Remote Sensing of Environment 212 (2018) 176–198

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities

Sujan Ghimire*, Ravinesh C. Deo*, Nathan J. Downs, Nawin Raj

School of Agricultural Computational and Environmental Sciences, International Centre for Applied Climate Sciences, Institute of Agriculture and Environment, University of Southern Queensland, Springfield, QLD 4300, Australia

ARTICLE INFO

Keywords:
Satellite solar prediction model
Particle swarm optimization
Neural network
Genetic algorithm
Support vector machine
Grid search
Genetic programming
Giovanni
ECMWF
Extreme learning machine

ABSTRACT

Designing predictive models of global solar radiation can be an effective renewable energy feasibility studies approach to resolve future problems associated with the supply, reliability and dynamical stability of consumable energy demands generated by solar-powered electrical plants. In this paper we design and present a new approach to predict the monthly mean daily solar radiation (*GSR*) by constructing an extreme learning machine (ELM) model integrated with the Moderate Resolution Imaging Spectroradiometer (MODIS)-based satellite and the European Center for Medium Range Weather Forecasting (ECMWF) Reanalysis data for solar rich cities: Brisbane and Townsville, Australia. A self-adaptive differential evolutionary ELM (i.e., SaDE-ELM) is proposed, utilizing a swarm-based ant colony optimization (ACO) feature selection to select the most important predictors for *GSR*, and the SaDE-ELM is then benchmarked with nine different data-driven models: a basic ELM, genetic programming (GP), online sequential ELM with fixed (OS-ELM) and varying (OSVARY-ELM) input sizes, and hybridized model including the particle swarm optimized-artificial neural network model (PSO-ANN), genetic algorithm optimized ANN (GA-ANN), PSO-support vector machine model (PSO-SVR), genetic algorithm optimized-SVR model (GA-SVR) and the SVR model optimized with grid search (GS-SVR). A comprehensive evaluation of the SaDE-ELM model is performed, considering key statistical metrics and diagnostic plots of measured and forecasted *GSR*. The results demonstrate excellent forecasting capability of the SaDE-ELM model in respect to the nine benchmark models. SaDE-ELM outperformed all comparative models for both tested study sites with a relative mean absolute and a root mean square error (*RRMSE*) of 2.6% and 2.3% (for Brisbane) and 0.8% and 0.7% (for Townsville), respectively. Majority of the forecasted errors are recorded in the lowest magnitude frequency band, to demonstrate the preciseness of the SaDE-ELM model. When tested for daily solar radiation forecasting using the ECMWF Reanalysis data for Brisbane study site, the performance resulted in an *RRMSE* ≈ 10.5%. Alternative models evaluated with the input data classified into El Niño, La Niña and the positive and negative phases of the Indian Ocean Dipole moment (considering the impacts of synoptic-scale climate phenomenon), confirms the superiority of the SaDE-ELM model (with *RRMSE* ≤ 13%). A seasonal analysis of all developed models depicts SaDE-ELM as the preferred tool over the basic ELM and the hybridized version of ANN, SVR and GP model. In accordance with the results obtained through MODIS satellite and ECMWF Reanalysis data products, this study ascertains that the proposed SaDE-ELM model applied with ACO feature selection, integrated with satellite-derived data is adoptable as a qualified tool for monthly and daily *GSR* predictions and long-term solar energy feasibility study especially in data sparse and regional sites where a satellite footprint can be identified.

1. Introduction

The prediction of global solar radiation (*GSR*) plays a very crucial role in real world since solar energy continues to be an increasingly important and a credible alternative to fossil fuel. There is growing unanimity driven by healthy arguments for the adoption of solar resources as an alternative for carbon-based fuels, not only from a global perspective, but also in an Australian context where solar energy has an immense potential due to the high solar insolation, generally low rainfall and a relatively smaller fraction of clouds due to a sub-tropical

* Corresponding authors.
E-mail addresses: u1081587@umail.usq.edu.au (S. Ghimire), ravinesh.deo@usq.edu.au (R.C. Deo).

https://doi.org/10.1016/j.rse.2018.05.003
Received 14 November 2017; Received in revised form 27 April 2018; Accepted 1 May 2018

location (Beath, 2012; Yusaf et al., 2011). A report published by the Australian photovoltaic institute mentioned that rooftop-based solar capacity has now reached about 5.6 GW with a large-scale solar capacity of 496 MW. These totals are able to meet about 3.3% of the total Australian electricity demand, in which the Sunshine state, Queensland is leading the rooftop solar energy supply, with a generation capacity of about 1.72 GW (Parkinson, 2017a). The report also emphasizes that by 2040, the amount of solar capacity will have risen by about 10-fold from its current level. Besides this, the Australian Government is constantly supporting further opportunities for scientific research in renewable energies (Haidar et al., 2015), firstly, to model the solar prospectivity in Australia's diverse spatial locations (i.e., metro and remote) with acceptable accuracy by high-performance models and secondly, to harness this energy where it is economically sustainable. Furthermore, *GSR* is stochastic and it is influenced by factors such as the absorption and scattering by clouds, and the particulate matter, including dust and pollutants that vary on this time scales from minutes to years (ARENA, 2015). Therefore, *GSR* forecasting has a key role in successfully managing the impacts on electricity demand/supply market, renewable energy generation and integration technologies to make them economically viable options. In light of this need, this paper designs and presents a new approach to model solar energy remotely, using MODIS satellite data for the forecasting of long-term (i.e., monthly averaged daily) global solar radiation for sub-tropical, solar rich cities (Brisbane and Townsville) located in Queensland, a state that support large local populations and the consumer energy demand continues to rise.

The prediction of *GSR* is typically executed by employing one of the three (i.e., purely statistical, standalone machine learning evolutionary process or statistical learning tool and hybridized machine learning) algorithms (Hussain and Al-Alili, 2016). The first approach is based on statistical relationships between inputs and target (i.e., GSR) and may include models such as autoregressive integrated moving average model (ARIMA), coupled auto-regressive and dynamical system (CARDS) model or support vector machine regression (SVR) model (Antonanzas et al., 2015a). The second approach involves biologically inspired machine learning (ML) models such as the particle swarm optimization (PSO), neuro-fuzzy, genetic programming (GP) regression and the artificial neural networks model (ANN) (Voyant et al., 2017). These ML methods have been progressively developed very rapidly in recent years, providing varied methodologies and accuracies under different contexts (Guo et al., 2014). Several ML models have been successfully applied to predict *GSR*, including ANNs (Aaboud et al., 2017; Bae et al., 2017; Baser and Demirhan, 2017; Chiang et al., 2017; da Silva et al., 2017; Dennis et al., 2017; Kim et al., 2017; Ma et al., 2017; Quej et al., 2017; Salcedo-Sanz et al., 2017; Sun and Zhang, 2017; van der Ven and Bongers, 2017; Voyant et al., 2017; Wang et al., 2017; Zou et al., 2017), PSO (Cheng et al., 2016; Gong et al., 2017; Kumar et al., 2017; Mansour et al., 2017; Rai, 2017; Ray et al., 2017), GA, GP (Asrari et al., 2017; Sameti et al., 2017; Saxena and Kumar, 2016; Viegas et al., 2016; Yuce et al., 2016) and the adaptive neuro-fuzzy inference systems (ANFIS) (Jha and Srivastava, 2016; Jović et al., 2016; Mohammadi et al., 2016; Mohan et al., 2017; Quej et al., 2017; Ruiz et al., 2016; Salisu et al., 2017). Comparison shows that their accuracies are generally better than empirical models (Ssekulima et al., 2016). The third approach is the hybrid based model that employs a combination of statistical and biological methods such as the PSO-ANN, GA-SVR, GA-ANN, PSO-SVR etc. Among these, machine learning approaches have specifically been well suited for *GSR* prediction due to their robustness, high-performance, and the capacity to be crossbred with alternative prediction methods (Voyant et al., 2017) for optimal accuracy. Many of these ML approaches include different inputs based on meteorological and geographical parameters such as the latitude, longitude, temperature, wind speed and direction, sunshine duration, precipitation, cloud cover, etcetera (Antonanzas et al., 2015b; Aybar-Ruiz et al., 2016; Hussain and Al-Alili, 2016; Mellit and Kalogirou, 2008; Olatomiwa et al., 2015). According to Belu (2014), large and medium scale solar energy facilities require long–term meteorological input datasets for the efficient energy system design and the control, operation and long-term sustainability of solar energy power plant investments. Often, solar power plants need to be located in remote regions to support regional communities residing in solar-rich dwellings but the requirement of sufficient data to model and justify the capital investment remains a major issue. Despite this requirement, very few studies have investigated the potential of freely available, remotely-sensed satellite data used for energy feasibility studies and the prediction of *GSR* in such locations prior to any major capital investments.

ANN has been used for *GSR* prediction (Marzo et al., 2017) involving different meteorological and geographical parameters as the predictors. Deo and Sahin (2017) employed a single satellite input parameters obtained from the Earth Orbiting System - Moderate Resolution Imaging Spectroradiometer (EOS-MODIS), the land-surface temperature (LST), to train an ANN model, with results showing that an ANN model outperformed both the multiple linear regression (MLR) and the ARIMA model. Other ML based regression approaches have successfully been employed for *GSR* prediction. Ramedani et al. (2014) employed a support vector regression (SVR) technique to develop a model for prediction of *GSR* in Tehran. In Olatomiwa et al. (2015), a support vector machine model coupled with the firefly algorithm (SVM-FFA) was used to predict monthly mean horizontal *GSR*. Similarly, Gala et al. (2016) proposed a hybrid ML method using an SVR, employing the gradient boosted regression (GBR) and the random forest regression (RFR) to improve the initial radiation forecasts provided by the state-of-the-art European Center for Medium Range Weather Forecasting (ECMWF) model. Importantly, the ECMWF datasets applied to the SVR method led to enhanced accuracy of the solar radiation forecasts. In addition to this, several hybrid models such as the GA combined with a multi-model framework (Wu et al., 2014), combined hidden Markov models and the generalized fuzzy model (Bao et al., 2013; Bhardwaj et al., 2013), Coral Reef Optimization–Extreme Learning Machine (CRO–ELM) (Kai-Min et al., 2002; Salcedo-Sanz et al., 2017), hybrid SVR-Wavelet model (Mohammadi et al., 2015), combined self-organizing maps (SOM), SVR and PSO (Dong et al., 2015), and a modified ANN model known as a non-linear autoregressive recurrent exogenous neural network (NARX-NN) using recursive filtering (Hussain and Al-Alili, 2016) has been used to predict the *GSR* data.

Despite its usage, major drawbacks of ANN model include the requirement for iterative tuning of its model parameters, slow prediction speed and a generally low performance compared to some of the other ML models (Acharya et al., 2014; Deo et al., 2017). To overcome this issue, the extreme learning machine (ELM) algorithm was developed by Huang (Huang et al., 2004) in a way that the ELM model attained a faster speed by random assignment of hidden neurons without compromising its final accuracy (Shamshirband et al., 2016). In recent years, the ELM model has been applied in various scientific contexts (Wang et al., 2014; Wang and Han, 2014) including a number of notable studies on *GSR* prediction (Salcedo-Sanz et al., 2014a; Salcedo-Sanz et al., 2018; Salcedo-Sanz et al., 2017). In particular, the study of Şahin et al. (2014) applied an ELM approach to predict *GSR* using satellite measurements. Some studies have also shown better performance of ELM in respect to an ANN model developed with classical training approaches (Alharbi, 2013) and an SVR model used for *GSR* prediction (Salcedo-Sanz et al., 2014b). Considering the merits demonstrated in many studies performed elsewhere in different contexts, the present utilization of ELM model integrated with remotely-sensed input data for *GSR* prediction is an interesting research endeavor particularly for solar-rich cities in Australia.

The purpose of this paper is fourfold: (1) to design a new approach for *GSR* prediction by applying self-adaptive differential evolutionary ELM (SaDE-ELM) hybridized with ant colony optimization (Shunmugapriya and Kanmani, 2017) as the *GSR* feature selection algorithm from remotely sensed MODIS satellite and ECMWF Reanalysis