fitting lattice and validation locations are shown in Fig. 1 (a).

2.2. Dimension-reduced lattices

To investigate the kriging accuracy under spatial dimension reduction, I consider two types of dimension-reduced lattice. The first type has regular grids with lower spatial resolutions, as shown in Fig. 1 (b)–(d). It is noted that a grid cell is only considered to be valid if its center falls within the geographic boundary of California. The spatial resolutions for these dimension-reduced lattices are 0.2°, 0.3°, and 0.4°, respectively. As a result, the numbers of fitting locations in these lattices are 1039, 454, and 259, respectively.

The second type of lattice is irregular. To generate these lattices, random locations are chosen from the initial 4100 locations. For the purpose of comparison, the number of random locations in type-two lattices follows the previous number of fitting locations. Voronoi diagrams are used for visualization—the fitting locations are the centers of the partitions, as shown in Fig. 2 (a)–(c). Lastly, a lattice with only 50 points, see Fig. 2 (d), is used to test an extreme case, where the number of fitting locations is rather small. In the subsequent text, the lattices shown in the two figures are referred to as LATTICE1, $\cdots$, LATTICE8.

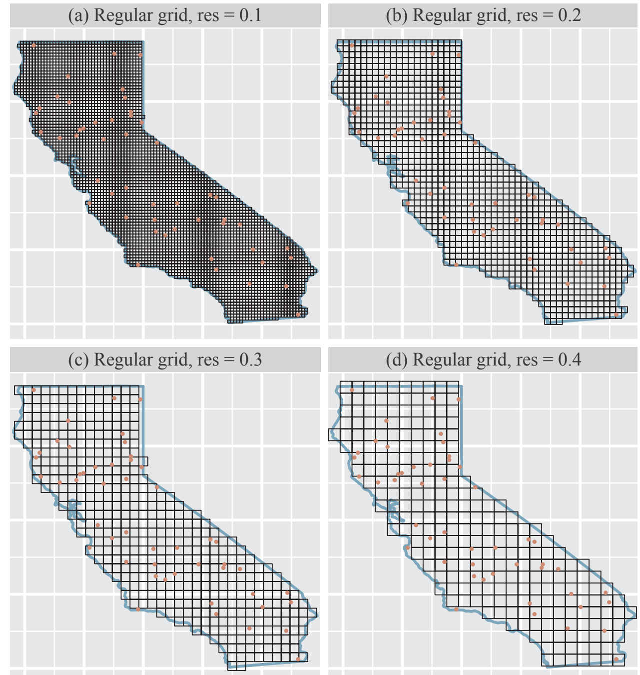

Fig. 1. Lattices of fitting data at 4 different resolutions (in degrees) in regular grids. The numbers of locations used in kriging are 4100, 1039, 454, and 259 in (a), (b), (c), and (d), respectively, whereas the 50 randomly chosen validation locations are marked with dots.

Fig. 2. Same as Fig. 1, but the lattices are irregular. The number of locations in (a)–(c) follows the size of the previous lattices (see Fig. 1), whereas (d) uses only 50 locations to picture an extreme scenario.

[1] A python version is provided by NREL, and an R version is provided with this paper.