Renewable Energy 89 (2016) 351–359

Contents lists available at ScienceDirect

Renewable Energy

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

A regime-dependent artificial neural network technique for short-range solar irradiance forecasting

T.C. McCandless [a, b, *], S.E. Haupt [a, b], G.S. Young [b]

[a] National Center for Atmospheric Research, 3450 Mitchell Lane, Boulder, CO 80301, USA
[b] The Pennsylvania State University, Department of Meteorology, 503 Walker Building, University Park, PA 16802-5013, USA

ARTICLE INFO

Article history:
Received 8 October 2015
Received in revised form
8 December 2015
Accepted 10 December 2015
Available online xxx

Keywords:
Solar irradiance
Artificial neural network
Regime-dependent prediction
m-Means clustering
Irradiance variability

ABSTRACT

Solar power can provide substantial power supply to the grid; however, it is also a highly variable energy source due to changes in weather conditions, i.e. clouds, that can cause rapid changes in solar power output. Independent systems operators (ISOs) and regional transmission organizations (RTOs) monitor the demand load and direct power generation from utilities, define operating limits and create contingency plans to balance the load with the available power generation resources. ISOs, RTOs, and utilities will require solar irradiance forecasts to effectively and efficiently balance the energy grid as the penetration of solar power increases. This study presents a cloud regime-dependent short-range solar irradiance forecasting system to provide 15-min average clearness index forecasts for 15-min, 60-min, 120-min and 180-min lead-times. A k-means algorithm identifies the cloud regime based on surface weather observations and irradiance observations. Then, Artificial Neural Networks (ANNs) are trained to predict the clearness index. This regime-dependent system makes a more accurate deterministic forecast than a global ANN or clearness index persistence and produces more accurate predictions of expected irradiance variability than assuming climatological average variability.

1. Introduction

The proliferation of photovoltaic (PV) power production has made accurate short-range solar irradiance forecasts a necessity for utility companies to ensure reliable and efficient integration of solar power into the energy grid. An accurate forecast for irradiance is necessary; however, a prediction of the variability of the irradiance is also helpful in maintaining reliable energy production with increased levels of solar power integration. The amount of solar irradiance reaching the PV panels depends on both the diurnal cycle and on the atmospheric state. While the diurnal cycle is easily forecast, the stochastic element of cloud formation makes that component of irradiance forecasting a challenge. This problem can be mitigated to some extent if one can forecast the cloud type. The identification of cloud types, i.e. cloud regimes, is a valuable tool in short-range solar irradiance forecasting because each cloud regime is associated with particular cloud properties such as cloud optical depth, cloud growth rate, and cloud dissipation rate; and therefore, have various degrees of irradiance attenuation. The cloud type also impacts the short-range temporal and spatial irradiance variability and the corresponding irradiance forecast uncertainty.

The optimal method for solar irradiance prediction depends on several factors, including the forecast lead time, with statistical techniques and cloud advection techniques most effective for short-range irradiance forecasting. Short-range forecasting is defined here as solar irradiance predictions from 15 min out to 180 min. Predicting solar power through statistical techniques has gained the attention of researchers in recent years [22]. found that a Support Vector Machine approach to post-processing Numerical Weather Prediction (NWP) Models' forecasts produced lower Global Horizontal Irradiance (GHI) forecast error compared to linear regression post-processing techniques [8] and [25] found that AutoRegressive Integrated Moving Average (ARIMA) models produced lower solar irradiance and solar power errors compared to other time-series short-range prediction techniques while [17] used a Markov process to predict sunshine and cloud cover [15]. reported that Artificial Neural Networks (ANNs) have been used in modeling and predicting solar radiation more than any other non-linear technique. More recently, several studies determined that

* Corresponding author. National Center for Atmospheric Research, 3450 Mitchell Lane, Boulder, CO 80301, USA.
E-mail addresses: mccandle@ucar.edu (T.C. McCandless), haupt@ucar.edu (S.E. Haupt), g3y@psu.edu (G.S. Young).

http://dx.doi.org/10.1016/j.renene.2015.12.030
0960-1481/

models based on ANNs improve solar irradiance or solar power forecast accuracy compared to various baseline techniques [2,3,6,12,14,23]. Several studies have examined the performance of these statistical forecast models in various weather conditions [18]. found the accuracy of an ANN optimized with a Genetic Algorithm had a strong seasonal dependence [13]. correlated total sky images, infrared data, and solar radiation observations at the surface to use as input to an ANN and found the variability of solar radiation to be strongly dependent on the amount of cloud cover. Each day was classified as sunny, partly sunny or cloudy and an ANN was used to forecast the daily profile of the power produced by a PV plant [16]. [4] concluded that the ANN model has lower errors for days characterized by direct irradiance (clear days) and for days characterized by diffuse irradiance (cloudy days) than for days characterized by a mix of direct and diffuse irradiance (partly cloudy days).

This current work seeks to improve two major facets of short-range solar irradiance forecasting via regime-dependent statistical forecasting: deterministic irradiance forecast accuracy and irradiance variability estimates. We first classify cloud regimes with a k-means algorithm and then apply ANNs to each regime to produce a more accurate GHI forecast with variability estimates. The GHI prediction is generally a necessary step in the prediction process before converting to power. The k-means algorithm statistically classifies the cloud regime based on surface weather and irradiance observations. This approach parallels that of [5]; who classified weather regimes with Principal Component Analysis (PCA) in order to apply regime-dependent optimal weights to ensemble temperature forecasts. After k-means clustering, ANNs are implemented for each weather regime independently with the intention of modeling each weather regime's inherent predictability, and thus, each regime's different causal relationships between predictors and predictand. Predictions are made for the clearness index (Kt), which is the ratio of the observed GHI at the surface to the Top Of Atmosphere (TOA) expected GHI. The prediction of Kt is important for utility companies because it quantifies the amount of attenuation from aerosols and clouds at a particular location [13]. Although it is common to classify days at being cloudy, partly cloudy or clear, we classify cloud regimes that are specific to the short-range development of clouds and therefore improve the forecast error for lead-times up to 180-min.

We wish to make short-range predictions for multiple sites near Sacramento, California for 15-min intervals out to 180 min. In operational forecasting, these short-range predictions are blended with forecasts from NWP models and a satellite based cloud advection technique then converted to power in the National Center for Atmospheric Research SunCast System that predicts solar power out to 168-h [7].

Section 2 provides an overview of our approach. In section 3, we discuss the data and the additional predictors derived from the initial datasets, which are the Sacramento Municipal Utility District (SMUD) irradiance network and the METAR network. In section 4, the prediction techniques of the ANNs and the baseline clearness index persistence forecast are described. In section 5, we summarize the k-means algorithm for cloud regime classification and the selection of optimal inputs. We describe the prediction methods before the regime classification method because we use the ANN prediction method to inform our decision on the best selection of inputs for the k-means regime classification. In section 6, we present and discuss the prediction results. The final section, 7, summarizes and poses potential future work.

2. Approach

The goal of this work is to develop a cloud regime-dependent short-range solar irradiance forecast system in order to not only improve the deterministic forecast accuracy, but also to provide a quantification of the expected solar irradiance variability and corresponding forecast uncertainty. This section outlines our classify-then-predict process; the details are described in the following sections. Our methodology begins by classifying the cloud regime with the k-means algorithm. We then train a separate ANN to make predictions for each individual regime as depicted in Fig. 1. This novel work goes beyond [16] and [12]; and others in the sophistication and automation in identifying regimes with the k-means algorithm and in the regime-dependent configuration of the ANNs that are specific to improving the final prediction algorithm. The process begins by selecting the optimal set of inputs for cloud regime classification that corresponds to the final model with the lowest forecast error. The selected set of inputs is then used by the k-means algorithm to classify and partition the datasets into an optimal number of cloud regime subsets. Finally, ANNs are constructed on each of the cloud regime datasets independently. This classify-then-predict process (with k-means then ANN) is repeated for each forecast lead-time.

A cloud regime-dependent Artificial Intelligence (AI) system requires dividing the cases into distinct regimes for which the fundamental relationship between predictors and predictand is expected to differ, and therefore, to allow more accurate short-range forecasts. Thus, careful sensitivity studies determined the optimal configurations of the AI models in order to match the complexity of the relationships among the predictors in the regimes. After all data are quality controlled and additional variables are derived, the datasets are randomly split 2/3 for training and 1/3 for testing. All of the results shown are from the testing datasets; however, the sensitivity tests conducted to determine the optimal configurations of the system were performed on the training datasets. The ANN and k-means sensitivity studies similarly split the training dataset into 2/3 for training and 1/3 for testing and the optimal configuration was determined based on this 1/3 independent test set. This approach avoids compromising the independence of the initial test dataset. We show results in this study for four forecast lead-times: 15 min, 60 min, 120 min, and 180 min. These predictions are for the 15-min average clearness index

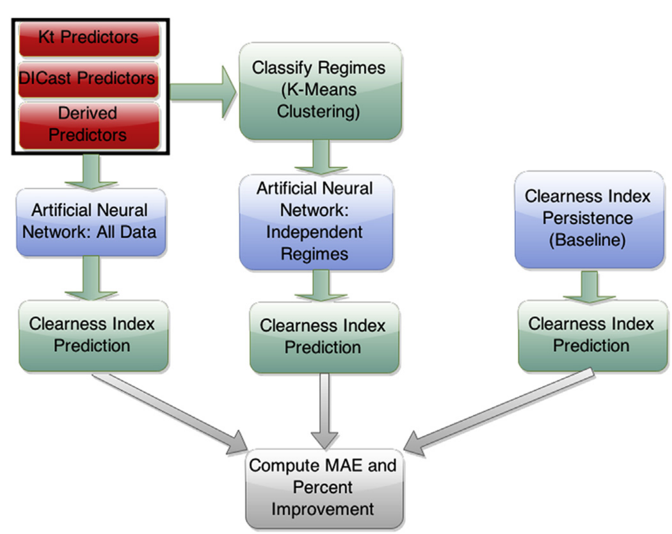

Fig. 1. Process design: first classify cloud regimes on the optimal set of potential inputs shown in the red rectangles outlines in the black box, then apply ANN models to predict the clearness index on each regime independently. An ANN is also applied on all data (i.e. without regime identification), and compared to the clearness index persistence prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)