Renewable Energy 80 (2015) 153–165

Contents lists available at ScienceDirect

Renewable Energy

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

Objective framework for optimal distribution of solar irradiance monitoring networks

Athanassios Zagouras [a], Alexander Kolovos [b], Carlos F.M. Coimbra [a,*]

[a] Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, Center of Excellence in Renewable Energy Integration and Center for Energy Research, University of California San Diego La Jolla, CA 92093, USA
[b] SpaceTimeWorks LLC, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 10 October 2014
Accepted 20 January 2015
Available online

Keywords:
Solar data
Cluster analysis
Adaptive affinity propagation
Bayesian maximum entropy
Spatiotemporal prediction
Geostatistics

ABSTRACT

Time-resolved characterization of solar irradiance at the ground level is a critical element in solar energy analysis. Siting of nodes in a network of solar irradiance monitoring stations (MS) is a multi-faceted problem that directly affects the determination of the solar resource and its spatio-temporal variability. The present work proposes an objective framework to optimize the deployment of solar MS over a sub-continental region. There are two main components in the proposed methodology. The first employs cluster analysis using the affinity propagation algorithm, to select the optimal number of clusters (regions with coherent solar microclimates) upon internal coherence criteria. The second component employs stochastic prediction and validation, through the use of a Bayesian maximum entropy method, and selects the optimal MS configuration, according to geostatistical criteria, among the solutions recommended by the cluster analysis. We apply this two-pronged methodology to determine clusters and optimal locations for global horizontal irradiance monitoring across the state of California. In this proof-of-concept study, 3 disparate MS configurations are examined within the cluster partition. The subsequent geostatistical analysis indicates that all configurations rank almost equally well based on different statistical error measures. The optimal configuration can be singled out depending on desired criteria of choice.

1. Introduction

Ground-based solar irradiance monitoring stations (MS) are the most reliable source to provide accurate assessment of ground solar radiation. Ground-based monitoring networks have been installed throughout the world for different purposes, such as the Baseline Surface Radiation Network [1], the European Solar Radiation Atlas [2], the California Irrigation Management Information System (CIMIS). Common limitations in such networks are (i) the limited spatial coverage compared to satellite modeled irradiance data in high spatial resolution gridded domains, and (ii) the lack of prior knowledge regarding the MS spatial representativeness that could enable the development of operational plans about MS installation locations. Traditional network design analysis relies mainly on topological design to determine locations for MS.

In recent years, advanced data mining methods (e.g., cluster analysis) have provided insight on the coherence of spatial groups. This feature has helped clustering methods find application in designing monitoring station networks; see, e.g., [3]. At the same time, such methods often require a relatively large number of observations, which typically exceeds the number of monitoring stations within a region. Geophysical clustering methods can be enhanced substantially by remote sensing, which provides extensive attribute coverage in increasingly detailed spatial and temporal resolutions. A shortcoming of satellite data is the relatively low accuracy and spatio-temporal resolution levels. Most radiation models from remote sensing depend strongly on calibration sites and dynamic atmospheric parameters; for example, satellite data of global horizontal irradiance (GHI) and direct normal irradiance (DNI) incur errors of the order of 5–40% depending on location and time granularity [4,5].

Several studies assess the representativeness of measuring points by comparing solar observations measured from both satellites and ground sites [6,7,8]. In related studies, ground-based stations are members of existing sensor networks and they are

* Corresponding author.
E-mail addresses: azagouras@eng.ucsd.edu (A. Zagouras), alexander.kolovos@spacetimeworks.com (A. Kolovos), ccoimbra@ucsd.edu (C.F.M. Coimbra).

http://dx.doi.org/10.1016/j.renene.2015.01.046
0960-1481/

used to define the *ground–truth*. Zagouras et al. [9,10,11] were the first to use cluster analysis of features extracted from satellite-derived solar data to determine the appropriate number of coherent clusters within different domains of interest. The results of these studies were based on criteria related to the convergence of the clustering methods to cluster partitions, validated with clustering validity metrics. An appropriate number of clusters is determined by assessing the clustering quality among a variety of clustering solutions.

In addition to clustering, geostatistics is also used in network design studies. For example, in the presence of sparse measurements [12], show how geostatistical predictive techniques in conjunction with simulation iterations can assist in determining optimized locations. More commonly, though, geostatistics is utilized to model an attribute in space and/or time from a set of measured values. Specifically, in solar literature solar irradiance is often modeled by using parametric models, e.g. [13], autoregressive techniques for solar irradiance time series, e.g. [14], and geostatistics. Geostatistics has the benefit of considering solar irradiance as a stochastic random field [15], thus providing insight about the solar irradiance field structure, homogeneities/stationarity and internal characteristics. In this process, geostatistics enables integration of relevant information about the solar irradiance field, and utterly enables prediction on the basis of spatial and/or temporal correlation.

Most often, methodologies used for geostatistical prediction are linear model-based techniques such as the kriging family of predictors. Kriging techniques have been previously used to predict solar irradiance at unsampled locations, both within purely spatial and spatiotemporal contexts (e.g., [16,17,18]). More recently [19], exhibited a more in-depth kriging example in the joint spatiotemporal continuum, and indicated the importance of spatiotemporal prediction for solar irradiance forecasting. Despite their mainstream character, these geostatistical techniques are built on restrictive assumptions and known limitations. For example, linear models are used to describe phenomena that are inherently nonlinear, and Gaussian assumptions are required for the data distributions by the linear interpolators. An additional major weakness of mainstream techniques is the inability to account rigorously for nontrivial uncertainty in the data set. Thus, in the presence of uncertain measurements such as probabilistic distributions or interval data, measurements are either skipped or being used by reducing their informational content to single values.

We propose the Bayesian maximum entropy (BME) method as an advanced alternative to the mainstream geostatistical methodologies. Being free of limitations and weaknesses like the above, BME additionally features very attractive characteristics for solar irradiance studies, such as the ability to rigorously incorporate uncertain data. This methodology has been applied previously broadly and successfully in fields such as environmental sciences, atmospheric monitoring, and environmental health risk and assessment; see, e.g., [20,21,22,23]. BME as a valuable alternative for solar irradiance prediction was first introduced by Kolovos in Ref. [24] in a limited study across the USA. An extensive application of BME in accurately predicting subhourly solar photovoltaic output over state-size areas was presented more recently by Lee et al. in Ref. [25]. On the side of teaming with cluster analysis in the proposed framework, our study extends the previous geostatistical solar literature by applying BME prediction for solar irradiance on a spatially large scale in the state-wide domain of California, and a temporally systematic scale at subhourly 30-minute intervals over a period of days across a calendar year. Moreover, to the best of our knowledge it is the first time geostatistics is implemented as an effective tool in a cross-disciplinary solar irradiance analysis in a double role; that is, both for energy resources assessment at this detailed space-time level, and as a critical component for planning and decision-making in the deployment of state-wide solar projects.

In the following, we begin with a descriptive overview of our proposed framework in Section 2, alongside with information about our case study in the state of California. Then, Section 3 provides a closer view to the collaborating methodologies in our framework prior to stepping through our analysis in detail. In Section 4 we discuss the results of this work, followed by our conclusions.

2. A comprehensive outlook

We approach the topic of selecting the number and installation locations for an MS network by introducing a 2-segment analysis framework. Our proposed framework is the composition of two complementary segments that combine cluster analysis schemes and geostatistics. The first segment is comprised by Optimal Cluster Selection (OCS), which is a clustering process and validation methodology to select an optimal number of clusters (NC) from an initial given set of GHI measurements. NC is the number of MS to be installed, by assigning a MS to each cluster area. Therefore, selecting the NC depends, for example, on the available resources to be invested for the creation of a MS network. Given the level of resources, a client can specify a feasible numeric window of MS. In response, OCS provides the analysis to determine an optimal NC, where optimality is deemed in terms of clustering validity assessment by adhering to statistical coherence features. In all, the OCS analysis improves MS placement by determining a structured station network scheme on the basis of (i) cluster analysis, and (ii) internal coherence criteria rooted in the cluster geometry.

OCS begins by performing cluster analysis of GHI temporal vectors that represent the solar irradiance temporal activity over the nodes of a gridded domain. The clustering algorithm employed in this study is the Adaptive Affinity Propagation (adAP), and yields a potential range of NC controlled by the algorithm's convergence criteria. From this range of NC we determine the optimal NC using a knee point detection scheme [26] on scientific criteria that pertain to cluster validity assessment [27,28]. Namely, in the following we estimate the change of gradient (knee-point) of an evaluation graph of clustering quality metrics calculated through the measures of compactness and separation among the derived clusters.

Using the above scientific criteria, OCS produces possible MS configurations on the basis of the derived spatial segmentation into the optimal NC. For illustration, our work examines 3 such different location configurations; (A) the exact cluster centers as the obtained from the clustering process, (B) the clusters centers indicated by the median vector among each cluster's vector set, and (C) a random selection of location within the region of each cluster. Clearly, the concept of determining the alternative cluster locations in B and C is limited by the fixed area of a cluster, as well as the position of the exact center. We investigate additional characteristics of these configurations A, B, and C in the analysis Section 4. OCS makes configuration recommendations by examining the coherence of a cluster in terms of its intra-variance. In that sense, good quality clustering is expected to produce highly coherent cluster regions with low variance among the data members within a cluster. Consequently, we aim to explore the extent to which the derived clustering is able to capture clusters with low intra-variance and, thus, how representative the main cluster center is when compared to other candidate locations within the cluster.

After an optimal NC is determined and candidate MS configurations are suggested, the OCS methodology results are wired to the second analysis segment of our framework. The second segment uses geostatistical spatiotemporal analysis to sift through candidate