Applied Energy 209 (2018) 79–94

Contents lists available at ScienceDirect

Applied Energy

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

An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia

Sancho Salcedo-Sanz[a], Ravinesh C. Deo[b,*], Laura Cornejo-Bueno[a], Carlos Camacho-Gómez[a], Sujan Ghimire[c]

[a] *Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain*
[b] *School of Agricultural, Computational and Environmental Sciences, International Center for Applied Climate Sciences, Institute of Agriculture and Environment, University of Southern Queensland, Springfield Campus, QLD 4300, Australia*
[c] *School of Agricultural, Computational and Environmental Sciences, Institute of Agriculture and Environment, University of Southern Queensland, Springfield Campus, QLD 4300, Australia*

HIGHLIGHTS

- Solar radiation (R_n) is predicted via neuro-evolutionary approach in Sunshine Coast, Australia.
- Coral Reefs Optimization, Grouping Genetic Algorithm and Extreme Learning Machines are integrated.
- CRO-hybrid and GGA-hybrid mechanism precision is assessed thoroughly.
- ELM is evaluated against MARS, SVR and MLR algorithm.
- CRO-(ELM)-ELM is an appealing approach to simulate daily R_n.

ARTICLE INFO

Keywords:
Solar radiation estimation
Coral Reefs Optimization
Extreme Learning Machines
Grouping Genetic Algorithms
Hybrid modelling system

ABSTRACT

This research paper aims to develop a hybrid neuro-evolutionary wrapper-based model for daily global solar radiation estimation in the solar-rich Sunshine State of Queensland, Australia. To design a robust hybrid modelling mechanism, the Interim-ERA European Centre for Medium Range Weather Forecasting (ECMWF) Reanalysis data are employed to train and cross-validate the estimation model that is formulated by an evolutionary-type algorithm: the Coral Reefs Optimization (CRO) integrated with an Extreme Learning Machine (ELM) model. The hybrid CRO-(ELM) algorithm is applied in two stages: first for the feature selection process guided by an ELM algorithm (a class of fast training neural network tool) as the model's fitness function to screen an optimal set of predictor variables and second, for the estimation of the solar radiation using the optimally screened variables by the final hybrid CRO-(ELM)-ELM system. To benchmark the performance of the hybrid CRO-ELM algorithm for this estimation problem we apply an alternative, yet a similar feature screening approach: the Grouping Genetic Algorithm encoded into the ELM-based model (GGA-(ELM) used as the predictor mechanism). After the feature selection process is performed by the CRO algorithm that extracts the data patterns for accurate estimation the alternative objective algorithm is applied (in this case the ELM again) to formulate the hybrid CRO-(ELM)-ELM modelling system. To provide a rigorous evaluation of the CRO-(ELM)-ELM hybrid system, alternative estimation approaches are considered: the Multivariate Adaptive Regression Splines (MARS), Multiple Linear Regression (MLR) and the Support Vector Regression (SVR). The hybrid CRO-(ELM)-ELM system is tested in a real problem where the results are evaluated by means of several statistical score metrics and diagnostic plots of the observed and the estimated daily global solar radiation in the testing phase. We show that the hybrid CRO-(ELM)-ELM model is able to yield promising results; thus improving those attained by the 7 alternative models (i.e., hybrid CRO-(ELM)-MARS, CRO-(ELM)-MLR and CRO-(ELM)-SVR and the GGA equivalent models). The study ascertains that the CRO-based hybrid system where a large pool of predictor data are carefully screened through a wrapper-based modelling system and the ELM model is applied as a objective estimation tool can be adopted as a qualified stratagem in solar radiation estimation problems. The proposed hybrid CRO-(ELM)-ELM system exhibits clear advantages over the alternative machine learning approaches

* Corresponding author.
E-mail address: ravinesh.deo@usq.edu.au (R.C. Deo).

http://dx.doi.org/10.1016/j.apenergy.2017.10.076
Received 21 July 2017; Received in revised form 7 October 2017; Accepted 23 October 2017

tested and also over the other machine learning algorithms used without the feature selection tool; thus advocating its scientific utility in renewable energy applications.

1. Introduction

Australia is a leading country with the highest level of the annual global solar radiation [1]. In spite of this, solar energy only accounts for only 3% of the total power consumed in the country, which remains behind wind energy (5%) and hydro-power (7%) [2]. Fossil fuels are still the main source of generating electricity in Australia for over 80% of the total generated power used by different consumers. Taking into account that the pollution from fossil fuel power stations is considered as the main contributor to global warming at global scale, the necessity of supporting renewables as a source of the consumable energy and ensuring their availability through the forecasting of energy potentials is a primary objective in energy policy development [3]. In terms of renewable resources, solar power should be the first renewable source in Australia by far, but the most important solar energy resource corresponds to zones of the country without connection with the national electricity grid [1]. The current government's energy policy (particularly in the sunshine State of Queensland, Australia) plans to support the development of solar power with massive investments that are planned in the deployment of large-scale solar energy facilities. It is estimated that by 2030, the percentage of energy consumption in Australia emanating from solar resources is likely to reach 30% [2].

The current boom in solar energy utilization is culminating new engineering problems; some of them being associated with a key question of how to integrate this renewable resource into real grid systems. Solar energy is intrinsically stochastic by its very nature, like some of the other renewable resources. The energy availability is affected by the presence of cloud cover and atmospheric attenuations mechanisms driven by atmospheric particles, dusts, ozone, etc. that act to moderate the surface level solar radiation. The estimation of solar radiation (by means of accurate and reliable models) where the physical interactions between solar radiation's variability and the upper atmospheric variables are incorporated into such models for predicting solar-driven power is thus a key task for integration of the energy into a power grid system. In the case of future investments for large-scale energy facilities, Global Solar Radiation (*GSR*) estimation received at a site of a solar plant is an extremely important factor and this energy completely depends on atmospheric variables [4]. Solar radiation estimation is, therefore, a major problem in the proper management of large-scale solar facilities, which has been tackled with very different methods [5–7]. Among these, machine learning (or data-intelligent) approaches have specifically been well-suited due to their robustness, high-performance and the capacity to be hybridized with numerical estimation methods [8]. Current machine learning techniques consider a number of input variables based on meteorological and geographical parameters (e.g., sunshine duration hours, land surface and air temperature, relative humidity, wind speed, wind direction, cloud cover and precipitation) [9,10]. According to Belu [11], the design, control and operation of large and medium-scale solar energy facilities requires long-term time-series of meteorological data to be fed into the primary estimation algorithm. Therefore, the utilization of several atmospheric variables for energy estimation remains a pivotal tasks in energy optimization problems.

A brief review of machine learning approaches used for solar energy estimation problems is now carried out. Well-established technique of Artificial Neural Network (ANN) has been the most popular method for energy estimation. In the study of Yadav and Chandel [12], an exhaustive review on solar radiation estimation using ANNs were presented in which different case studies that utilized atmospheric input parameters, were described. Photovoltaic (PV)-based applications were studied by Mellit and Kalogirou [13] where a number of artificial intelligence techniques (ANN, Fuzzy Logic, Genetic Algorithm, Expert Systems, etc.) were utilized to validate their utility in the sizing of PV power systems, power simulations, control of PV systems and the estimation of PV-based power using atmospheric and meteorological datasets. In the study of Sozen et al. [14], the ANN model was constructed with geographic parameters to estimate the solar energy potential in Turkey whereas the research conducted by Behrang et al. [15] incorporated different combinations of input variables into Multi-Layer Perceptron (MLP) and Radial Basis Function Neural Network ((RBF-NN) models. Dorvlo et al. [16] compared MLP and RBF-NN models and Bou [17] assessed the performance of ANN models in Kuwait whereas in Alsina et al. [18] investigated the performance of ANN for estimation of monthly daily solar radiation. In that work, the estimation was carried out by considering 45 measuring stations in Italy. Recently, the application of an Extreme Learning Machine (ELM) model, which is considered to be a fast and an efficient data-intelligent model in comparison to the ANN model, has been explored mainly due to the high accuracy attained in a number of problems. For example, the study of Sahin et al. [19] utilized ELM integrated with satellite dataset to conclude that ELM outperformed ANN model in terms of the estimation accuracy and computational workload. Other than this, the ELM model's estimation performance has been described in recently published works [20–25]. However, the application of the ELM model for solar radiation estimation in the present study region (Sunshine State of Australia) is yet to be explored although a recent study has validated the ELM model for global and diffuse solar ultraviolet index estimation [26].

Apart from the ELM model, a number of other machine learning algorithms have successfully been applied in solar radiation estimation. Lou et al. [27] designed a data-intelligent model with boosted regression tree to estimate diffuse solar radiation in Denver and Hong-Kong and Mohammadi et al. [28] used Support Vector Regression (SVR) integrated with Wavelet Transform (WT) to improve the performance of the former tool. The research performed by Olatomiwa et al. [29] combined Firefly Meta-heuristic Algorithm with SVR to validate its suitability for solar radiation estimation in Nigeria and the study of Antonanzas et al. [30] tackled a solar irradiation mapping problem with an SVR model using exogenous predictor data, while Monteiro et al. [31] compared the performance of SVR and ANN for photovoltaic power generation. Belaid et al. [32] tackled a problem of solar radiation estimation using SVR, considering different forecast horizons and Chen et al. [33–35] applied least-square SVR using atmospheric data. In the study of Yacef et al. [36], a Bayesian Network in comparison with ANN models was applied whereas a new approach based on temporal Gaussian Processes (GP) was presented by Salcedo-Sanz et al. [37]. In the latter the GP model was seen to improve alternative regressors (e.g., ANN, SVR and regression trees). Another approach for solar radiation estimation used 'all-sky' conditions and satellite images. For example, Fu and Cheng [38] estimated solar radiation with features extracted from all-sky image, cloud pixels, frame differences, gradient magnitude, intensity levels, accumulated intensity along the vertical line of the Sun and the number of corners in a cloud image. A recent study [39,40] also tackled the problem of solar radiation estimation in the sunshine State of Queensland, Australia using satellite images and an ANN model although their study did not employ a specific feature selection algorithm, as it has been attempted in the present study.

In order to attain accurate estimation of *GSR* from a large predictor set from atmospheric and ground-based sources that has been carefully screened, hybrid modelling systems (i.e. systems that merge two or