


Adaptive Neuro-Fuzzy Inference System integrated with solar zenith angle for forecasting sub-tropical Photosynthetically Active Radiation

Ravinesh C. Deo¹  | Nathan J. Downs¹ | Jan F. Adamowski² | Alfio V. Parisi¹

¹Centre for Applied Climate Sciences, University of Southern Queensland, Springfield, Queensland, Australia

²Department of Bioresource Engineering, Faculty of Agricultural and Environmental Science, McGill University, Sainte Anne de Bellevue, Québec, Canada

Correspondence

Ravinesh C. Deo, Centre for Applied Climate Sciences, University of Southern Queensland, QLD, Australia.
Email: ravinesh.deo@usq.edu.au

Funding information

Academic Development and Outside Studies Program (ADOSP), USQ Centre for Applied Climate Sciences

Abstract

Advocacy for climate mitigation aims to minimize the use of fossil fuel and to support clean energy adaptation. While alternative energies (e.g., biofuels) extracted from feedstock (e.g., micro-algae) represent a promising role, their production requires reliably modeled photosynthetically active radiation (PAR). PAR models predict energy parameters (e.g., algal carbon fixation) to aid in decision-making at PAR sites. Here, we model very short-term (5-min scale), sub-tropical region's PAR with an Adaptive Neuro-Fuzzy Inference System model with a Centroid-Mean (ANFIS-CM) trained with a non-climate input (i.e., only the solar angle, θ_Z). Accuracy is benchmarked against genetic programming (GP), M5Tree, Random Forest (RF), and multiple linear regression (MLR). ANFIS-CM integrates fuzzy and neural network algorithms, whereas GP adopts an evolutionary approach, M5Tree employs binary decision, RF employs a bootstrapped ensemble, and MLR uses statistical tools to link PAR with θ_Z . To design the ANFIS-CM model, 5-min θ_Z (01–31 December 2012; 0500H–1900H) for sub-tropical, Toowoomba are utilized to extract predictive features, and the testing accuracy (i.e., differences between measurements and forecasts) is evaluated with correlation (r), root-mean-square error ($RMSE$), mean absolute error (MAE), Willmott (WI), Nash–Sutcliffe (E_{NS}), and Legates & McCabes (E_{LM}) Index. ANFIS-CM and GP are equivalent for 5-min forecasts, yielding the lowest $RMSE$ (233.45 and 233.01 $\mu\text{mol m}^{-2}\text{s}^{-1}$) and MAE (186.59 and 186.23 $\mu\text{mol m}^{-2}\text{s}^{-1}$). In contrast, MLR, M5Tree, and RF yields higher $RMSE$ and MAE [$RMSE = 322.25 \mu\text{mol m}^{-2}\text{s}^{-1}$, $MAE = 275.32 \mu\text{mol m}^{-2}\text{s}^{-1}$], [$RMSE = 287.70 \mu\text{mol m}^{-2}\text{s}^{-1}$, $MAE = 234.78 \mu\text{mol m}^{-2}\text{s}^{-1}$], and [$RMSE = 359.91 \mu\text{mol m}^{-2}\text{s}^{-1}$, $MAE = 324.52 \mu\text{mol m}^{-2}\text{s}^{-1}$]. Based on normalized error, ANFIS-CM is considerably superior ($MAE = 17.18\%$ versus 19.78%, 34.37%, 26.39%, and 30.60% for GP, MLR, M5Tree, and RF models, respectively). For hourly forecasts, ANFIS-CM outperforms all other methods ($WI = 0.964$ vs. 0.942, 0.955, 0.933 & 0.893, and $E_{LM} = 0.741$ versus 0.701, 0.728, 0.619 & 0.490 for GP, MLR, M5Tree, and RF, respectively). Descriptive errors support the versatile predictive skills of the ANFIS-CM model and its role in real-time prediction of the photosynthetic-active energy to

explore biofuel generation from micro-algae, studying food chains, and supporting agricultural precision.

KEYWORDS

ANFIS-CM, genetic programming, M5 Tree, multiple linear regression, photosynthetically active radiation, Random Forest, real-time forecasting

1 | BACKGROUND REVIEW

Besides rare extremophile chemoautotrophs, the survival of life on the earth is dependent upon incident solar radiation. Scattering and absorption by atmospheric constituents, both natural and anthropogenic in nature, can alter the spectral distribution of solar radiation reaching the Earth. Essentially, the short wavelengths within the high energy ultraviolet (UV) spectrum ($290 \text{ nm} \leq \lambda_{\text{UV}} \leq 400 \text{ nm}$) are preferentially absorbed by ozone, while low energy, optically transparent visible, and infrared wavelengths (Figure 1) are passed through.

Food chains are driven by the solar energy component termed as the photosynthetically active radiation (PAR; $400 \text{ nm} \leq \lambda_{\text{PAR}} \leq 700 \text{ nm}$), which is therefore critical to biosphere functions. Microcellular organisms draw on PAR for energy to develop organic constituents that are considered the building blocks of marine and terrestrial food webs. These, in turn, regulate and sustain biomass production systems. Accordingly, by regulating the natural carbon cycle,

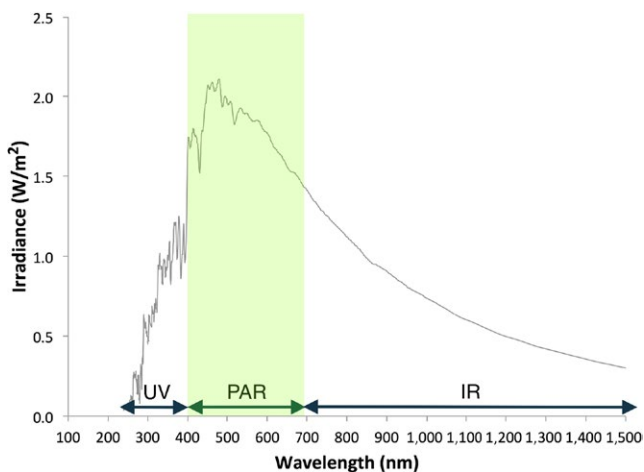


FIGURE 1 Solar spectrum before absorption by the atmosphere, as measured by the Spectral Irradiance Monitor (SIM) on-board the Solar Radiation and Climate (SOURCE) satellite [1], indicating the solar ultraviolet (UV) and solar infra-red (IR) bands. Note that the photosynthetically active radiation (PAR) lying in-between the spectrum is affected by the atmospheric absorption at near normal incidence to the Earth's surface at the solar noon

PAR availability can influence crop yield (Monteith & Moss, 1977) and CO_2 , water, and energy exchanges in the atmospheric continuum of plant-atmospheric systems (Fernández-Martínez et al., 2014; Hetherington & Woodward, 2003; Muir, Wallace, Done, & Aguirre, 2015). Near real-time PAR forecasts are a key parameter for renewable energy production (Eltbaakh et al., 2011; Jacovides, Tymvios, Boland, & Tsitouri, 2015), including global biomass systems that generate approximately 46 EJ of bioenergy per year. These provide food, fodder, fiber, hydrocarbons for heating, electrical power, liquid fuels, and other chemicals (IPCC 2007). Consequently, PAR is considered crucial for supporting the demand of sustainable and environmentally friendly energy production solutions.

Energy production by means of solar photovoltaics and algae biofuels is impacted by the feasibility of PAR energy “harvesting” sites. Biofuels, as a renewable energy driver, have the potential to benefit the current and future growth of cleaner productions. According to the World Economic Forum (WEF), the conversion of biomass into energy and other chemical bio-products can contribute US\$295 billion per year (WEF 2010) to the global economy. By 2014, global biofuel production had expanded from less than $30 \times 10^9 \text{ L}$ per year in the previous decade to $120 \times 10^9 \text{ L}$ per year. Therefore, this sector is expected to reach US\$1.13 trillion per year by 2022 (REN 2016). By 2020, attendant revenues are forecasted to attain US\$80 billion per year, while bio-products such as plastics and chemicals may reach a value of US\$15 billion per year. Considering the rapid expansion of the biofuels market since the 2010 WEF report, the optimization of PAR availability mapping can play an important role in addressing the cost-effectiveness of energy production and scientific and policy-driven incentives for investments in renewable energies and their supply chain (Akgul, Shah, & Papageorgiou, 2012; Brennan & Owende, 2010; Dupraz et al., 2011; Simionato, Basso, Giacometti, & Morosinotto, 2013). Over 60 nations have instigated mandates and targets for inclusion of biofuels in the development of renewable energies (Lane, 2016). However, the use of land for biofuel production continues to place pressures on available land resources since biomass production competes for land resources dedicated to sustaining global food crop production (Rajagopal, Sexton, Roland-Holst, & Zilberman, 2007). Consequently, the identification of alternative solar-rich sites for energy production

remains a challenging task for climate scientists, energy engineers, and climate-energy policymakers.

Globally, Brazil and the United States have directed significant research and investment efforts toward industrial biotechnology, particularly in biofuels (Kumar, Shrestha, & Salam, 2013; WEC 2017; WEF 2017). In addition, many European nations, traditionally associated with low levels of sunshine, have established national action plans that introduce biofuels and mandatory renewable energies (EC 2017; EIC 2017). These include implementation of tax relief policies, loans, subsidies, and research grants to encourage greater co-investment in the development of bio-products (Rajagopal et al., 2007; WEF 2010). Furthermore, the state of Queensland (QLD), Australia, presents an ideal combination of sub-tropical and tropical climates with a mature and modern agricultural industry and supply chain system (TIQ 2017). Therefore, consistent with their 10-year roadmap, the Australian government has allocated \$Aus1.0 billion to satisfy several bio-futures projects (State of Queensland, 2016). The need to expand research on biomass energy systems and develop accurate and reliable models as mapping tools for potential sites is a key motivation of the present study.

The identification of land surface networks that provide dedicated PAR can assist in the optimization and efficiency of industries; however, such networks are limited due to the lack of availability of meteorological data at surface monitoring sites at a global scale. To date, most PAR modeling has been undertaken at specific meteorological sites, where measurements of global (total) solar radiation have been used to develop radiative transfer equations (Qin, Yang, Liang, & Tang, 2012; Wang et al., 2015). Neglecting the influence of cloud cover, approximately 40%–50% of global solar radiation that is considered PAR, can be represented through models that include daily sunshine duration and global solar radiation data (Guofeng, Leeuw, Skidmore, Yaolin, & Prins, 2010; Sudhakar, Srivastava, Satpathy, & Premalatha, 2013). To address limitations in PAR surface predictions, alternative approaches incorporate atmospheric boundary conditions (where available), including air density, vapor, cloud fraction and optical depth, and particulate and aerosol concentrations. Such parametric models have the flexibility to be customized to different locations given a set of known atmospheric conditions (Clough et al., 2005); however, they do present significant challenges in terms of complexity and adaptability to diverse sites for bio-investment. Moreover, these parametric models are subject to logistic constraints such as collection of meteorological data difficulties over complex terrains where PAR potentials are high.

Meteorologically derived PAR estimations, achieved through radiative transfer (i.e., deterministic) approaches, pose challenges given their imprecise knowledge of conditions and complexity of differential equations required to simulate them. These constraints restrict the ability to generate

spatial predictions of PAR for locations earmarked for PAR harvesting (Fu & Rich, 2002). For instance, the attenuation of solar irradiance is linked to spatiotemporally variable atmospheric absorption and scattering processes. Semi-empirical models are computationally efficient at estimating surface PAR but can be significantly disturbed by local climatic and geographic conditions (Parisi, Wong, & Randall, 1998; Wang, Kisi, Zounemat-Kermani, Hu, & Gong, 2016; Wang et al., 2015). As a result, deterministic methods require recalibration to characteristic local conditions (e.g., cloudiness, vapor, and aerosols) (Xia et al., 2008). Considering the difficulties faced by semi-empirical and radiative models, versatile and inexpensive data-driven models are beneficial as they have less extensive requirements, can accommodate urban and remote sites, and overcome other limitations posed by the paucity of experimental facilities (Jacovides et al., 2015).

Satellite (and reanalysis) data sources can fill potential gaps in analysis of PAR availability for sites where insufficient locally measured atmospheric parameters are available (Deo & Sahin, 2017; Ghimire, Deo, Downs, & Raj, 2018; Rubio, Lopez, Tovar, Pozo, & Batlles, 2005). However, satellite swaths at daily (or longer) timescales do not offer the benefit of near real-time (i.e., very short-term) PAR estimations, considered vital for accurately describing a plant's photosynthetic behavior at a micro-level (Escobedo, Gomes, Oliveira, & Soares, 2011; Ge, Smith, Jacovides, Kramer, & Carruthers, 2011; Wang et al., 2015). Data-driven approaches using input datasets (initial conditions and mathematical equations are not required as with empirical and physical models) can be trained and adapted to variable climates for modeling PAR at the local scale. As a result, data-driven models have recently emerged as pertinent tools in energy systems (e.g., (Deo & Sahin, 2017; Deo, Downs, Parisi, Adamowski, & Quilty, 2017; Deo, Wen, & Feng, 2016; Ghimire et al., 2018; Hu, Wang, & Zeng, 2013)). Such models offer added opportunity to forecast future PAR on very short- and long-term scales, without needing extensive atmospheric model initialization conditions. Yet, even though much research has been performed for solar, wind, and atmospheric modeling applications (e.g., (Deo & Sahin, 2017; Deo et al., 2016; Dokur, Kurban, & Ceyhan, 2015; Ghorbani, Khatibi, Hosseini, & Bilgili, 2013; Islam, Mohandes, & Rehman, 2016; Mohammadi et al., 2015; Salcedo-Sanz, Deo, Cornejo-Bueno, Camacho-Gómez, & Ghimire, 2018)), modeling of very short-term (i.e., near real-time) PAR in sub-tropical regions remains scarce. These regions have less cloud cover than tropical or temperate climate areas and therefore enhance the suitability of sites for solar and bio-energy generation.

In Spain, Lopez, Rubio, Martinez, and Batlles (2001) applied artificial neural network (ANN) modeling to estimate PAR, utilizing sunshine durations and global solar data recorded at radiometric stations at separate occasions. Similarly, with extensive inputs (e.g., global ultraviolet,

photosynthetic photon flux density (PPFD), broadband and extraterrestrial flux, temperature, humidity, sunshine, precipitable water, and ozone), Jacovides et al. (2015) used an ANN model to estimate daily solar global UV radiation, PAR, and broadband fluxes at a Mediterranean site. Their comparison between ANN and regression models determined the former to be more promising tools. On the other hand, Wang et al. (2016) estimated hourly PAR in various ecosystems in China using ANN-based modeling approaches (e.g., multilayer perceptron, generalized regression neural network, and radial basis neural network). The model employed meteorological input data such as atmospheric temperature, humidity, dewpoint, water vapor, and air pressure. Yu and Guo (2016) modeled hourly PAR data in the Midwestern United States with ANN and regression models. However, although they provide decision-makers instantaneous predictions to broaden our understanding of photosynthesis at micro-levels, no study has estimated PAR at near real-time forecasting periods (i.e., very short-term intervals).

Given the ANN modeling's limitations (e.g., iterative tuning of parameters, slow convergence rates, local maxima or minima issues, and inability to attain a global solution), advances can take place to accurately estimate PAR by testing other machine learning approaches with unique data extraction features. This can address the shortfalls of standalone ANN (or other equivalent) methods. A robust alternative tool for PAR estimation is the Adaptive Neuro-Fuzzy Inference System (ANFIS). This combines fuzzy set and ANN theory, offering advantages over a standalone ANN (Ali, Deo, Downs, & Maraseni, 2018; Yaseen et al., 2018). First utilized by Jang (1995), ANFIS consists of a hybrid learning system that processes data features *via* a set of rules and membership functions (MF). Principally, the ANFIS model has 5 layers where neuro-fuzzy modeling adapts learning techniques from neural networks to the fuzzy inference system (FIS) (Zadeh, 1965). The structure of FIS consists of a rule base and a database that defines MFs, including reasons to perform the inferences based on rules selected to derive an output (Ali et al., 2018). By combining the two, modelers can address the difficulties and limitations in standalone ANN and FIS paradigms (i.e., the weaknesses of FIS can be compensated for by the complementary capacities of neural networks (Vieira, Dias, & Mota, 2004)). Due to its theoretical basis and improved accuracy, the ANFIS model has been adopted in solar energy and drought studies (Ali et al., 2018; Chaabene & Ammar, 2008; Khatib, Mohamed, & Sopian, 2012; Mellit, Arab, Khorissi, & Salhi, 2007; Mellit & Kalogirou, 2011; Şen, 1998). However, to the best of the authors' knowledge, the application of ANFIS for generating PAR estimates in solar-rich regions (i.e., tropical or sub-tropical) remains unexplored.

In this study, ANFIS C-Means Clustering (ANFIS-CM) was utilized to model PAR in the state of Queensland (QLD), a

region that has an abundance of solar resources and is demonstrating an upward trend in bio-future investments. The state government's energy policies and Bio-futures Roadmap will encourage the development of technologies to harness solar and biomass energy (State of Queensland, 2016), backed by a 10-year action plan (State of Queensland, 2016) to invest \$Aus1 billion in export-oriented industrial biotechnology and high-value knowledge-intensive jobs.

QLD is located in a sub-tropical belt with less cloud cover compared to other tropical regions. It is an adequate location for year-round production of high-quality feedstock and biomass. The production of energy-dense biomass and productive feeds is needed for biofuels. Production of energy-rich crops such as sugarcane (*Saccharum officinarum* L.), eucalyptus (*Eucalyptus* sp. L'Hér), algae, sorghum [*Sorghum bicolor* (L.) Moench], native grasses, crop stubble, cassava (*Manihot esculenta* Crantz), agave (*Agave* sp. L.), and pongamia [*Millettia pinnata* (L.) Panigrahi] (Bahadori & Nwaoha, 2013; Martin & Rice, 2012; QREP 2009; State of Queensland, 2016) require accurate modeling of PAR.

The present study is valuable due to the lack of dedicated monitoring stations in the Southern Hemisphere, which restrict the ability individual's ability to make surface level PAR predictions. Much of this sub-tropical region (11°S–30°S lat.) benefits from less daily cloud cover and a greater abundance of dry, clean air, making it ideal for harnessing freely available solar radiation for either solar or biofuel energy (Shafiullah, 2016) production.

In accordance with the necessity to model PAR in QLD, this paper has three aims:

- 1 to model very short-term (i.e., 5 min) and hourly PAR, in a sub-tropical latitude region for the first time, using ANFIS-CM,
- 2 to explore the accuracy of ANFIS-CM using limited input data comprised of solar zenith angle (θ_2) and measured photosynthetic photoflux density (PPFD) during the austral summer solstice period, and
- 3 to compare the ANFIS-CM model's performance to other types of data-driven models (i.e., genetic programming, multiple linear regression, M5 Model Tree, and Random Forest) following previous studies (e.g., (Liu & Jordan, 1960; Sudhakar et al., 2013)).

In developing these models, measured data were obtained from University of Southern Queensland, Toowoomba, Australia (27.5°S, 151°E), that has been gathered since 2007 as part of the university's Atmospheric and Ultraviolet Monitoring Program, initiated in 2003. Testing phase model performances, including the ANFIS-CM model employing limited input data (i.e., solar zenith angle alone), were evaluated and compared for their ability to estimate PAR at multiple timescales. In addition, visual assessment tools (e.g., scatter plots of observed

vs. predicted PAR) and statistical model accuracy metrics were utilized.

2 | THEORETICAL BACKGROUND

2.1 | Photosynthetic photon flux density and the solar zenith angle

PAR energy absorbed by photosynthetic organisms fuels the production of organic compounds necessary to sustain the development of green biomass. It is reported using the radiometric unit of irradiance (W m^{-2}). In practice, surface PAR is quoted in terms of photosynthetic photon flux density (PPFD) rather than the radiometric equivalent. PPFD represents the number of visible light photons ($400 \text{ nm} \leq \lambda \leq 700 \text{ nm}$), from the solar spectrum, incident upon a unit of area in a unit of time, absorbed by a plant's photosynthetic apparatus (Figure 1). PPFD is a measure of the unweighted PAR photon flux density, equivalent to $6.02 \times 10^{17} \text{ photons m}^{-2}\text{s}^{-1}$ ($=1 \mu\text{mol m}^{-2}\text{s}^{-1}$). Near the solar noon, the atmospheric PAR absorption is minimal. However, upper atmospheric PAR can be attenuated by more than 17% depending on the optical depth of the atmosphere which is greater in higher latitudes and at the sunrise and sunset times (Ooms, Dinh, Sargent, & Sinton, 2016). Moreover, PAR is scattered and absorbed by aerosols, including smoke, dust, marine salt, and dissolved water vapor. Clouds can also significantly reduce the effective PAR that reaches the surface due to multiple scattering effects. Hence, the modeling of PAR within a sub-tropical region (Liu, Liang, He, Liu, & Zheng, 2008) where complicated atmospheric effects are minimal allows for a more insightful study of the key factors that impact solar and biomass energy production systems.

Total incident PAR (i.e., under cloud-free conditions) is determined mathematically by an astronomical function of the solar zenith angle (θ_z) (e.g., (Michalsky, 1988)) for any location and time without the need for a specific apparatus. Accordingly, it is a useful tool for future mapping at a wide range of spatial and temporal scales, including sites located in remote and complex geographic terrains. θ_z is an ideal model input for the prediction of solar radiation under cloud-free conditions. A previous study used this astronomical property for ultraviolet index forecasting [36] and determined θ_z values for the region at 5-min intervals. For this study, the astronomical Almanac's algorithm (Michalsky, 1988), valid until 2050, provided the solar position with an accuracy of $\pm 0.01^\circ$.

2.2 | Adaptive Neuro-Fuzzy Inference System

This is the first paper to study the application of an ANFIS model, proposed by Jang and Sun (1995), to forecast very short-term PAR with a limited input dataset, comprised of

only the θ_z values. ANFIS model's potential advantages over the standalone ANN arise from its integration of fuzzy set theory into a neural network approach (Kurtulus & Razack, 2010; Tabari, Kisi, Ezani, & Talaei, 2012a; Tabari, Talaei, & Abghari, 2012b). This generates a hybrid system where each sub-layer of a feed-forward neural network can be identified as a neuro-fuzzy component. Although ANN is a powerful decision-support tool for modeling natural processes that contain feature data (Deo & Sahin, 2017; Deo & Şahin, 2015), it is not without deficiencies. Consequently, the combination of a neural network and the fuzzy logic system can result in improved speed, greater fault tolerance, and better compliance compared to a standalone ANN network or a standalone fuzzy logic model (Setlak, 2008).

In the fuzzy-enhanced ANFIS model, simulations employ Takagi–Sugeno–Kang (TSK) fuzzy rule type 3 along with linear combinations of model inputs and a constant. The output is weighted as the mean of the fuzzy rule's output (Sugeno & Kang, 1988). A significant merit of ANFIS, over standalone fuzzy logic or ANN, is its ability to extract features via a fuzzy logic mechanism, where a multi-valued logic that approximates all the features is implemented. The degree of truth related to the features ranges between 0 and 1. It contrasts propositional logic which is not constrained to these truth values (Von Altrock, 1995). Due to a hybrid system, fuzzy logic considers imprecision and uncertainty, while neural network provides versatility for feature extraction. Implementation of an ANFIS model involves an initial fuzzy model with derived inputs from fuzzy rules extracted from input-target data. In addition, the neural network is executed to fine-tune the rules of the fuzzy system to generate an optimal ANFIS model.

ANFIS is designed with grid partitioning, subtractive clustering, and C-Fuzzy Means clustering algorithms based on the Mamdani, Sugeno, or Tsumoto systems (Kisi, 2013; Mamdani, 1976; Sugeno & Kang, 1988; Takagi & Sugeno, 1985). Our study adopted an ANFIS model with C-Fuzzy Means (CM) clustering based on the Sugeno system (hereafter, "ANFIS-CM") (Bezdek, 2013; Bezdek, Ehrlich, & Full, 1984; Dunn, 1973). Compared to the Mamdani or Tsumoto, the Sugeno system provides ANFIS-CM with a greater computational efficiency and adaptability, feature optimization, better provision for the continuity of outputs, and robust mathematical analysis of the input features (Kaur & Kaur, 2012; Zaher, Kandil, & Fahmy, 2014). ANFIS-CM approach has the ability to determine and iteratively update the membership values with pre-defined clusters, where its inputs are members of all clusters with corresponding membership values (Chattopadhyay, Pratihari, & De Sarkar, 2011), optimized by minimizing the cost function (Bezdek et al., 1984; Collazo-Cuevas et al., 2010):

$$\Lambda = \sum_{k=1}^{k=n} \sum_{i=1}^{i=c} \eta_{ik}^m |x_k - \lambda_i|^2 \quad (1)$$

where c is the total number of clusters; i is the cluster number; k is the datum number; m is a constant, where $m > 1.0$, typically $m = 2.0$; n is the total number of datum points; x_k is the k^{th} datum point; λ_i is the i^{th} cluster center, and η^m is the degree of membership of the k^{th} datum point in the i^{th} cluster.

The degree of membership function is (Bezdek et al., 1984; Chattopadhyay et al., 2011):

$$\eta_{ik}^m = \left[\sum_{j=1}^n \left(\frac{|x_k - \lambda_i|}{|x_k - \lambda_j|} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (2)$$

The number of grid clusters (c) can be set with an arbitrary guess for each cluster center λ_i (where $\lambda_i = 1, 2, 3, \dots, c$) for the algorithm to converge to a solution. In accordance with Bezdek et al. (1984), the optimal value for λ_i will represent either a local minimum or a saddle point of the cost function.

In contrast to a model constructed with grid partitioning, ANFIS-CM utilizes fuzzy partitioning where data are allocated to more than one cluster with membership values bounded by 0 or 1. ANFIS-CM applies fuzzy rules to the input-target data (i.e., “the IF-THEN” rule) based on their antecedents (“IF” part) and consequents (“Then” part). In the Sugeno ANFIS system, a rule is constituted by a weighted linear combination of crisp inputs represented with two inputs (x and y) and one output (f) as observed in Equations 3 and 4.

$$\text{Rule 1: IF } x \text{ is } P_1 \text{ and } y \text{ is } Q_1, \text{ then } f_1 = p_1x + q_1y + r_1 \quad (3)$$

$$\text{Rule 2: IF } x \text{ is } P_2 \text{ and } y \text{ is } Q_2, \text{ then } f_2 = p_2x + q_2y + r_2 \quad (4)$$

where f_i represents the output within the fuzzy region; p_i , q_i , and r_i are the design parameters determined during model training ($i = 1, 2$), and P_i and Q_i are fuzzy sets.

Figure 2a outlines a schematic view of an ANFIS model with two inputs, one output, and five layers where the zeroth layer is the input node. The function of other layers are as follows (Wen et al., 2015):

- Layer 1 is the fuzzification region, with nodes representing a membership value to which they identify using fuzzy sets:

$$O_{1,i} = \mu_{A_i}(x) \quad \text{for } i = 1, 2 \quad (5)$$

$$O_{1,i} = \mu_{B_{i-2}}(y) \quad \text{for } i = 1, 2 \quad (6)$$

where x and y are the two crisp inputs for node I , and A_i and B_i are fuzzy sets associated with this node and characterized by the shape of the membership function, occupying appropriately parameterized membership function.

For an optimal ANFIS model, appropriate first-layer membership functions must be selected.

- Layer 2 consists of nodes providing the strength of fuzzy rules by a multiplicative operator on the incoming signal which is released as the product (i.e., $O_{2,i}$ represents the firing strength of a rule, W_i) and is computed as:

$$O_{2,i} = W_i = \mu_{A_i}(x) \times \mu_{B_{i-2}}(y) \quad \text{for } i = 1, 2 \quad (7)$$

- In Layer 3, every node is a fixed node (denoted as N) where the i^{th} node calculates the ratio between the i^{th} rule's firing strength (W_i) to the sum of all rules' firing strengths ($W_1 + W_2$) to yield the normalized firing strength (\bar{W}_i):

$$O_{3,i} = \bar{W}_i = \frac{W_i}{W_1 + W_2} \quad \text{for } i = 1, 2 \quad (8)$$

- For Layer 4, node i determines the contribution of the i^{th} rule toward the outputs in terms of:

$$O_{4,i} = \bar{W}_i f_i (p_i x + q_i y + r_i) \quad (9)$$

where $\{p_i, q_i, r_i\}$ are the set parameters, and \bar{W}_i is the weighted output of Layer 3.

- Layer 5, the summation layer, is used to compute the overall strength (i.e., output $O_{5,i}$):

$$O_{5,i} = \sum_i \bar{W}_i f_i = \frac{\sum_i W_i f_i}{W_i} \quad (10)$$

2.3 | Genetic Programming

Genetic programming (GP; Figure 2b) used evolutionary principles to analyze patterns (Koza, 1992). GP achieves a systematic and domain-independent model with optimally tuned equations (Poli, Langdon, McPhee, & Koza, 2008). GP generates mathematical formulas that resemble input features (Deo & Samui, 2017). A fitness function is applied by (i) minimizing mean square error between the trained model and target data, creating functional and terminal sets and (ii) deducing the parameters for operators (e.g., crossover, mutation probability) (Sreekanth & Datta, 2011). GP is optimized to reach an acceptable level of agreement between the target and inputs. Functional nodes employing arithmetic operations (e.g., +, −, ×, ÷), Boolean logic (e.g., AND, OR, NOT), conditional (IF, THEN, ELSE), and mathematical functions (SIN, EXP, COS, TAN) are applied to map training features with target data. Functions and terminals are deduced randomly to formulate tree structures with a root node and a branch (Gandomi & Alavi, 2012; Guven & Kişi, 2011) (Mehr, Kahya, & Olyaie, 2013). A population of equations is constructed where

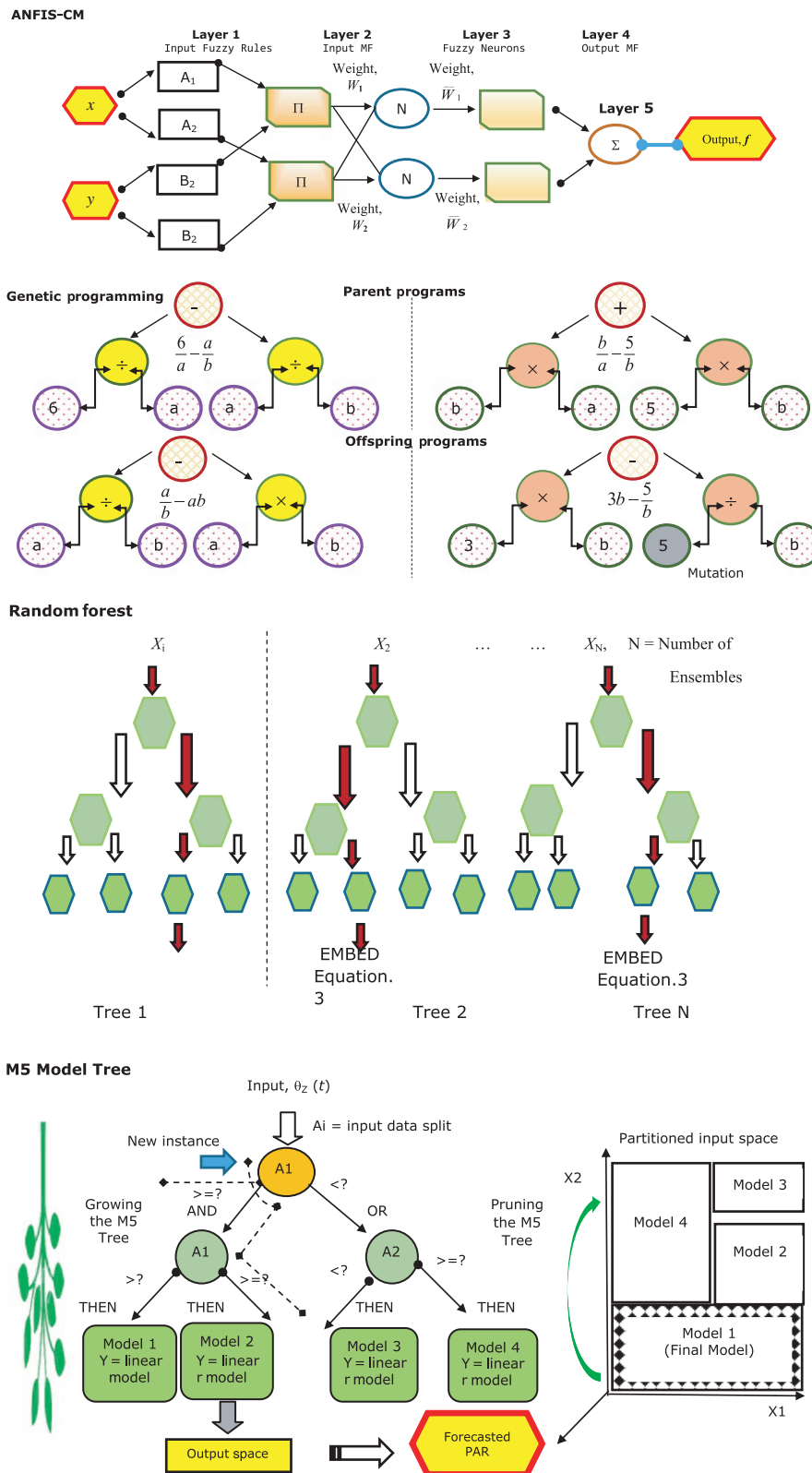


FIGURE 2 A schematic structure of the Adaptive Neuro-Fuzzy Inference System with Fuzzy C-Means Clustering (ANFIS-CM), Genetic Programming (GP), Random Forest (RF), and the M5 Model Tree algorithm

“parents” are selected out of individuals and “off-springs” are created from “parents” through the processes emulating reproduction, mutation, and crossover (Deo & Samui, 2017). As

this follows a non-deterministic process, the GP model should be run several times with the termination defined regardless of how many initial populations have been set.

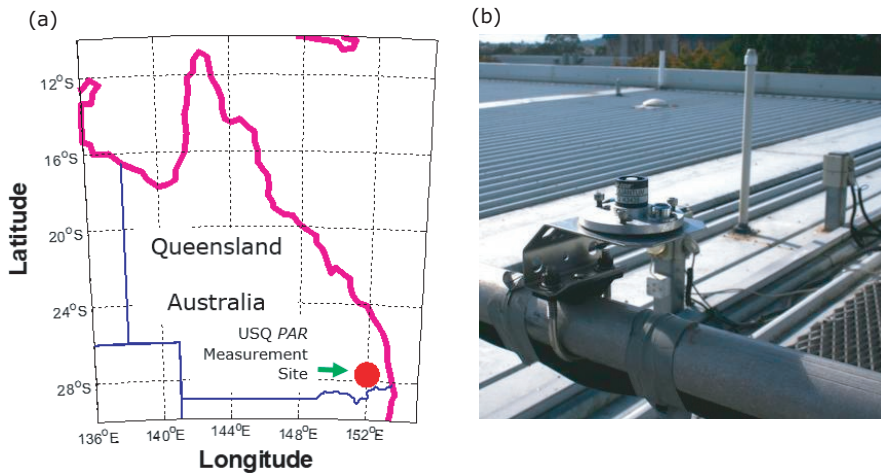


FIGURE 3 (a) Location of experimental facility in sub-tropical Queensland. (b) Roof top image of the LI-COR Quantum sensor connected to the CR100 Campbell Data logger at The University of Southern Queensland (USQ) Atmospheric and Ultraviolet monitoring station

2.4 | M5 tree model

M5 Tree algorithm (Quinlan, 1992) is based on a binary decision process (Deo et al., 2017) (Figure 2c). It splits input-target data and applies linear regression at the terminal (leaf) nodes to develop a set of predictive relationships (Mitchell, 1997). Data are split in subsets to create a decision tree and pruned to maintain the best data features (Rahimikhoob, Asadi, & Mashal, 2013). Because M5 Tree establishes relationships between the inputs and target (Bhattacharya & Solomatine, 2005), the “divide-and-conquer” rule is applied to generate models. The model is selected based on the standard deviation of class values where a consequent reduction in the model’s training error, σ_R , is monitored (Bhattacharya & Solomatine, 2005; Kisi, 2015):

$$\sigma_R = \sigma(\Lambda) - \sum_{i=1}^{i=N} \frac{\Lambda_i}{\Lambda} \cdot \sigma(\Lambda_i) \quad (11)$$

where Λ represents a set of examples that reach the node, and Λ_i is the subset of examples that have the i^{th} outcome of the training set.

When the maximum number of splits has been attained, a regression is applied to select the optimal trees (i.e., those that reach a minimum value of σ_R). Splitting ceases when the class value reaches the node with the lowest σ_R . Further details of the M5 Tree model are stated in Ref. (Bhattacharya & Solomatine, 2005; Kisi, 2015). (Quinlan, 1992; Witten & Frank, 2005).

2.5 | Multiple linear regression

MLR is a statistical model that draws out the cause and effects between objective and predictor variables (Deo & Sahin, 2017). For N observations of k predictors, the MLR can be expressed as a regression equation (Draper & Smith, 1981; Montgomery, Peck, & Vining, 2012):

$$Y = C + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (12)$$

where C is the ordinate intercept; X is a vector of predictor variable(s), (i.e., the $(N \times k)$ matrix of θ_Z); Y is an $(N \times 1)$ matrix of the objective variable (i.e., the PAR data in the training phase of the present study), and β is the regression coefficients (Civelekoglu, Yigit, Diamadopoulos, & Kitis, 2007; Şahin, Kaya, & Uyar, 2013), estimated by a least-squares method (Ozdamar, 2004).

Details regarding MLR can be found in the works of Draper and Smith (1981) and Montgomery et al. (2012).

2.6 | Random forest

Random Forest (RF) is an efficient algorithm based on model aggregations (Breiman, 2001). It belongs in the family of ensemble models used to generate accurate predictions with minimal overfitting (Breiman, 2001). For an RF model, a pair of data points [$D = \{(X_1, Y_1), \dots, (X_N, Y_N)\}$] is composed of N identically distributed observations of the vector (X, Y) . In this study, X is the input θ_Z , and Y is the response variable PAR .

The RF seeks to combine binary trees built with bootstrapped samples from the original learning sample D , where a subset of explanatory variables X have been chosen randomly at each node. A combination of random trees (typically 2,000) is grown based on the input data. Each tree is generated by a bootstrap sample, leaving about a third of the overall samples for validation. These are also referred to as out-of-bag (OOB) predictions. The splits (branching) of trees are determined with a randomized subset of predictors at each node, and the outcome is determined as the average of all trees (Breiman, 2001; Cutler et al., 2007). RF uses OOB samples to determine the errors that serve as independent observations used to grow the tree. Accordingly, no cross-validation data are required (Prasad, Iverson, & Liaw, 2006). RF involves these four steps:

- 1 If N represents the number of data cases in the training set, a sample of these cases is drawn randomly with replacement.

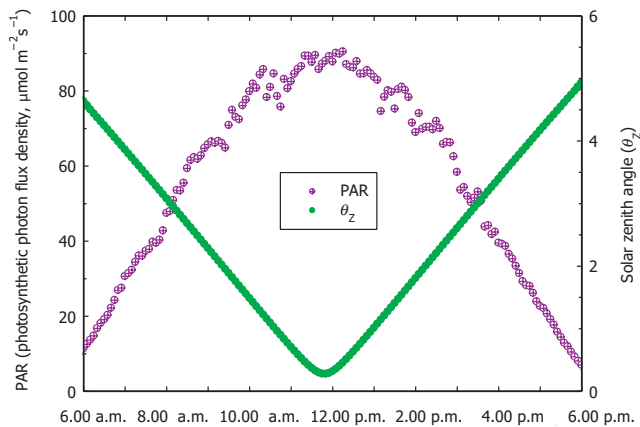


FIGURE 4 The 5-min interval photosynthetically active radiation (PAR) averaged over the austral summer study period (1–31 XII 2012), with the corresponding time-averaged solar zenith angle (θ_z)

These are used for training (growing the original trees).

- 2 If there are M predictors (or inputs), the RF will specify a number, $m < M$ such that at each node, a total of m variables are drawn randomly out of the M data. The best split on these m variables is adopted to split the node, while holding the value of m as a constant as the forest continues to grow.
- 3 During the execution of the RF algorithm, each resulting tree is grown to its maximum extent without considering any pruning applied to its structure.
- 4 The new data are predicted by aggregating the predictions of n trees (i.e., the mean value is determined for the case of a regression problem). OOB estimate of the error rate can be quite accurate provided a sufficient number of trees have been grown (Bylander, 2002).

3 | STUDY AREA, MATERIALS, AND METHODOLOGY

3.1 | Apparatus and data acquisition

At the Toowoomba Campus of the University of Southern Queensland (USQ), 120 km west of Brisbane, a quality-controlled atmospheric and ultraviolet monitoring station has measured PAR (and weather conditions) continuously since 2011 (Figure 3a). Located at an elevation of 690 m AMSL, Toowoomba is a regional city with high solar potential. It is classified as a regional center for agricultural activities, making PAR forecast models desirable (Section 3.1). Moreover, as a sub-tropical center situated inland of the Pacific Ocean, the site has limited anthropogenic and aerosol effects. The specific USQ study site has a large number of sunshine days with a clear hemispheric view of the solar horizon (Sabburg, 2000), making it an ideal site for the investigation of PAR potentials.

High-quality cloud-affected values of PPFD data were acquired over the austral summer solstice period from 1 to 31 December 2012 using a PAR Quantum sensor (LI-190R; LI-COR, Lincoln, NE, USA) connected to a CR100 Campbell Scientific data logger (Logan, UT, USA) (Figure 3b). The LI-190R was installed on an unobstructed rooftop site at USQ to monitor PAR at 5-min intervals over a 24-hr period. Used in a number of studies (Gill, Ming, & Ouyang, 2017; Johnson et al., 2015) and designed for long-term outdoor use, the LR-190R has a stated uncertainty of $\pm 5\%$ (US National Institute of Standards and Technology standard). In the present study, PAR data for the daytime hours (0500H–1900H) were used since irradiance is most likely to be intercepted by plants during the day and nighttime levels are nearly zero.

The resulting 197 PAR data points were matched with 5-min interval calculated θ_z values to create a matrix comprised of θ_z and PAR (i.e., input-target pairs). The matrix served to train, validate, and test the PAR models. Mean variation in θ_z for globally averaged PAR at 5-min intervals is shown in Figure 4 and the associated descriptive statistics described in Table 1.

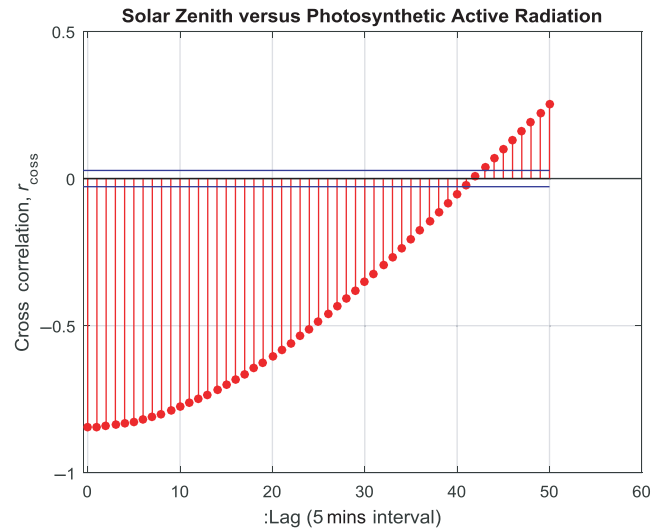
3.2 | PAR forecast model development

Very short-term (5 min) and hourly forecast models were developed with MATLAB subroutines. Given that the purpose of this study was to use limited input data, the ANFIS-CM model drew on θ_z data deduced empirically (Michalsky, 1988) as a single predictor. The model's performance was evaluated against GP, MLR, M5 Tree, and RF models (Sections 2.2–2.5). Data pairs (θ_z and PAR, Table 1) were partitioned in the ratio of 80:10:10, where the first partition (01–25 December) was used for model design, the middle (26–28 December) for validation, and the remaining (29–31 December) for testing. As a preliminary check on the exploratory and response relationships between θ_z and PAR, all data were examined via cross-correlations based on the training set. The cross-correlation coefficient (r_{cross}) was computed using the covariance (Φ) function at lagged time-scales to assess the similarity between predictor and time-shifted copies of the objective variable (i.e., θ_z vs. PAR). Figure 5 shows the r_{cross} value for different lags of θ_z vs. PAR data.

The high r_{cross} value, at a lag of 40, reaffirms the notion that a PAR model could be developed from the features present in the solar zenith angle data. Therefore, the input–output matrix was scaled between [0, 1] to avoid large fluctuations in data attributes from dominating patterns represented in the smaller attributes (Hsu, Chang, & Lin, 2003). The scaling also permitted optimizing the algorithm by avoiding additional iterations required for features with large numerical values (Deo et al., 2016; Samui & Dixon, 2012):

TABLE 1 Descriptive statistics of photosynthetically active radiation data for the study site located at the University of Southern Queensland, USQ Toowoomba

Data acquisition period (d/m/y)	Data partition	Forecasting horizon	Records	Photosynthetic photon flux density statistic ($\mu\text{ mol m}^{-2}\text{ s}^{-1}$)						
				Mean	SD	Median	Min	Max	Skew	Flatness
1–25 XII 2012	Training	5 min	5197 (~80%)	983.08	983.04	938.08	0.84	2603.10	0.51	1.68
		1 hr	434 (~80%)	42.13	28.61	41.14	0.39	91.54	0.05	-1.36
26–28 XII 2012	Validation	5 min	507 (~10%)	848.46	649.05	649.05	2.48	2772.20	0.61	2.15
		1 hr	42 (~10%)	38.30	28.13	33.57	0.29	89.42	0.34	-1.17
29–31 XII 2012	Testing	5 min	507 (~10%)	1117.60	1234.00	1234.00	7.62	2410.70	-0.13	1.64
		1 hr	42 (~10%)	48.66	30.03	54.12	1.31	90.05	-0.27	-1.40

**FIGURE 5** Cross correlation coefficient (r_{cross}) computed between the solar zenith angle (θ_z) and the photosynthetically active radiation (PAR) at the 5-min interval over the austral summer study period (1–31 XII 2012). The blue lines denote the 95% confidence interval

$$Z_{\text{norm}} = \frac{(Z - Z_{\text{min}})}{(Z_{\text{max}} - Z_{\text{min}})} \quad (13)$$

where Z is any data series (input or output); Z_{max} is the maximum value in data series Z ; Z_{min} is the minimum value in data series Z ; Z_{norm} is the normalized data series.

While the ANFIS model is based on grid partitioning, subtractive clustering, fuzzy C-means (CM) clustering, or other CM methods (Guillaume, 2001), the previous studies' (Abdulshahed, Longstaff, Fletcher, & Myers, 2015) methodology, which utilized CM for fuzzy MFs and rules, was followed. Introduced by Dunn (Dunn, 1973) and improved by Bezdek (Bezdek et al., 1984), CM is a clustering tool where each datum is assigned to a cluster with a specified membership. Compared to other clustering methods, CM aims to create a small but sufficient number of fuzzy rules such that input-target data pairs (θ_z , PAR) allow training of the ANFIS model. With CM, MFs have the flexibility to be altered through the learning process and subsequent adjustments can be made by a supervised learning process.

Rules that modeled the input-target data behaviors were deduced for antecedents and consequents. The type of FIS was set to "sugeno," allowing specified clusters for rules and MFs. Each subset was trained by two approaches, according to Park et al. (Park, Kim, Lim, & Kang, 2005), to determine the solar zenith angle clusters for PAR error compensations in the trained model:

1. In the training and validation period, several clusters (between 2 and 6) were tested and the model's

TABLE 2 Design of Adaptive Neuro-Fuzzy Inference System (ANFIS), genetic programming (GP), M5 Model Tree, Random Forest (RF), and multiple linear regression (MLR) model, including the model's training and validation performance statistics for 5-min forecasting horizons: correlation coefficient (r), root-mean-square error (RMSE) ($\mu\text{mol m}^{-2} \text{s}^{-1}$)

Model	Design parameters	Training		Validation		
		r	RMSE	r	RMSE	
ANFIS	ANFIS Type	“sugeno”	0.858	358.93	0.744	488.85
	Epoch	100				
	Clusters	“auto” subclust (3)				
	Radii	0.5				
	Membership Function	“gaussmf” (Gaussian)				
GP	Population size	100	0.855	362.46	0.741	491.01
	Max. generations	150				
	Generations elapsed	150				
	Tournament size	200				
	Elite fraction	0.7				
	Probability of pareto tournament	1				
	Max. tree depth	4				
	ERC probability	0.05				
	Crossover probability	0.84				
	Function set	$\times, \div, +, -, ^2,$ $\tanh, e, \log,$ $\times 3, + 3, \sqrt{}$ ^3				
Model ensembles	1,000 runs					
M5 tree	Number of rules	237	0.879	333.42	0.747	492.52
	Minimum cases	5				
	Smoothing	15				
	Split threshold	0.05				
RF	Number of trees	50, 100, 200, 400, 800, 1600 (optimal value 800)	0.917	279.15	0.742	510.09
	Leaf	5				
	Foot	1				
	Surrogate	On				
	Delta criterion decision split	0.1007				
	Number of predictor split	2248				
MLR	Y-intercept (C)	0.704	0.851	367.44	0.738	493.54
	Model coefficient (β)	-0.7477				

validation error and Legates & McCabes Index (E_{LM}) were monitored to select the optimal cluster and MFs.

- The “auto” option was set using the “subclust” algorithm (radii = 0.5) for cross-checking.

Clustering was performed at each iteration, and the objective function was minimized to deduce the best number and location of clusters.

Sugeno type FIS, comprised of a combination of least-squares and backpropagation gradient descent methods, was applied in the ANFIS model. The model's training was halted when the designated epoch (raised incrementally from 5, 25, 50, 75, 100, and 150 to 200) and training error goal (set to 0) were attained. To achieve optimal accuracy and prevent overfitting, an independent validation set was

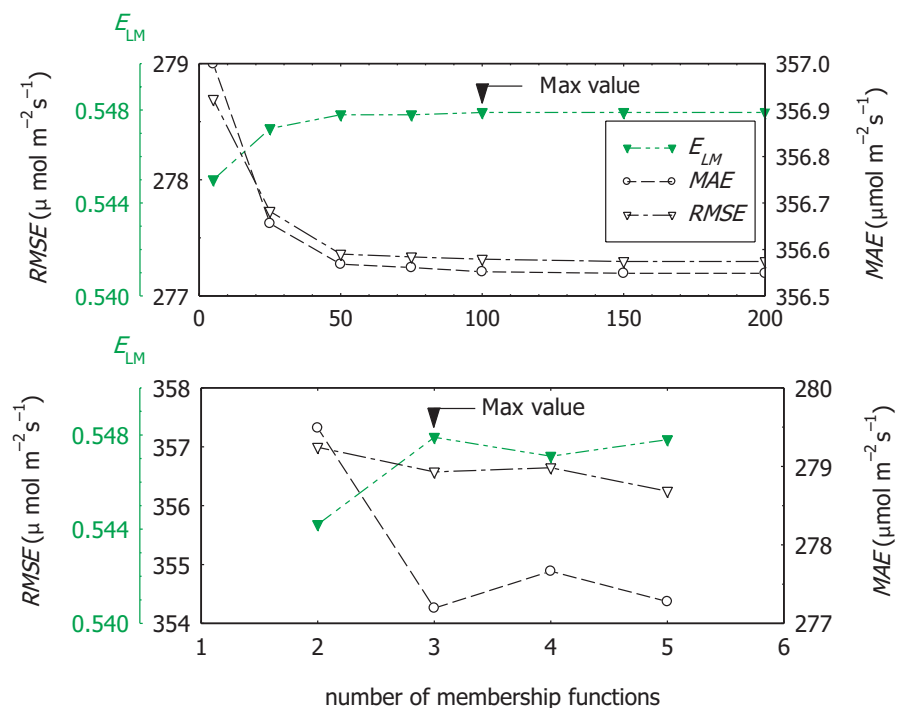


FIGURE 6 The training performance of the ANFIS-CM models with different training epochs (top sub-panel) and the number of membership functions, MF (bottom sub-panel) in terms of the root mean square error ($RMSE$), mean absolute error (MAE), and the Legates & McCabes Index (E_{LM}) (Note that an optimal epoch of 100 and 3 MFs are indicated, as utilised in the final ANFIS model)

used. Subsequently, overfitting was detected when the error (i.e., difference between simulated and observed output) increased while the training error decreased. A Gaussian MF, employed in previous studies (Abdulshahed et al., 2015), was adopted to describe the membership degree as it is a smooth and nonzero function (Abdulshahed, Longstaff, Fletcher, & Myers, 2013)

Table 2 depicts ANFIS training and validation performances in terms of the root-mean-square error ($RMSE$) and correlation coefficient (r). Figure 6(a) shows the performance with the different epochs and MFs tested through the mean absolute error (MAE) and Legates & McCabes Index (E_{LM}). The optimal epoch was found to be 100 with 3 MFs leading to the highest r value and lowest $RMSE$ value for the training and validation dataset.

To serve as a benchmark for the ANFIS-CM model's performance, a genetic programming (GP) approach was used (Deo & Samui, 2017; Searson, 2015). A symbolic representation with a configuration set to evolve multigene individuals and consisting of one or more genes (each as a traditional tree) was employed since more accurate models can be obtained via this method (Searson, 2015). With θ_Z as the predictor variable, the required number of genes was acquired incrementally to improve the fitness (i.e., reduce the error of input-target data). The weights for the genes were obtained through standard least-square functions to regress them against the target (i.e., PAR) data. The resulting pseudo-linear model captured non-linear behaviors embedded in the predictor data. Considering the model's non-deterministic nature, the GP was run with a population size of 100, a large termination time (~800 s), and 10

ensemble models. Table 2 shows the details of optimal GP model configurations.

ANFIS-CM model was also benchmarked against an M5 Tree model (Table 2). Several training parameters were set for tree initialization. After testing several values, the minimum split, smoothing, and split threshold were set to 2.0, 15, and 0.05, respectively. These values concur with those assigned in earlier studies (Deo et al., 2017; Wang, 1997). Furthermore, the pruning option was applied to implement the "divide-and-conquer rule" (Kisi, 2015; Rahimikhoob et al., 2013). The optimum number of decision trees (or "rules") was obtained from the model with lowest $RMSE$ and largest r^2 , confirmed through the standard deviation of class values and highest reduction in training error. The model was then fine-tuned to yield the best set of M5 Tree rules.

RF and MLR models (Deo & Sahin, 2017; Prasad, Deo, Li, & Maraseni, 2018) also served as benchmarks for the ANFIS-CM model. The RF uses bootstrap aggregation ("bagging") to construct an ensemble of decision trees that regress the exploratory and response relationships between θ_Z and PAR data. The number of decision trees was optimized through several trials (50–1600 in twofold increments). The optimal model was attained with 800 decision trees, with the leaf number set to 5. The $fboot$ property, set to a default value 1, is the fraction of observations to be randomly selected with replacement for each bootstrap replica. Note that the size of each replica is given by $N \times fboot$ where N is the number of observations in the training set. The surrogate option was initiated to allow predictive measures of variable association to fill each tree, averaged over the split. Using bootstrap

replicas, trees were erected independently from sampled data. Observations that were not included were deemed OOB. The error of ensembles was estimated through computed predictions for each tree on its OOB averaged response over all ensembles, where predicted OOB observations were compared with true values. Table 2 shows the characteristics of an RF model with 2,248 predictor splits and 0.1007 as the decision split value. Lastly, an MLR model was developed (Myers, 2000), where the regression coefficients and the ordinate intercept were determined (Table 2).

3.3 | Forecast Model Evaluation

3.3.1 | Mathematical formulation of performance metrics

Direct comparisons of observed vs. forecasted PAR were employed, in conjunction with statistical score metrics (Dawson, Abraham, & See, 2007; Krause, Boyle, & Bäse, 2005), to evaluate the accuracy of ANFIS-CM and other models employed for comparative purposes. The correlation coefficient (r), $RMSE$, MAE , their normalized equivalents ($RRMSE$ and $RMAE$), and Willmott's, Nash–Sutcliffe, and Legates & McCabes Index (Krause et al., 2005; Legates & McCabe, 1999, 2013; Nash & Sutcliffe, 1970; Willmott, 1981, 1984) were chosen to compare model performances and defined by the following equations:

$$r = \frac{\sum_{i=1}^N (PAR_i^{obs} - \overline{PAR^{obs}}) (PAR_i^{for} - \overline{PAR^{for}})}{\sqrt{\sum_{i=1}^N (PAR_i^{obs} - \overline{PAR^{obs}})^2} \sqrt{\sum_{i=1}^N (PAR_i^{for} - \overline{PAR^{for}})^2}} \quad (14)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (PAR_i^{for} - PAR_i^{obs})^2} \quad (15)$$

$$RRMSE = 100 \times \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (PAR_i^{for} - PAR_i^{obs})^2}}{\frac{1}{N} \sum_{i=1}^N (PAR_i^{obs})} \quad (16)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N \left| (PAR_i^{for} - PAR_i^{obs}) \right| \quad (17)$$

$$MAPE = 100 \times \frac{1}{N} \sum_{i=1}^N \left| \frac{(PAR_i^{for} - PAR_i^{obs})}{PAR_i^{obs}} \right| \quad (18)$$

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^N (PAR_i^{obs} - PAR_i^{for})^2}{\sum_{i=1}^N (PAR_i^{obs} - \overline{PAR^{obs}})^2} \right], -\infty \leq E_{NS} \leq 1 \quad (19)$$

$$WI = 1 - \left[\frac{\sum_{i=1}^N (PAR_i^{for} - PAR_i^{obs})^2}{\sum_{i=1}^N (|PAR_i^{for} - \overline{PAR^{for}}| + |PAR_i^{obs} - \overline{PAR^{obs}}|)^2} \right], 0 \leq WI \leq 1 \quad (20)$$

$$E_{LM} = 1 - \left[\frac{\sum_{i=1}^N |PAR_i^{obs} - PAR_i^{for}|}{\sum_{i=1}^N |PAR_i^{obs} - \overline{PAR^{obs}}|} \right], 0 \leq E_{LM} \leq 1 \quad (21)$$

where N is the number of data points in the testing period (i.e., 29–31 December 2012); PAR_i^{obs} is the i^{th} observed PAR value; $\overline{\{PAR^{obs}\}}$ is the mean observed PAR value; PAR_i^{for} is the i^{th} forecasted PAR value; $\overline{PAR^{for}}$ is the mean forecasted PAR value.

3.3.2 | Physical interpretation

Several model performance measures were adopted to enable a robust and complementary evaluation of the forecasting models' accuracy in a manner designed to overcome the merits and constraints of single metrics. Further diagnostic and graphical tools (e.g., scatterplots and error distribution) (Willmott, 1984) were used in conjunction with statistical indices to garner information on the versatility of the tested models for very short-term and hourly PAR forecasting, as implemented in the past (Legates & Davis, 1997; Legates & McCabe, 1999).

The widely used r , MAE , and $RMSE$ values enabled model evaluation by a direct comparison of PAR forecasts with observed PAR. A perfect model would be expected to achieve r , $RMSE$, and MAE values of 1.0, 0.0, and 0.0, respectively. However, since r is a linear model metric that considers covariance, it standardizes differences between observed and predicted means and variances for forecasted and observed PAR. Because r is insensitive to additive and proportional differences between forecasts and observations, studies have illustrated that the correlation-based measure has higher sensitivity to outliers than to observations near the mean value (Moore, Notz, & Notz, 2006). Therefore, regardless if $r = 1.0$ between observed and forecasted values, a model's accuracy is defined by whether its slope and ordinate intercept approach 1 and 0, respectively.

In accordance with insights provided by Legates & McCabes Index (1999), oversensitivity to outliers can lead to a bias toward extremes if only correlation-based metrics are

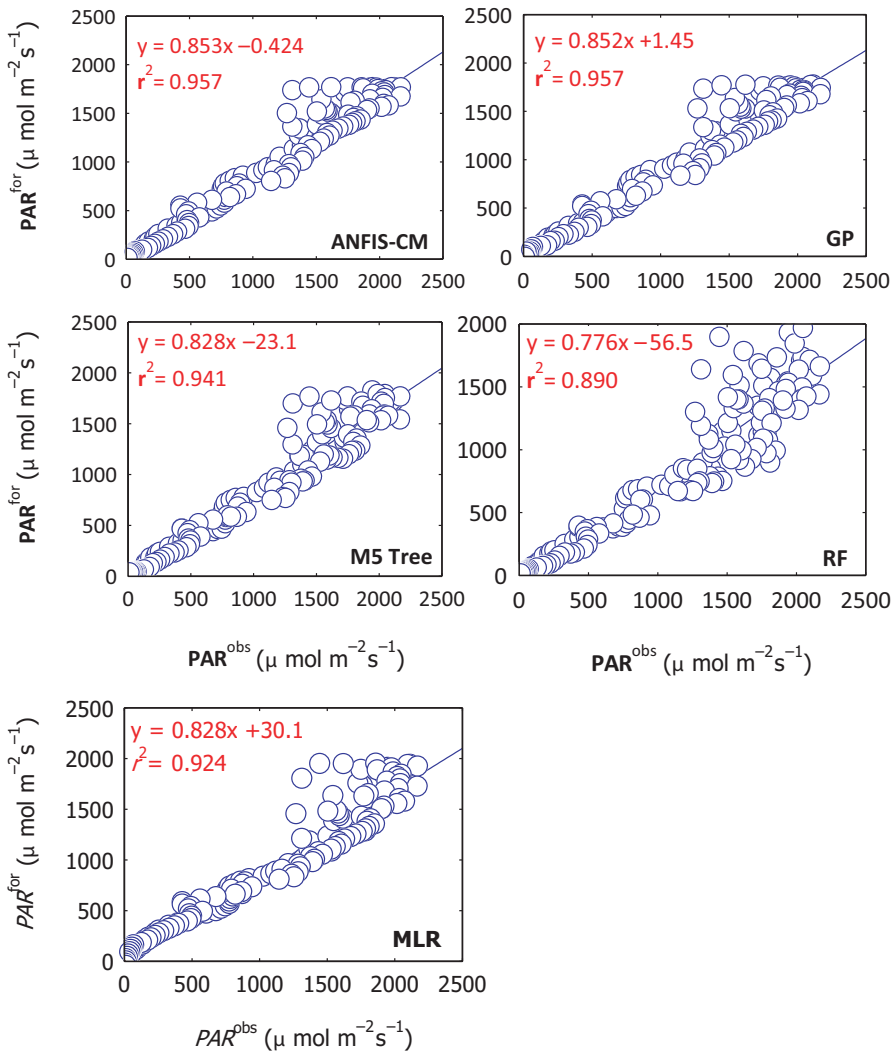


FIGURE 7 Scatterplots of the very short-term (5 min interval) observed (i.e., PAR^{obs}) and forecasted (i.e., PAR^{for}) photosynthetically-active radiation (PAR) generated through the ANFIS-CM model compared with the Genetic Programming (GP), M5 Model Tree, Random Forest (RF) and Multiple Linear Regression (MLR) algorithms tested for a 3-day average test simulation from Dec. 29–31, 2012. A least square regression line, $y = mx + C$ and the coefficient of determination (R^2) are inserted in each sub-panel

employed in model evaluation. The widely reported *MAE* and *RMSE* satisfy the triangle inequality requirement as a distance metric and provide an objective assessment if the error distribution is Gaussian (Chai & Draxler, 2014); however, they are non-normalized metrics (i.e., describe the forecasting error in absolute terms). Accordingly, their relative percentage values (*RRMSE* and *RMAE*) are also employed in comparing model accuracy in geographically and climatologically diverse sites (Deo & Sahin, 2017; Deo et al., 2016; Mohammadi et al., 2015).

As alternative metrics, the Nash Sutcliffe's coefficient (Nash & Sutcliffe, 1970) and Willmott's Index (Willmott, 1981, 1984) bounded by $[-\infty, 1.0]$ and $[0, 1.0]$, respectively, were also adopted to assess the congruence between forecasted and observed PAR. On the other hand, the E_{NS} value assesses the deviation from unity of the ratio of mean square error to the variance in observed data. That is, if the term $(PAR_i^{obs} - PAR_i^{for})$ is as large as the variability in observed data (i.e., $(PAR_i^{obs} - PAR_i^{obs})$), E_{NS} is expected to be zero (i.e., a poor model), whereas if $(PAR_i^{obs} - PAR_i^{for}) \ll (PAR_i^{obs} - PAR_i^{obs})$, then the forecast model is considered

accurate as the value is expected to be closer to 1 (Legates & McCabe, 1999). By virtue of its normalized structure and its sensitivity to the differences in observed and simulated means and variances, the E_{NS} is considered an advancement over r , *MAE*, and *RMSE* as they are sensitive to extreme values from squared error differences. In contrast, the *WI* considers the ratio of errors; therefore, it provides an alternative to overcome the insensitivity issues associated with E_{NS} and r (Willmott, 1981). To further investigate the models' relative accuracy, the Legates & McCabe's Index ($0 \leq E_{LM} \leq 1.0$) was calculated. E_{LM} can prove more informative than the *WI* or E_{NS} when relatively large forecasted values are expected, including for a poorly fitted model (Willmott, 1984). Through this index, the errors and differences are given appropriate weight and are not inflated by their square values. As outliers may lead to a relatively higher value of *WI* due to the squaring of the differences (Willmott, 1981), the use of E_{LM} (Legates & McCabe, 1999), as a modified form of the *WI*, is deemed to provide complementary information (Legates & Davis, 1997; Legates & McCabe, 1999, 2013).

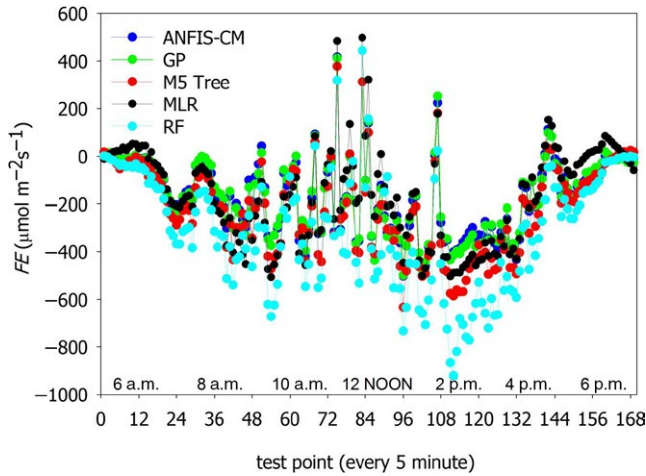


FIGURE 8 The model forecasting error, FE (i.e., forecasted minus observed value of the very short-term, 5-min photosynthetically-active radiation), generated through the ANFIS-CM model compared with the GP, M5 Tree, MLR and RF algorithm for a 3-day average test simulation from 29 to 31 XII 2012

4 | RESULTS AND GENERAL DISCUSSION

4.1 | Very short-term (5-min) PAR prediction

For a 3-day simulation period (29–31 December 2012), a direct graphical comparison (i.e., coefficient of determination, r^2 , and slope, m) of observed (i.e., PAR^{obs}) to very short-term ANFIS-CM-forecasted (i.e., PAR^{for}) data was created. Similar comparisons were made for equivalent GP, M5 Tree, RF, and MLR model-forecasted PAR (Figure 7). While both the ANFIS-CM and GP models presented high level of accuracy ($r^2 \approx 0.957$), the slope of the ANFIS-CM model ($m = 0.853$) was marginally better than the GP model's ($m = 0.852$) value. The accuracy level of the M5 Tree, MLR, and RF models was progressively inferior ($r^2 \approx 0.941$, 0.924, and 0.890, respectively). All forecasting models' accuracy was greatest with relatively small observed values of

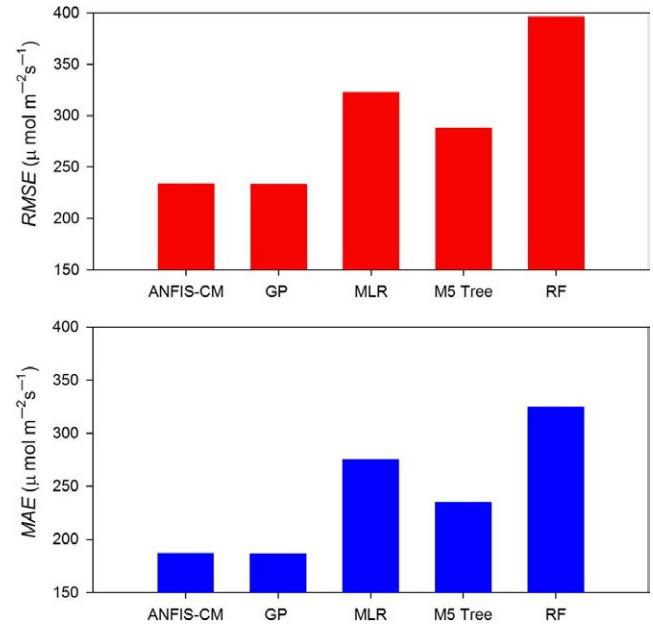


FIGURE 9 Evaluation of the ANFIS-CM model against GP, M5 Tree, MLR and RF models for forecasting the very short-term, 5-min interval photosynthetic active radiation as measured by the root mean square error ($RMSE$) and mean absolute error (MAE) in a 3-day average simulation (29 to 31 XII 2012) period

PAR (i.e., $<1250 \mu \text{mol m}^{-2} \text{s}^{-1}$). However, for large levels of PAR ($>1500 \mu \text{mol m}^{-2} \text{s}^{-1}$) and occasional-wide fluctuations caused by the intrusion of cloud cover, the accuracy of the ANFIS-CM and GP models remained largely unchanged, while the RF model performed poorly. Despite the high coefficients of determination that indicated good accuracy for the ANFIS-CM and other models, their inability to forecast large fluxes in PAR/PPFD data was likely attributable to the exclusion of clouds in the models (visible as an upward trend from the 1:1 lines, Figure 7), which were trained strictly with θ_z data.

Figure 8 shows the 3-day average forecasting error, $FE = PAR_i^{for} - PAR_i^{obs}$, for every 5 min across the testing

TABLE 3 Performance of ANFIS-CM versus comparative models (i.e., GP, MLR, M5 Tree and RF) evaluated for the 3-day test period 29–31 December 2012 in terms of correlation coefficient (r), Willmott's Index (WI), Nash–Sutcliffe coefficient (E_{NS}), and Legates & McCabe's Index (E_{LM}) and including relative (%) root-mean-square error ($RRMSE$) and relative mean absolute error ($RMAE$) for very short-term (i.e., 5 min) PAR forecasting

Model	Testing performance					
	r	WI	E_{NS}	E_{LM}	$RRMSE, \%$	$RMAE, \%$
ANFIS-CM	0.978	0.953	0.885	0.697	20.34	17.18
GP	0.978	0.953	0.885	0.697	20.38	19.78
MLR	0.965	0.921	0.804	0.578	26.56	34.37
M5 tree	0.970	0.932	0.825	0.618	25.11	26.39
Random forest	0.944	0.883	0.669	0.473	34.55	30.60

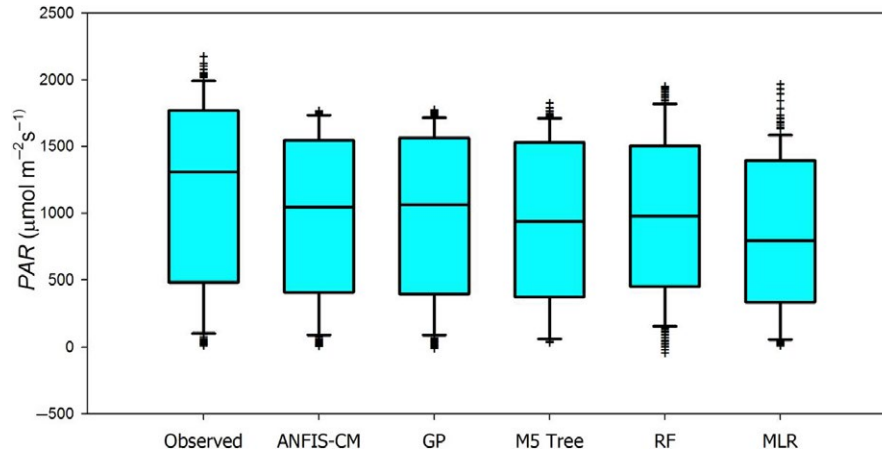


FIGURE 10 Distribution of very short-term, 5-min forecasted photosynthetically-active radiation generated by the ANFIS-CM model against GP, M5 Tree, MLR and RF algorithms for all simulation data pooled for a 3-day (29 to 31 XII 2012) period. Whiskers denote the location of outliers (i.e., extreme error), while horizontal line within the box shows median error, upper/lower boundaries of the box denote upper and lower quartiles and upper/lower horizontal lines at first whisker shows the first outlier of the extreme error outside the respective quartile value

Error statistic	ANFIS-CM	GP	M5 model tree	RF	MLR
25th percentile, p_{25}	64.74	48.75	85.82	132.98	51.08
50th percentile, p_{50}	170.56	169.37	208.48	313.56	174.68
75th percentile, p_{75}	317.41	319.84	378.07	504.77	351.92
Minimum	0.47	0.16	0.90	0.35	1.24
Maximum	502.26	498.66	633.11	920.27	506.71
Variance	19800.79	19730.06	27816.88	51730.24	25588.55
Skewness	0.26	0.23	0.30	0.34	0.41
Flatness	-1.19	-1.26	-1.02	-0.82	-1.26

TABLE 4 Distribution statistics of the absolute forecasting error, $|FE|$ ($\mu \text{ mol m}^{-2} \text{ s}^{-1}$) in testing period (29–31 December 2012) for very short-term (i.e., 5 min) PAR forecasting

period. Consistent with Figure 7, the quality of the ANFIS and GP model performances was similar, whereas the M5 Tree, MLR, and RF models displayed significantly larger errors in forecasted PAR data. Particularly, larger errors were observed for the results between 0800H and 1700H, when a relative increase in the intensity of cloud cover generated large fluctuations in PAR values recorded at the Earth's surface. Nonetheless, the ANFIS-CM and GP models' performances remained superior, with maximum forecasting errors generally lower than the three other models and the latter models also under-predicted PAR more frequently than they over-predicted it, primarily due to the repetitive presence of cloud cover.

A more detailed assessment of model performance for very short-term PAR forecasting (Table 3, Figure 9) continued to indicate higher accuracy for the ANFIS-CM and GP models. Consistent with the scatterplot, the RF model generated the least accurate results for $RMSE$ and MAE (Figure 9) and other statistical indices ($WI = 0.883$,

$E_{NS} = 0.669$, and $E_{LM} = 0.473$). When relative forecasting errors were considered over the test period, the ANFIS-CM model ($RRMSE = 20.34\%$ and $MAE = 17.18\%$) performed slightly better than the GP model ($RRMSE = 20.38\%$ and $RMAE = 19.78\%$), while for the other data-driven models were in the ranges of $25.11 \leq RRMSE \leq 34.55\%$ and $26.39 \leq RMAE \leq 34.37\%$. The forecasting error distribution (not depicted) was non-Gaussian; hence, in accordance with earlier studies (e.g., (Chai & Draxler, 2014; Mohammadi et al., 2015)), the magnitude of the relative MAE value (normally used for error distributions) was less than 20%. In concurrence with other studies (Mohammadi et al., 2015), it is evident that the present ANFIS-CM model is the most suitable for very short-term PAR prediction.

The statistical distribution of the observed and forecasted PAR from all data-driven models is presented in Figure 10 and Table 4. While all data-driven models produced widely disparate results when compared with the observed PAR

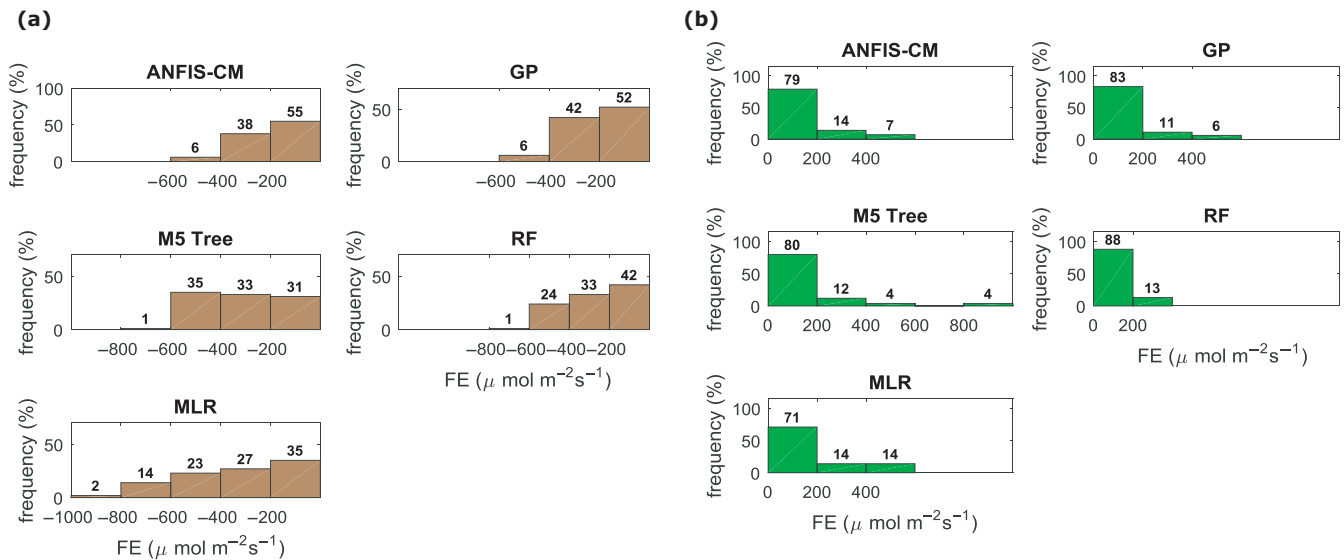


FIGURE 11 Detailed evaluation of the cumulative forecast errors generated by ANFIS-CM compared the GP, M5 Tree, MLR and RF algorithms in a 3-day average simulation (29 to 31 XII 2012). (a) Under-predictions (shown in brown), (b) Over-predictions (shown in green)

distribution, the ANFIS-CM, GP, and MLR models showed similar medians and lower and upper quartiles. For each of these three models, the lower quartiles of their PAR^{for} were similar to PAR^{obs} , whereas the upper quartiles were significantly under-predicted (a likely consequence of cloud cover). In terms of the extreme values of PAR^{obs} , the RF model was more accurate as its upper whiskers extended beyond those generated by other models. Interestingly, the GP model was relatively inaccurate compared to the ANFIS-CM model as it generated negative (i.e., false) values of PAR (leading to an elevation of the lower whiskers).

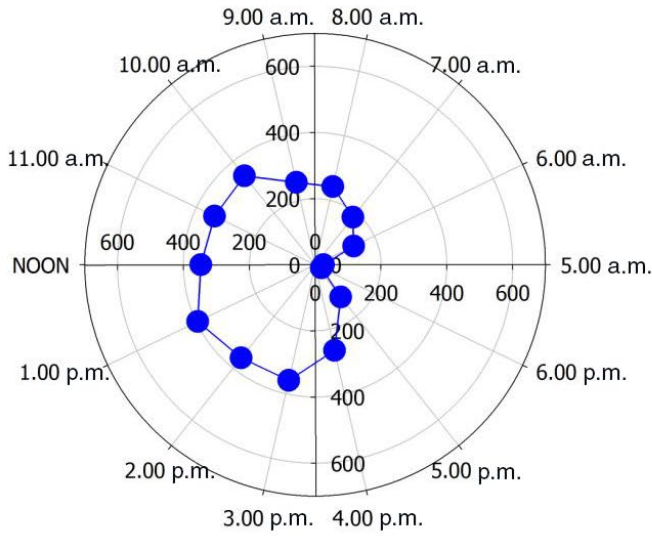
On this basis, the ANFIS-CM and the GP model appeared to outstrip the MLR, M5 Tree, and RF models in terms of accuracy. The data-driven models' cumulative forecast errors are shown in increments of $200 \mu \text{ mol m}^{-2} \text{ s}^{-1}$ for both under- and over-predictions (Figure 11). The RF model achieved more accurate forecasts for over-predicted PAR as almost 88% of all forecasted errors were within the smallest error bracket. This could be attributed to the ensemble approach applied through the bootstrapped aggregation of models used to develop the RF model. In terms of under-predicted PAR data, the ANFIS-CM model was marginally superior, with approximately 55% of all forecasted errors within $\pm 200 \mu \text{ mol m}^{-2} \text{ s}^{-1}$ compared to approximately 52% for GP, 31% for M5 Tree, 42% for RF, and only 35% for MLR models. Since the percentage of errors within the smallest error bracket was relatively low for the four comparative models, a greater proportion of errors are expected to be accumulated in larger error brackets, indicating a relatively inferior performance. Notably, the MLR and M5 Tree models were very imprecise for very short-term PAR predictions as the errors within $\pm(400\text{--}600 \mu \text{ mol m}^{-2} \text{ s}^{-1})$ and $\pm(600\text{--}800 \mu \text{ mol m}^{-2} \text{ s}^{-1})$ amounted to approximately

23% and 14%, for the MLR, and 35% and 1%, for the M5 Tree models. Similar results were obtained for the RF model, indicating that the generation of underestimated PAR values compared to the observed values was highly likely.

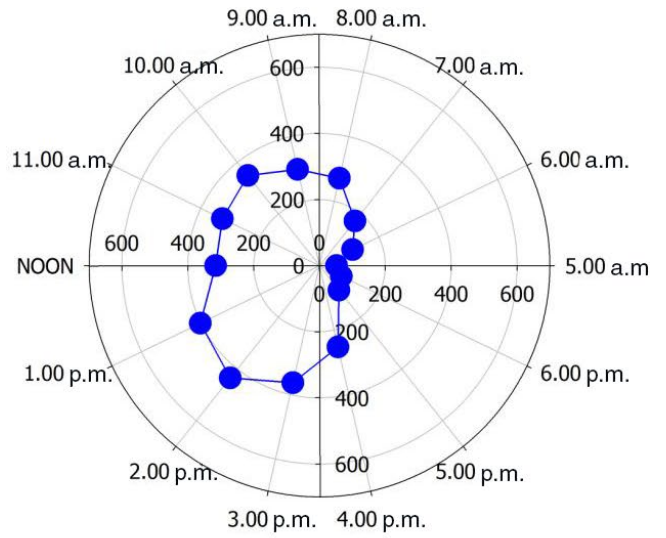
To further evaluate the accuracy of the ANFIS-CM and its counterparts, errors for very short-term PAR^{for} were accumulated over hourly periods (summed 5-min interval PAR^{obs}) from 0500H to 1900H, pooled for a 3-day test period, and plotted as a polar plot for all model simulations (Figure 12). The ANFIS model outperformed all other data-driven models, including the GP model. The patterns of errors in hourly PAR^{for} vs. PAR^{obs} showed that models are most accurate in the morning and late afternoons, although the errors generated by the RF, MLR, and M5 Model Tree were relatively larger than ANFIS-CM or GP models' results. A closer inspection of the forecasted errors by the ANFIS-CM and the GP models revealed subtle, yet significant differences in their predictive abilities. For example, at about 1400H, the accumulated hourly forecasted error for the GP model was $400 \mu \text{ mol m}^{-2} \text{ s}^{-1}$, whereas ANFIS-CM model indicated a lower threshold value. A similar trend was also found for PAR^{for} at 0900H. This analysis confirms that, although the statistical score metrics (Table 3) suggest a similar performance between ANFIS-CM and GP models, the hourly forecasted errors for the very short-term PAR^{for} justify the conclusion that the ANFIS-CM model is a superior tool when compared to the other models.

The radial error plot (for all models) may also be used to explain the Toowoomba measurement site's local climatology. Uncertainty at 1400H (and to a lesser extent at 0900H) is an effect of cloud cover. Figure 12 depicts when PAR^{for} values were most likely to be affected by cloud cover during the solstice test period by modeling PAR over short timescales

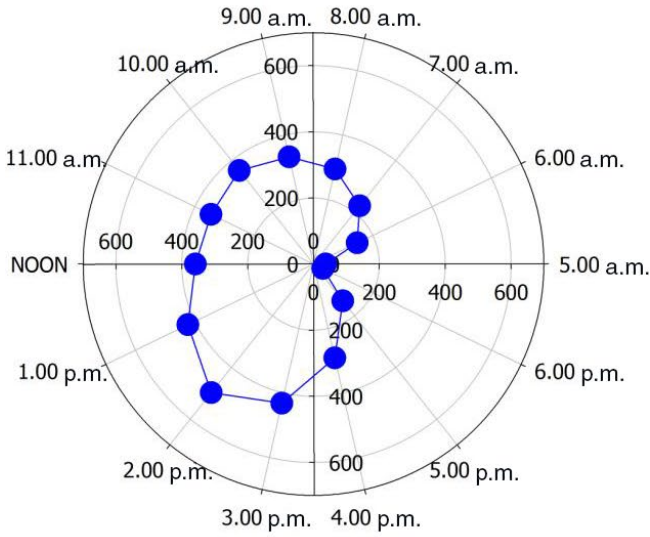
ANFIS-CM



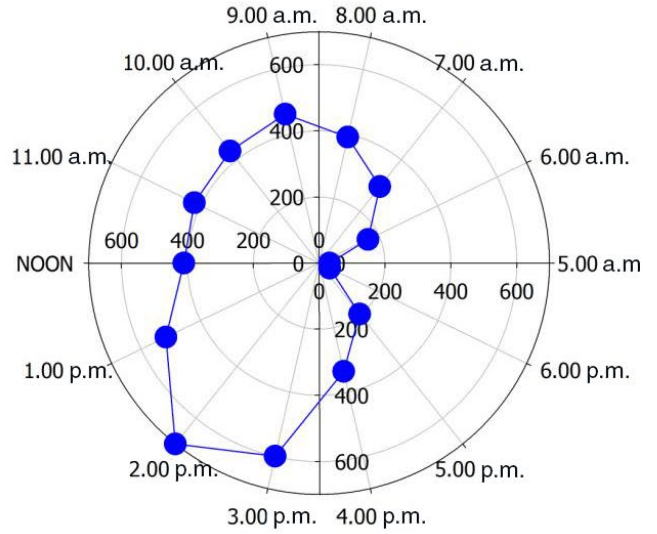
GP



M5 Tree



RF



MLR

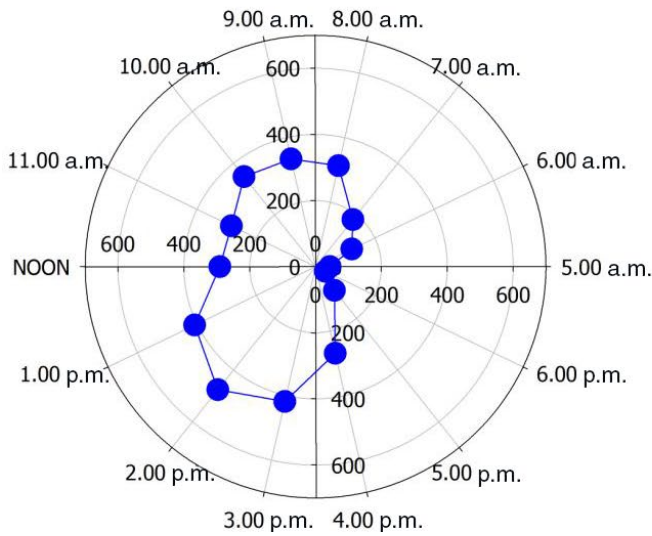


FIGURE 12 Hourly errors generated by ANFIS-CM model for very short-term forecasting of PAR versus Genetic Programming, M5 Model Tree, Random Forest, and Multiple Linear Regression (MLR) averaged for all simulation data pooled for the 3-day test period (29 to 31 XII 2012, 0500H–1900H). Outward radial axis from the origin denotes the size of the forecasted error

using θ_z as the primary input for comparison with PAR^{obs} . This demonstrates the value of very short-term PAR prediction compared to other commonly derived PAR measurements such as the daily light integral (DLI). An assessment of hourly (or shorter) PAR/PPFD-derived predictions, using ANFIS-CM plots with local daily averages by month (or season), could reveal further useful information on the temporal nature of surface PAR.

4.2 | Hourly PAR prediction

To extend the scope and relevance of this study and to ensure comparability of the results with previous studies that focussed on hourly prediction (Lopez et al., 2001; Wang et al., 2016; Yu & Guo, 2016), hourly PAR was modeled with the present ANFIS-CM and the other data-driven models. Accurate estimation of PAR is essential

for correct description of moderate timescale photosynthetic dynamics and prediction of plant biochemical processes at a micro-cellular level (Ge et al., 2011; Wang et al., 2015). Accordingly, PAR^{for} , predicted from the 5-min θ_z , was averaged over hourly intervals and employed to generate forecasts over this horizon. This required training of the models with the input-target data matrix (1–25 December 2012), the validation matrix (from 26 to 28 December 2012), and the testing matrix from (29 to 31 December 2012). Figure 13 shows the scatterplot of the hourly forecasted and observed PAR values (in $\mu\text{mol m}^{-2}\text{hr}^{-1}$).

The ANFIS-CM model outperformed all other data-driven models, as indicated by exhibiting the highest coefficient of determination (r^2) between PAR^{for} and PAR^{obs} of 0.963 compared to 0.936, 0.955, 0.902, and 0.924 for the GP, M5 Tree, RF, and MLR models, respectively. In fact, the

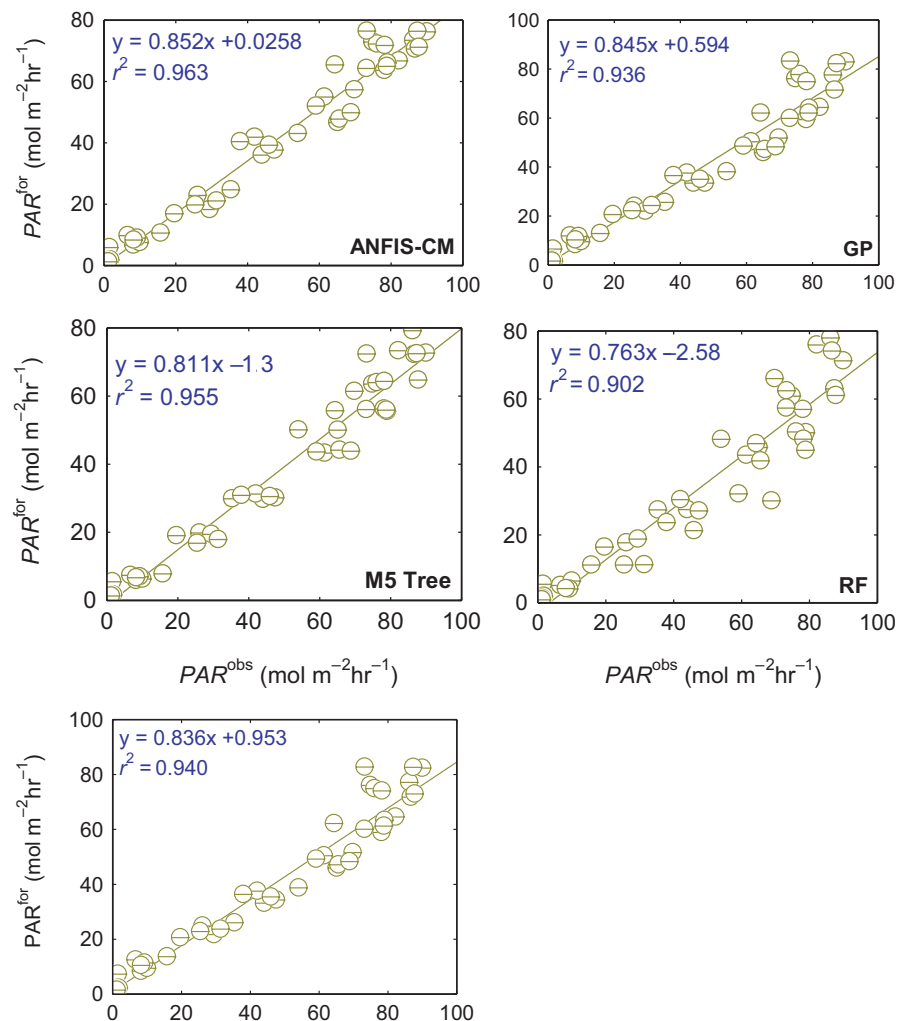


FIGURE 13 Scatterplots of hourly observed (PAR^{obs}) and forecasted (PAR^{for}) photosynthetically-active radiation (PAR) generated through ANFIS-CM compared with GP, M5 Tree, MLR and RF algorithm tested for a 3-day average test simulation (29–31 XII 2012). The least square regression line, $y = mx + C$ and coefficient of determination (R^2) are inserted in each sub-panel

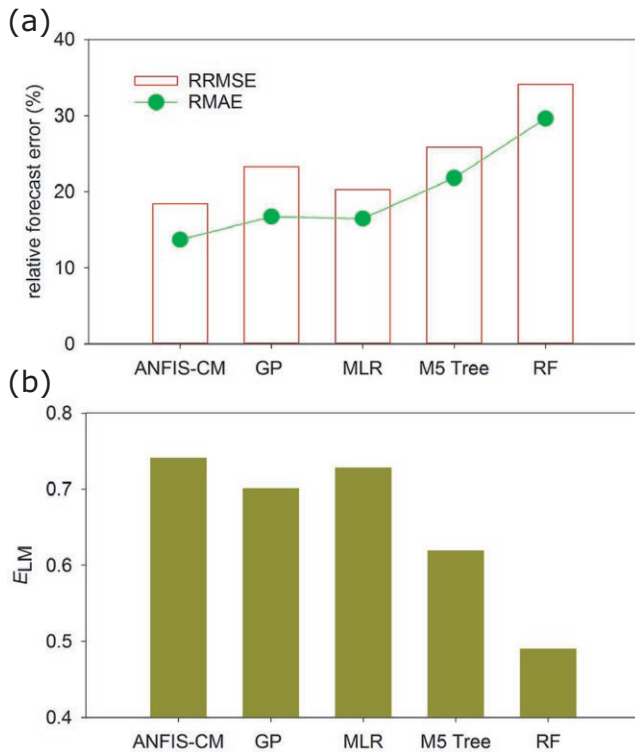


FIGURE 14 Evaluation of the ANFIS-CM model for hourly PAR forecasting versus the GP, M5 Tree, MLR and RF algorithm: (a) relative root mean square error (RRMSE) and relative mean absolute error (RMAE); (b) Legates & McCabes Index (E_{LM}) for data averaged in the 3-day testing period

maximum deviation of PAR^{for} data from its PAR^{obs} equivalent was lowest for the ANFIS-CM model, which further displayed values of slope and ordinate intercept closest to 0.0 and 1.0, respectively. While the level of agreement between PAR^{for} and PAR^{obs} was relatively high for $PAR < 60 \mu \text{mol m}^{-2} \text{hr}^{-1}$, a larger degree of scatter was evident for higher PAR values. This is consistent with the presence of greater cloud cover during the daytime period.

Further evaluation of model performances on an hourly scale was undertaken by comparing the relative forecasting error (i.e., $RMAE$ & $RRMSE$) (Figure 14 a) and E_{LM} (Figure 14b) for the ANFIS-CM model to other forecasting models, averaged over the 3-day test period. Both portions of Figure 15 confirm that the ANFIS-CM model achieved greater accuracy in forecasting hourly PAR values than the other models.

Based on comparisons of hourly PAR^{for} prediction accuracy, the present results justify that the ANFIS-CM model could be considered a suitable model for hourly PAR forecasting. The hourly forecasted PAR data generated by the MLR model yielded a relatively smaller error value compared to the GP and M5 Tree model, which stands at odds with the very short-term (5 min) forecasting horizon results. Overall, at an hourly scale, the RF model stood out

as the least accurate model (Figure 14a), consistent with its performance at a very short-term horizon (see Figure 9; Table 3). The superiority of the ANFIS-CM model against the comparable data-driven models is confirmed by an E_{LM} of 0.741 that was registered for hourly predictions of PAR, compared to the values of 0.728, 0.701, 0.619, and 0.490 for the MLR, GP, M5 Tree, and RF models, respectively (Figure 14b).

The present ANFIS-CM model ($RMSE = 8.64 \mu \text{mol m}^{-2} \text{hr}^{-1}$, $MAE = 6.95 \mu \text{mol m}^{-2} \text{hr}^{-1}$) far exceeded the accuracy of the GP, MLR, M5 Tree, and RF models (Table 5). While the MLR model's relative errors, $RRMSE$ and $RMAE$, represented only 10.0% and 4.7%, the equivalent relative errors for the GP, M5 Tree, and RF models were 26.4% and 15.4%, 40.5% and 46.8%, and 85.1% and 96.5%, respectively. In agreement with correlation coefficients of the PAR^{obs} and PAR^{for} (Table 5), the present ANFIS-CM model stood as the most accurate model when applied to hourly PAR forecasting.

The present models were compared to multilayer perceptron artificial neural network (MLP-ANN), radial basis artificial neural network (RBNN-ANN), and all-sky PAR (i.e., no cloud) models applied in China to predict hourly PAR in several ecosystems (i.e., farmland, wetland, forest, bay, grassland and desert and lake stations). These utilized global solar radiation (E_g), air temperature (T_A), relative humidity (R_H), dewpoint (T_D), water vapor pressure (V_W), air pressure (P_A), clearness index (K_τ), and the cosine of solar zenith angle (μ) as input variables (Wang et al., 2016). These data-expensive modeling methods, which use several meteorological variables to predict hourly PAR^{for} , were compared to the present ANFIS-CM model that relies on a single-input datum (θ_z). The single-input ANFIS-CM model was only marginally less accurate than the best multi-input models developed in the prior study (Wang et al., 2016); the $RMSE$ value for the ANFIS-CM model was only 4.6% greater than MLP-ANN's or the all-sky deterministic model tested in a forested location. However, the ANFIS model was tested in a regional city, and therefore, any direct comparison with literature data must be made with caution. Nonetheless, the ANFIS-CM model outperformed the different ANN models applied for hourly PAR prediction in farmlands, wetlands, and bay and grassland locations. Due to the present models' limited input data requirements, this study highlights the pre-eminence of the ANFIS-CM model over other models that utilize several meteorological input datasets which may not be available at remote locations potentially suitable for biomass production. Meteorological measurements could pose a significant logistic constraint for costly installation, maintenance and monitoring of apparatus, and reduction of sources of error in real PAR measurements (Ross & Sulev, 2000). For this reason, the main benefit of this study's approach, using the θ_z as

TABLE 5 Performance of ANFIS-CM versus comparative models (i.e., GP, MLR, M5 Tree, and RF) evaluated for the 3-day test period 29–31 December 2012 in terms of correlation coefficient (r), root-mean-square error ($RMSE$), and mean absolute error for hourly PAR forecasting, along with a comparison of multiple predictor PAR forecasting models developed by Wang et al. (2016)

Study	Model	Predictor Variable (s) ^a	Model accuracy metric			
			r	$RMSE$ (mol m ⁻² hr ⁻¹)	MAE (mol m ⁻² hr ⁻¹)	
Present Study	ANFIS-CM	θ_z	0.992	8.64	6.95	
	GP		0.944	10.92	8.02	
	MLR		0.981	9.51	7.28	
	M5 Tree		0.989	12.14	10.20	
	Random Forest		0.985	15.99	13.66	
Wang et al. (2016)	AKA station (farmland)	MLP-ANN	E_g, T_A, R_H, P_A, T_D and V_W	0.989	7.87	6.37
		all-sky PAR model (ALSKY)	K_τ, μ	0.988	16.55	13.33
	SJM station (wetland)	RBNN-ANN	E_g, T_A, R_H, P_A, T_D and V_W	0.979	18.18	11.79
		all-sky PAR model (ALSKY)	K_τ, μ	0.983	18.18	12.29
	CBF station (forest)	MLP-ANN	E_g, T_A, R_H, P_A, T_D and V_W	0.993	8.23	5.63
		all-sky PAR model (ALSKY)	K_τ, μ	0.993	8.23	5.63
	JBZ station (bay)	MLP-ANN	E_g and V_W	0.991	11.07	7.840
		all-sky PAR model (ALSKY)	K_τ, μ	0.991	11.07	8.30
	NMG station (grasslands)	MLP-ANN	E_g, T_A, R_H, P_A, T_D and V_W	0.990	11.18	8.65
		all-sky PAR model (ALSKY)	K_τ, μ	0.984	11.18	8.29

^aGlobal solar radiation (E_g), air temperature (T_A), relative humidity (R_H), dewpoint (T_D), water vapor pressure (V_W) and air pressure (P_A), clearness index (K_τ), cosine of solar angle (μ).

The best model is boldfaced

a predictor, is the larger adaptability and portability of the models if utilized over complex geographic terrains and data-scarce regions. Given the adequate performance for hourly forecasting, the present ANFIS-CM model could potentially be improved by employing available meteorological-based input data in its development, validation, and testing phases.

5 | FURTHER DISCUSSION, LIMITATIONS, AND FUTURE WORK

To promote green energy while addressing environmental, social, and economic issues associated with climate change, the utilization of renewable energy is receiving increased attention in many parts of the world. In accordance with REN21, a multi-stakeholder network collectively sharing insight and knowledge to generate the *Renewables Global Status Report*, bioenergy utilization has continued to increase in an effort to meet rising global energy demands while supporting environmental sustainability (REN 2016). Notably, bio-power generation has risen roughly 8%

annually, primarily in China, Japan, Germany, and the United Kingdom. Ethanol production has increased by 4% globally, with record levels being achieved in the USA and Brazil. Other global bio-investments, including those to encourage the production of fuels by hydrogenating vegetable oils, have increased significantly in 2015 (UPM 2016). Tall oil production in New Zealand has recently commenced (NZIC 2016). Total, a French energy corporation, has invested USD 220 million to convert the La Mède oil refinery into a bio-refinery for renewable diesel from feedstock (Total 2016). In the present study region of sub-tropical QLD, a 10-year action plan has been established to support biomass initiatives. The state is home to many bio-energy projects such as the Mackay Renewable Bio-commodities Pilot Plant. It is a pilot-scale development with an integrated bio-refinery for conversion of cellulosic biomass into bioethanol and other high-value bio-commodities. It is also home to the Solar Bio Fuels Research Centre, a research, development, and demonstration plant where macro-algal biomass serves as feedstock for renewable fuel generation (State of Queensland, 2016). In accordance with government projections, by 2035, the

industrial bio-products sector could contribute \$1.8 billion to QLD's annual gross state product and support approximately 6,640 full-time jobs. Given the requirement of reliable predicted PAR at bio-investment sites, forecasting models are likely to become important tools for energy policy analysts, government, and non-government organizations.

The models presented represent an advancement in terms of methodological development and feasibility of creating data-inexpensive models for forecasting near real-time PAR. The ANFIS-CM model validated in this research represents an initial attempt to apply an integrated hybrid neural network with fuzzy inference system to achieve a high level of accuracy for 5-min and hourly PAR predictions. While this study adds to existing research utilizing data-driven models largely restricted to artificial neural network models (e.g., (Jacovides et al., 2015; Lopez et al., 2001; Wang, Gong, Lin, & Hu, 2014; Yu & Guo, 2016)), it is unique in employing minimal input data and an easily acquired solar parameter (θ_z) calculable for any location via well-known algorithms (e.g., (Michalsky, 1988)). Thus, this study is both a complementary contribution and a significant advancement to previous studies (e.g., (Lopez et al., 2001; Wang et al., 2016; Yu & Guo, 2016)). As the newly developed models could be applied to several sites identified for solar and bioenergy investment without the need for extensive meteorological datasets, they could additionally serve to predict very short-term and hourly PAR for energy and climate change studies. In terms of practical usage, these models which have a high predictive accuracy in generating short-term (i.e., 5 min and hourly) PPFD forecasts offer a new opportunity for continuous monitoring of PPFD availability and plants' biochemical processes (including photosynthesis by biofuel generation plants and algae) in sub-tropical regions with less cloud cover.

Studies have attempted to measure PPFD to investigate its role in carboxylation efficiency (Weber, Tenhunen, Gates, & Lange, 1987), photosynthetic rates, stomatal conductance, internal CO₂ concentration, and transpiration rates (Baligar, Bunce, Elson, & Fageria, 2012). Given the quantum nature of the photosynthetic light reactions, biologists may be interested in studying real-time changes in PPFD (e.g., (McCree, 1971)). Empirical and parametric models, which are more mathematically complex and less adaptable than data-driven models, have been developed to enable the monitoring of photosynthetic activity from available PPFD data (Ritchie, 2010). Other than the short-term monitoring of solar energy by plant photosynthetic processes, the ANFIS-CM model used for hourly PPFD predictions (resulting in less than 20% error) can be developed as an add-on for agricultural applications for monitoring of PPFD diurnal changes on a local scale (Ge et al., 2011; Warner & Caldwell, 1983).

Since PPFD, a short-term component of solar radiation, was measured and accurately forecasted in this study, a follow-up study could apply the ANFIS-CM model to

forecasting longer-term solar radiation components such as the daily light integral. The total energy per day could be modeled to examine a plant's productivity over longer-term seasonal and yearly time frames (Fausey, Heins, & Cameron, 2005). Although the ANFIS-CM model was efficient in forecasting PPFD over 5-min and hourly horizons with limited predictor data, the predictive accuracy could be improved by incorporating a more extensive set of data (including climate-based inputs), if available.

Publicly available satellite and reanalysis data, for example, may have an important role in the future for greater fine-tuning of data-driven approaches to predict PAR. Ground-based climate input acquisition, such as ones discussed in earlier studies (e.g., (Wang et al., 2016)), requires expensive investments for weather monitoring apparatus. To address such limitations, in a follow-up study, the presented models could be trained on freely available climate data from the European Centre for Medium Range Weather Forecasting (ECMWF) at 6-hr sample periods (Dee et al., 2011), or twice daily remotely sensed inputs sourced from moderate resolution imaging spectroradiometer (MODIS) (Deo & Sahin, 2017; Frouin & Pinker, 1995). Such data could be used for longer-term prediction of PAR over daily, monthly, and seasonal forecasting horizons which can provide insight to unpredictable local climate patterns, not presently captured in current global climate models (GCMs).

The exclusion of crucial variables (such as cloud cover, water vapor, and aerosol content) that act to attenuate the surface availability of incident solar radiation, could lead to limited accuracy of PAR forecasts; however, in our study region, the unavailability of such atmospheric inputs at the prescribed 5-min and hourly forecasting horizons presented an insurmountable limitation. Nevertheless, in a follow-up study, the inclusion of such data from sources such as the ECMWF and the MODIS satellite-based estimates could be employed to improve the accuracy of the ANFIS model, mainly for forecasts obtained between 0900H and 1600H (e.g., Figure 8).

The present study, the first of its kind to forecast surface level PAR without the use of climate data, was limited to a single sub-tropical location. However, the usage and wider adoption of the preferred ANFIS-CM method will require further independent studies for climatologically and geographically diverse regions based on their limited site characteristics or available temporal dataset. One particular climate input necessary to increase the credibility and to evaluate the greater value of the ANFIS method is to use cloud cover, a factor that varies significantly from region to region. It is therefore important to formulate a new study to develop and evaluate a range of data-driven models (with cloud covers and other attenuating or amplifying factors of PAR) using the ANFIS modeling framework to test its viability for broader applications especially in diverse climatic regions. Besides modeling land surface PAR for solar energy and climate applications, the

scope of the present ANFIS-CM model could be expanded to other marine applications (e.g., for regular monitoring of coral bleaching affected by changes in water temperature and PAR) using instantaneous and daily PAR data sourced from local and remote platforms including MODIS. Such datasets may represent total PAR incidents on the ocean surface at the time of the satellite footprint (MODIS 2016; NASA 2014). Hence, this study provides a new framework for forthcoming investigations that could apply the ANFIS-CM model to both land and marine sites where the solar potential is relatively high and climate change impacts on marine ecosystems require investigation. Therefore, the practical relevance of the present data-driven models can be extended to more than just sustainable energy applications.

6 | CONCLUSIONS

Using a set of limited input data defined by θ_z , the present paper established the utility of the Adaptive Neuro-Fuzzy Inference System (ANFIS) algorithm for modeling very short-term (5-min) and hourly values of PAR in a sub-tropical regional location in Australia. Five-min interval θ_z , derived from an empirical algorithm (Michalsky, 1988), and measured PAR data acquired as part of the University of Southern Queensland's Atmospheric and Ultraviolet Monitoring Program, served to develop an ANFIS and other distinct data-driven models (i.e., GP, M5 Tree, RF, and MLR) to serve as benchmarks. Models were evaluated through diagnostic plots of PAR^{obs} vs. PAR^{for} , $RMSE$, MAE , and their percentage equivalents $RRMS$, $RMAE$, E_{NS} , WI , and E_{LM} .

The study's findings can be enumerated as follows.

1. For a very short-term forecasting horizon, the performance of the ANFIS-CM and the GP model remained similar in terms of the correlation coefficients for the linear relationship between PAR^{obs} and PAR^{for} over the test period ($r \approx 0.978$), which concurred with the relatively large values of the E_{NS} , WI , and E_{LM} . The ANFIS-CM and GP models achieved a $WI \approx 0.953$, whereas values of 0.921, 0.932, and 0.883 were obtained for the MLR, M5 Tree and RF models, respectively. In the testing phase, the ANFIS-CM model's $RMAE$ was found to be only marginally better the GP model (i.e., 17.18% vs. 20.03%), but significantly better than the MLR (26.56%), M5 Tree (25.11%), and RF models (34.55%).
2. When the ANFIS-CM model was evaluated for an hourly forecasting horizon, its accuracy was superior than all other data-driven models tested, yielding a value $r \approx 0.992$ vs. 0.944, 0.981, 0.989, 0.985 for the GP, MLR, M5 Tree, and RF models, respectively. This was confirmed by the ANFIS-CM model's larger E_{LM} value of 0.7405 vs.

0.7047, 0.7282, 0.6192, and 0.4899 for the GP, MLR, M5 Tree, and RF models, respectively. It was also paralleled by similarly large values of the E_{NS} and WI assessing the closeness of PAR^{obs} and PAR^{for} data.

Overall, this pilot study highlights the appropriateness of using an ANFIS-CM model for very short-term and hourly forecasting of PAR without the need for meteorological-based inputs. This confers an advantage on the model for use in data sparse regions. However, given the contributory influence of atmospheric attenuating factors (e.g., cloud cover, water vapor, and aerosol), the results of our study provide an added opportunity to incorporate their influences and improve the performance of the ANFIS-CM model in different climates and seasons. Such improvements could be implemented on hourly, daily, or seasonal scales for which atmospheric data could be sourced from satellites and observed-modeled repositories (e.g., ECMWF fields). Despite the provision of this research paper providing baseline relevance of the ANFIS-CM model for PAR forecasting, the testing of the model in a wide range of remote and regional sites where solar energy, agricultural, and bio-investment works are targeted should be performed to confirm its practicality.

ACKNOWLEDGMENTS

We sincerely acknowledge The University of Southern Queensland Academic Division for providing funding through the Academic Development and Outside Studies Program (ADOSP) to Dr. Ravinesh C Deo and Dr. Nathan Downs. We thank reviewers and the journal's handling editor for constructive criticisms that have improved the clarity of this paper. The authors greatly acknowledge USQ Centre for Applied Climate Sciences for supporting the publication fees.

ORCID

Ravinesh C. Deo  <http://orcid.org/0000-0002-2290-6749>

REFERENCES

- Abdulshahed, A., Longstaff, A.P., Fletcher, S., & Myers, A. (2013). Comparative study of ANN and ANFIS prediction models for thermal error compensation on CNC machine tools. In, Lamdamap 10th International Conference (pp. 79–89): EUSPEN
- Abdulshahed, A. M., Longstaff, A. P., Fletcher, S., & Myers, A. (2015). Thermal error modelling of machine tools based on ANFIS with fuzzy c-means clustering using a thermal imaging camera. *Applied Mathematical Modelling*, 39, 1837–1852. <https://doi.org/10.1016/j.apm.2014.10.016>
- Akgul, O., Shah, N., & Papageorgiou, L. G. (2012). Economic optimisation of a UK advanced biofuel supply chain. *Biomass and Bioenergy*, 41, 57–72. <https://doi.org/10.1016/j.biombioe.2012.01.040>

- Ali, M., Deo, R. C., Downs, N. J., & Maraseni, T. (2018). An ensemble-ANFIS based uncertainty assessment model for forecasting multi-scalar standardized precipitation index. *Atmospheric Research*, 207, 155–180. <https://doi.org/10.1016/j.atmosres.2018.02.024>
- Bahadori, A., & Nwaoha, C. (2013). A review on solar energy utilisation in Australia. *Renewable and Sustainable Energy Reviews*, 18, 1–5. <https://doi.org/10.1016/j.rser.2012.10.003>
- Baligar, V., Bunce, J., Elson, M., & Fageria, N. (2012). Photosynthetic photon flux density, carbon dioxide concentration and temperature influence photosynthesis in *Crotalaria* species. *The Open Plant Science Journal*, 6, 1–7. <https://doi.org/10.2174/1874294701206010001>
- Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Berlin, Germany: Springer Science & Business Media.
- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10, 191–203. [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7)
- Bhattacharya, B., & Solomatine, D. P. (2005). Neural networks and M5 model trees in modelling water level–discharge relationship. *Neurocomputing*, 63, 381–396. <https://doi.org/10.1016/j.neucom.2004.04.016>
- Breiman, L. (2001). Random forests Machine Learning 45: 5–32. Figure Titles and Legends described only once and is significantly higher in sporadic ccRCC compared with other somatic tumour types. long-range and short-range interaction, such as mutations at, 79, R82
- Brennan, L., & Owende, P. (2010). Biofuels from microalgae—a review of technologies for production, processing, and extractions of bio-fuels and co-products. *Renewable and Sustainable Energy Reviews*, 14, 557–577. <https://doi.org/10.1016/j.rser.2009.10.009>
- Bylander, T. (2002). Estimating generalization error on two-class datasets using out-of-bag estimates. *Machine Learning*, 48, 287–297. <https://doi.org/10.1023/A:1013964023376>
- Chaabene, M., & Ammar, M. B. (2008). Neuro-fuzzy dynamic model with Kalman filter to forecast irradiance and temperature for solar energy systems. *Renewable Energy*, 33, 1435–1443. <https://doi.org/10.1016/j.renene.2007.10.004>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7, 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chattopadhyay, S., Pratihari, D. K., & De Sarkar, S. C. (2011). A comparative study of fuzzy c-means algorithm and entropy-based fuzzy clustering algorithms. *Computing and Informatics*, 30, 701–720.
- Civelekoglu, G., Yigit, N., Diamadopoulos, E., & Kitis, M. (2007). Prediction of bromate formation using multi-linear regression and artificial neural networks. *Ozone: Science and Engineering*, 29, 353–362. <https://doi.org/10.1080/01919510701549327>
- Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., ... Brown, P. (2005). Atmospheric radiative transfer modeling: A summary of the AER codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 91, 233–244. <https://doi.org/10.1016/j.jqsrt.2004.05.058>
- Collazo-Cuevas, J.I., Aceves-Fernandez, M.A., Gorrostieta-Hurtado, E., Pedraza-Ortega, J., Sotomayor-Olmedo, A., & Delgado-Rosas, M. (2010). Comparison between Fuzzy C-means clustering and Fuzzy Clustering Subtractive in urban air pollution. In, Electronics, Communications and Computer (CONIELECOMP), 2010 20th International Conference on (pp. 174–179): IEEE.
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88, 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Dawson, C. W., Abrahart, R. J., & See, L. M. (2007). HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling & Software*, 22, 1034–1052. <https://doi.org/10.1016/j.envsoft.2006.06.008>
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., ... Bauer, P. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. <https://doi.org/10.1002/qj.828>
- Deo, R. C., Downs, N., Parisi, A., Adamowski, J., & Quilty, J. (2017). Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. *Environmental Research*, 155, 141–166. <https://doi.org/10.1016/j.envres.2017.01.035>
- Deo, R. C., & Sahin, M. (2017). Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renewable and Sustainable Energy Reviews*, 72, 828–848. <https://doi.org/10.1016/j.rser.2017.01.114>
- Deo, R. C., & Şahin, M. (2015). Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. *Atmospheric Research*, 161–162, 65–81. <https://doi.org/10.1016/j.atmosres.2015.03.018>
- Deo, R. C., & Samui, P. (2017). Forecasting evaporative loss by least-square support-vector regression and evaluation with genetic programming, Gaussian process, and minimax probability machine regression: Case study of Brisbane City. *Journal of Hydrologic Engineering*, 22, 05017003. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001506](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001506)
- Deo, R. C., Wen, X., & Feng, Q. (2016). A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Applied Energy*, 168, 568–593. <https://doi.org/10.1016/j.apenergy.2016.01.130>
- Dokur, E., Kurban, M., & Ceyhan, S. (2015). Hybrid model for short term wind speed forecasting using empirical mode decomposition and artificial neural network. In, Electrical and Electronics Engineering (ELECO), 2015 9th International Conference on (pp. 420–423): IEEE.
- Draper, N., & Smith, H. (1981). *Applied regression analysis*. (709 pp). New York: John Wiley.
- Dunn, J.C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3, (3).
- Dupraz, C., Marrou, H., Talbot, G., Dufour, L., Nogier, A., & Ferard, Y. (2011). Combining solar photovoltaic panels and food crops for optimising land use: Towards new agrivoltaic schemes. *Renewable Energy*, 36, 2725–2732. <https://doi.org/10.1016/j.renene.2011.03.005>
- EC. (2017). European Commission (EC) 2009, Directive 2009/28/EC of the European Parliament and the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC. Official Journal of the European Union. pp. L140/16–L140/62. Retrieved from <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0028&from=EN>, Accessed 14 June 2107.

- EIC. (2017). European Commission (EC): Energy 2017.. In. National Action Plans, Retrieved from <https://ec.europa.eu/energy/en/topics/renewable-energy/national-action-plans>, Accessed 13 June 2017.
- Eltbaakh, Y. A., Ruslan, M. H., Alghoul, M., Othman, M. Y., Sopian, K., & Fadhel, M. (2011). Measurement of total and spectral solar irradiance: Overview of existing research. *Renewable and Sustainable Energy Reviews*, *15*, 1403–1426. <https://doi.org/10.1016/j.rser.2010.10.018>
- Escobedo, J. F., Gomes, E. N., Oliveira, A. P., & Soares, J. (2011). Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. *Renewable Energy*, *36*, 169–178. <https://doi.org/10.1016/j.renene.2010.06.018>
- Fausey, B. A., Heins, R. D., & Cameron, A. C. (2005). Daily light integral affects flowering and quality of greenhouse-grown *Achillea*, *Gaura*, and *Lavandula*. *HortScience*, *40*, 114–118.
- Fernández-Martínez, M., Vicca, S., Janssens, I. A., Luyssaert, S., Campioli, M., Sardans, J., ... Peñuelas, J. (2014). Spatial variability and controls over biomass stocks, carbon fluxes, and resource-use efficiencies across forest ecosystems. *Trees*, *28*, 597–611. <https://doi.org/10.1007/s00468-013-0975-9>
- Frouin, R., & Pinker, R. T. (1995). Estimating photosynthetically active radiation (PAR) at the earth's surface from satellite observations. *Remote Sensing of Environment*, *51*, 98–107. [https://doi.org/10.1016/0034-4257\(94\)00068-X](https://doi.org/10.1016/0034-4257(94)00068-X)
- Fu, P., & Rich, P. M. (2002). A geometric solar radiation model with applications in agriculture and forestry. *Computers and Electronics in Agriculture*, *37*, 25–35. [https://doi.org/10.1016/S0168-1699\(02\)00115-1](https://doi.org/10.1016/S0168-1699(02)00115-1)
- Gandomi, A. H., & Alavi, A. H. (2012). A new multi-gene genetic programming approach to nonlinear system modeling. Part I: Materials and structural engineering problems. *Neural Computing and Applications*, *21*, 171–187. <https://doi.org/10.1007/s00521-011-0734-z>
- Ge, S., Smith, R. G., Jacovides, C. P., Kramer, M. G., & Carruthers, R. I. (2011). Dynamics of photosynthetic photon flux density (PPFD) and estimates in coastal northern California. *Theoretical and Applied Climatology*, *105*, 107–118. <https://doi.org/10.1007/s00704-010-0368-6>
- Ghimire, S., Deo, R. C., Downs, N. J., & Raj, N. (2018). 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. *Remote Sensing of Environment*, *212*, 176–198. <https://doi.org/10.1016/j.rse.2018.05.003>
- Ghorbani, M., Khatibi, R., Hosseini, B., & Bilgili, M. (2013). Relative importance of parameters affecting wind speed prediction using artificial neural networks. *Theoretical and Applied Climatology*, *114*, 107–114. <https://doi.org/10.1007/s00704-012-0821-9>
- Gill, D., Ming, T., & Ouyang, W. (2017). Improving the Lake Erie HAB Tracker: A Forecasting & Decision Support Tool for Harmful Algal Blooms. Masters Project. Online: <https://deepblue.lib.umich.edu/handle/2027.42/136562?show=full>
- Guillaume, S. (2001). Designing fuzzy inference systems from data: An interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, *9*, 426–443. <https://doi.org/10.1109/91.928739>
- Guofeng, W., Leeuw, J., Skidmore, A. K., Yaolin, L., & Prins, H. H. (2010). Comparison of Extrapolation and Interpolation methods for estimating daily photo synthetically active radiation (PAR)—a case study of the Poyang lake national nature reserve, China. *Geo-Spatial Information Science*, *13*, 235–242.
- Güven, A., & Kışi, Ö. (2011). Daily pan evaporation modeling using linear genetic programming technique. *Irrigation Science*, *29*, 135–145. <https://doi.org/10.1007/s00271-010-0225-5>
- Hetherington, A. M., & Woodward, F. I. (2003). The role of stomata in sensing and driving environmental change. *Nature*, *424*, 901–908. <https://doi.org/10.1038/nature01843>
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector classification. <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- Hu, J., Wang, J., & Zeng, G. (2013). A hybrid forecasting approach applied to wind speed time series. *Renewable Energy*, *60*, 185–194. <https://doi.org/10.1016/j.renene.2013.05.012>
- IPCC (2007). Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change United Kingdom and New York, NY, USA: Cambridge University Press.
- Islam, M. S., Mohandes, M., & Rehman, S. (2016). Vertical extrapolation of wind speed using artificial neural network hybrid system. *Neural Computing and Applications*, *28*, 2369.
- Jacovides, C., Tymvios, F., Bolland, J., & Tsitouri, M. (2015). Artificial Neural Network models for estimating daily solar global UV, PAR and broadband radiant fluxes in an eastern Mediterranean site. *Atmospheric Research*, *152*, 138–145. <https://doi.org/10.1016/j.atmosres.2013.11.004>
- Jang, J.-S., & Sun, C.-T. (1995). Neuro-fuzzy modeling and control. *Proceedings of the IEEE*, *83*, 378–406. <https://doi.org/10.1109/5.364486>
- Johnson, D., Thomas, T., Heinicke, D., Peterson, R., Morgan, P., McDermitt, D., & Burba, G. (2015). A New Quantum Sensor for Measuring Photosynthetically Active Radiation. In, *AGU Fall Meeting Abstracts*.
- Kaur, A., & Kaur, A. (2012). Comparison of mamdani-type and sugeno-type fuzzy inference systems for air conditioning system. *International Journal of Soft Computing and Engineering (IJSCE)*, *2*, 323–325.
- Khatib, T., Mohamed, A., & Sopian, K. (2012). A review of solar energy modeling techniques. *Renewable and Sustainable Energy Reviews*, *16*, 2864–2869. <https://doi.org/10.1016/j.rser.2012.01.064>
- Kisi, O. (2013). Applicability of Mamdani and Sugeno fuzzy genetic approaches for modeling reference evapotranspiration. *Journal of Hydrology*, *504*, 160–170. <https://doi.org/10.1016/j.jhydrol.2013.09.043>
- Kisi, O. (2015). Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology*, *528*, 312–320. <https://doi.org/10.1016/j.jhydrol.2015.06.052>
- Koza, J.R. (1992). *Genetic programming: on the programming of computers by means of natural selection*. MIT press. Statistics and Computing. *4*(2), 87–112.
- Krause, P., Boyle, D., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97. <https://doi.org/10.5194/adgeo-5-89-2005>
- Kumar, S., Shrestha, P., & Salam, P. A. (2013). A review of biofuel policies in the major biofuel producing countries of ASEAN: Production, targets, policy drivers and impacts. *Renewable and Sustainable Energy Reviews*, *26*, 822–836. <https://doi.org/10.1016/j.rser.2013.06.007>

- Kurtulus, B., & Razack, M. (2010). Modeling daily discharge responses of a large karstic aquifer using soft computing methods: Artificial neural network and neuro-fuzzy. *Journal of Hydrology*, *381*, 101–111. <https://doi.org/10.1016/j.jhydrol.2009.11.029>
- Lane, J. (2016). “Biofuels mandates around the world:2016”, Biofuels Digest. In. January 2016, Retrieved from: <http://www.biofuels-digest.com/bdigest/2016/01/03/biofuels-mandates-around-the-world-2016/>, Accessed: 14 June 2017.
- Legates, D. R., & Davis, R. E. (1997). The continuing search for an anthropogenic climate change signal: Limitations of correlation-based approaches. *Geophysical Research Letters*, *24*, 2319–2322. <https://doi.org/10.1029/97GL02207>
- Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*, 233–241. <https://doi.org/10.1029/1998WR900018>
- Legates, D. R., & McCabe, G. J. (2013). A refined index of model performance: A rejoinder. *International Journal of Climatology*, *33*, 1053–1056. <https://doi.org/10.1002/joc.3487>
- Liu, B. Y., & Jordan, R. C. (1960). The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy*, *4*, 1–19. [https://doi.org/10.1016/0038-092X\(60\)90062-1](https://doi.org/10.1016/0038-092X(60)90062-1)
- Liu, R., Liang, S., He, H., Liu, J., & Zheng, T. (2008). Mapping incident photosynthetically active radiation from MODIS data over China. *Remote Sensing of Environment*, *112*, 998–1009.
- Lopez, G., Rubio, M., Martinez, M., & Batlles, F. (2001). Estimation of hourly global photosynthetically active radiation using artificial neural network models. *Agricultural and Forest Meteorology*, *107*, 279–291. [https://doi.org/10.1016/S0168-1923\(01\)00217-9](https://doi.org/10.1016/S0168-1923(01)00217-9)
- Mamdani, E. H. (1976). Advances in the linguistic synthesis of fuzzy controllers. *International Journal of Man-Machine Studies*, *8*, 669–678. [https://doi.org/10.1016/S0020-7373\(76\)80028-4](https://doi.org/10.1016/S0020-7373(76)80028-4)
- Martin, N. J., & Rice, J. L. (2012). Developing renewable energy supply in Queensland, Australia: A study of the barriers, targets, policies and actions. *Renewable Energy*, *44*, 119–127. <https://doi.org/10.1016/j.renene.2012.01.006>
- McCree, K. J. (1971). The action spectrum, absorptance and quantum yield of photosynthesis in crop plants. *Agricultural Meteorology*, *9*, 191–216. [https://doi.org/10.1016/0002-1571\(71\)90022-7](https://doi.org/10.1016/0002-1571(71)90022-7)
- Mehr, A. D., Kahya, E., & Olyala, E. (2013). Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *Journal of Hydrology*, *505*, 240–249. <https://doi.org/10.1016/j.jhydrol.2013.10.003>
- Mellit, A., Arab, A.H., Khorissi, N., & Salhi, H. (2007). An ANFIS-based forecasting for solar radiation data from sunshine duration and ambient temperature. In, Power Engineering Society General Meeting, 2007. IEEE (pp. 1–6): IEEE
- Mellit, A., & Kalogirou, S. A. (2011). ANFIS-based modelling for photovoltaic power supply system: A case study. *Renewable Energy*, *36*, 250–258. <https://doi.org/10.1016/j.renene.2010.06.028>
- Michalsky, J. J. (1988). The astronomical almanac’s algorithm for approximate solar position (1950–2050). *Solar Energy*, *40*, 227–235. [https://doi.org/10.1016/0038-092X\(88\)90045-X](https://doi.org/10.1016/0038-092X(88)90045-X)
- Mitchell, T.M. (1997). Machine learning. Computer Science Series (McGraw-Hill, Burr Ridge, 1997) MATH
- MODIS. (2016). MODIS (Moderate-Resolution Imaging Spectroradiometer). In NASA (Ed.). Retrieved from: http://modis.gsfc.nasa.gov/about/media/modis_brochure.pdf: NASA
- Mohammadi, K., Shamshirband, S., Tong, C. W., Arif, M., Petković, D., & Ch, S. (2015). A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. *Energy Conversion and Management*, *92*, 162–171. <https://doi.org/10.1016/j.enconman.2014.12.050>
- Monteith, J. L., & Moss, C. (1977). Climate and the efficiency of crop production in Britain [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *281*, 277–294. <https://doi.org/10.1098/rstb.1977.0140>
- Montgomery, D.C., Peck, E.A., & Vining, G.G. (2012). *Introduction to linear regression analysis*. (5th Ed). John Wiley & Sons.
- Moore, D.S., Notz, W.I., & Notz, W. (2006). *Statistics: Concepts and controversies*. New York: Macmillan.
- Muir, P. R., Wallace, C. C., Done, T., & Aguirre, J. D. (2015). Limited scope for latitudinal extension of reef corals. *Science*, *348*, 1135–1138. <https://doi.org/10.1126/science.1259911>
- Myers, R. H. (2000). *Classical and modern regression with applications (Duxbury Classic)*. Pacific Grove: Duxbury Press.
- NASA. (2014). NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group; MODIS-Terra Ocean Color Data. In. http://dx.doi.org/10.5067/TERRA/MODIS_OC.2014.0: NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group.
- Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- NZIC. (2016). Tall oil production and processing. Retrieved from <http://nzic.org.nz/ChemProcesses/forestry/4G.pdf>. New Zealand: New Zealand Institute of Chemistry
- Ooms, M. D., Dinh, C. T., Sargent, E. H., & Sinton, D. (2016). Photon management for augmented photosynthesis. *Nature Communications*, *7*, 12699. <https://doi.org/10.1038/ncomms12699>
- Ozdamar, K. (2004). *The Statistical Data Analysis with Software Packages*. Eskisehir: Kaan press.
- Parisi, A., Wong, J., & Randall, C. (1998). Simultaneous assessment of photosynthetically active and ultraviolet solar radiation. *Agricultural and Forest Meteorology*, *92*, 97–103. [https://doi.org/10.1016/S0168-1923\(98\)00094-X](https://doi.org/10.1016/S0168-1923(98)00094-X)
- Park, S.-H., Kim, S.-J., Lim, K.-J., & Kang, S.-H. (2005). Comparison of recognition rates between BP and ANFIS with FCM clustering method on off-line PD diagnosis of defect models of traction motor stator coil. In, Electrical Insulating Materials, 2005. (ISEIM 2005). Proceedings of 2005 International Symposium on (pp. 849–852): IEEE
- Poli, R., Langdon, W.B., McPhee, N.F., & Koza, J.R. (2008). A field guide to genetic programming. Lulu. com
- Prasad, R., Deo, R. C., Li, Y., & Maraseni, T. (2018). Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors. *Soil and Tillage Research*, *181*, 63–81. <https://doi.org/10.1016/j.still.2018.03.021>
- Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, *9*, 181–199. <https://doi.org/10.1007/s10021-005-0054-1>
- Qin, J., Yang, K., Liang, S., & Tang, W. (2012). Estimation of daily mean photosynthetically active radiation under all-sky conditions based on relative sunshine data. *Journal of Applied Meteorology and Climatology*, *51*, 150–160. <https://doi.org/10.1175/JAMC-D-10-05018.1>

- QREP. (2009). The Queensland Renewable Energy Plan (QREP). Brisbane: Queensland Government, The State of Queensland.
- Quinlan, J.R. (1992). Learning with continuous classes. In, *5th Australian joint conference on artificial intelligence* (pp. 343-348): Singapore. In Proceedings AI'92 (Adams & Sterling, Eds), 343-348, Singapore: World Scientific.
- Rahimikhoob, A., Asadi, M., & Mashal, M. (2013). A comparison between conventional and M5 model tree methods for converting pan evaporation to reference evapotranspiration for semi-arid region. *Water Resources Management*, *27*, 4815–4826. <https://doi.org/10.1007/s11269-013-0440-y>
- Rajagopal, D., Sexton, S. E., Roland-Holst, D., & Zilberman, D. (2007). Challenge of biofuel: Filling the tank without emptying the stomach? *Environmental Research Letters*, *2*, 044004. <https://doi.org/10.1088/1748-9326/2/4/044004>
- REN. (2016). Renewables 2016 Global Status Report: ISBN 978-3-9818107-0-7. In. REN21 Secretariat: Paris: Renewable Energy Policy Network for the 21st Century.
- Ritchie, R. (2010). Modelling photosynthetic photon flux density and maximum potential gross photosynthesis. *Photosynthetica*, *48*, 596–609. <https://doi.org/10.1007/s11099-010-0077-5>
- Ross, J., & Sulev, M. (2000). Sources of errors in measurements of PAR. *Agricultural and Forest Meteorology*, *100*, 103–125. [https://doi.org/10.1016/S0168-1923\(99\)00144-6](https://doi.org/10.1016/S0168-1923(99)00144-6)
- Rubio, M., Lopez, G., Tovar, J., Pozo, D., & Batlles, F. (2005). The use of satellite measurements to estimate photosynthetically active radiation. *Physics and Chemistry of the Earth, Parts A/B/C*, *30*, 159–164. <https://doi.org/10.1016/j.pce.2004.08.029>
- Sabburg, J.M. (2000). Quantification of cloud around the sun and its correlation with global UV measurement, PhD Thesis. In. Queensland University of Technology: Queensland University of Technology.
- Şahin, M., Kaya, Y., & Uyar, M. (2013). Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/AVHRR data. *Advances in Space Research*, *51*, 891–904. <https://doi.org/10.1016/j.asr.2012.10.010>
- Salcedo-Sanz, S., Deo, R. C., Cornejo-Bueno, L., Camacho-Gómez, C., & Ghimire, S. (2018). An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia. *Applied Energy*, *209*, 79–94. <https://doi.org/10.1016/j.apenergy.2017.10.076>
- Samui, P., & Dixon, B. (2012). Application of support vector machine and relevance vector machine to determine evaporative losses in reservoirs. *Hydrological Processes*, *26*, 1361–1369. <https://doi.org/10.1002/hyp.8278>
- Searson, D.P. (2015). GPTIPS 2: An open-source software platform for symbolic data mining. *Handbook of genetic programming applications* (pp. 551–573): Springer. <https://doi.org/10.1007/978-3-319-20883-1>
- Şen, Z. (1998). Fuzzy algorithm for estimation of solar irradiation from sunshine duration. *Solar Energy*, *63*, 39–49. [https://doi.org/10.1016/S0038-092X\(98\)00043-7](https://doi.org/10.1016/S0038-092X(98)00043-7)
- Setlak, G. (2008). The fuzzy-neuro classifier for decision support. *International Journal of Information Theory and Applications*, *15*, 21–26.
- Shafiullah, G. (2016). Hybrid renewable energy integration (HREI) system for subtropical climate in Central Queensland, Australia. *Renewable Energy*, *96*, 1034–1053. <https://doi.org/10.1016/j.renene.2016.04.101>
- Simionato, D., Basso, S., Giacometti, G. M., & Morosinotto, T. (2013). Optimization of light use efficiency for biofuel production in algae. *Biophysical Chemistry*, *182*, 71–78. <https://doi.org/10.1016/j.bpc.2013.06.017>
- Sreekanth, J., & Datta, B. (2011). Coupled simulation-optimization model for coastal aquifer management using genetic programming-based ensemble surrogate models and multiple-realization optimization. *Water Resources Research*, *47*.
- State of Queensland, (2016). Advance Queensland: Queensland Biofutures 10-Year Roadmap and Action Plan. City East, Queensland: Queensland Government, Department of State Development.
- Sudhakar, K., Srivastava, T., Satpathy, G., & Premalatha, M. (2013). Modelling and estimation of photosynthetically active incident radiation based on global irradiance in Indian latitudes. *International Journal of Energy and Environmental Engineering*, *4*, 21. <https://doi.org/10.1186/2251-6832-4-21>
- Sugeno, M., & Kang, G. (1988). Structure identification of fuzzy model. *Fuzzy Sets and Systems*, *28*, 15–33. [https://doi.org/10.1016/0165-0114\(88\)90113-3](https://doi.org/10.1016/0165-0114(88)90113-3)
- Tabari, H., Kisi, O., Ezani, A., & Talaee, P. H. (2012a). SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology*, *444*, 78–89. <https://doi.org/10.1016/j.jhydrol.2012.04.007>
- Tabari, H., Talaee, P. H., & Abghari, H. (2012b). Utility of coactive neuro-fuzzy inference system for pan evaporation modeling in comparison with multilayer perceptron. *Meteorology and Atmospheric Physics*, *116*, 147–154. <https://doi.org/10.1007/s00703-012-0184-x>
- Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, man, and Cybernetics*, *SMC-15 (1)*, 116–132. <https://doi.org/10.1109/TSMC.1985.6313399>
- TIQ (2017). Trade and Investment Queensland Australia (TIQ). In. Queensland Australia: Bio-industries, Queensland Government, Retrieved from: <http://www.tiq.qld.gov.au/wp-content/uploads/2015/12/TIQ-755-14-Bio-Industries-Brochure-6pp-EndOpps2.pdf>, Accessed 13 June 2017
- Total (2016). La Mède: Total's first biorefinery. Retrieved from <http://www.total.com/en/energy-expertise/projects/biomass/la-mede-total-first-biorefinery>. France: TOTAL
- UPM. (2016). UPM Biofuels, “Investment in the world's first biorefinery producing wood-based diesel. Retrieved from <http://www.upm-biofuels.com/biofuel-production/biorefinery/Pages/Default.aspx>
- Vieira, J., Dias, F.M., & Mota, A. (2004). Neuro-fuzzy systems: a survey. In, 5th WSEAS NNA International Conference.
- Von Altrock, C. (1995). *Fuzzy logic and neurofuzzy applications explained*. New Jersey: Upper Saddle River. ISBN 0-13-368465-2.: Prentice Hall.
- Wang, Y.W. (1997). IH: Inducing Model Trees for Predicting Continuous Classes. In, Proceedings of European Conference on Machine Learning, University of Economics, Prague
- Wang, L., Gong, W., Feng, L., Lin, A., Hu, B., & Zhou, M. (2015). Estimation of hourly and daily photosynthetically active radiation in Inner Mongolia, China, from 1990 to 2012. *International Journal of Climatology*, *35*, 3120–3131. <https://doi.org/10.1002/joc.4197>
- Wang, L., Gong, W., Lin, A., & Hu, B. (2014). Analysis of photosynthetically active radiation under various sky conditions in Wuhan,

- Central China. *International Journal of Biometeorology*, 58, 1711–1720. <https://doi.org/10.1007/s00484-013-0775-3>
- Wang, L., Kisi, O., Zounemat-Kermani, M., Hu, B., & Gong, W. (2016). Modeling and comparison of hourly photosynthetically active radiation in different ecosystems. *Renewable and Sustainable Energy Reviews*, 56, 436–453. <https://doi.org/10.1016/j.rser.2015.11.068>
- Warner, C. W., & Caldwell, M. M. (1983). Influence of photon flux density in the 400–700 nm waveband on inhibition of photosynthesis by UV-B (280–320 nm) irradiation in soybean leaves: Separation of indirect and immediate effects. *Photochemistry and Photobiology*, 38, 341–346. <https://doi.org/10.1111/j.1751-1097.1983.tb02681.x>
- Weber, J. A., Tenhunen, J. D., Gates, D. M., & Lange, O. L. (1987). Effect of photosynthetic photon flux density on carboxylation efficiency. *Plant Physiology*, 85, 109–114. <https://doi.org/10.1104/pp.85.1.109>
- WEC (2017). World Energy Council (WEC), 2010. Biofuels: Policies, Standards and Technologies, World Energy Council 2010. Retrieved from https://www.worldenergy.org/wp-content/uploads/2012/10/PUB_Biofuels_Policies_Standards_and_Technologies_2010_WEC.pdf, Accessed 13 June 2017.
- WEF. (2010). The future of industrial biorefineries. In. Retrieved from: http://www3.weforum.org/docs/WEF_FutureIndustrialBiorefineries_Report_2010.pdf, Accessed 13 June 2017. World Economic Forum, Switzerland.
- WEF. (2017). World Economic Forum (WEF) 2017. “These countries produce the most biofuels”. In. Retrieved from <https://www.weforum.org/agenda/2015/11/these-countries-produce-the-most-biofuels/>, Accessed 13 June 2017.
- Wen, X., Feng, Q., Yu, H., Wu, J., Si, J., Chang, Z., & Xi, H. (2015). Wavelet and adaptive neuro-fuzzy inference system conjunction model for groundwater level predicting in a coastal aquifer. *Neural Computing and Applications*, 26, 1203–1215. <https://doi.org/10.1007/s00521-014-1794-7>
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2, 184–194.
- Willmott, C.J. (1984). On the evaluation of model performance in physical geography. *Spatial statistics and models* (pp. 443–460): Springer <https://doi.org/10.1007/978-94-017-3048-8>
- Witten, I.H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann
- Xia, X., Li, Z., Wang, P., Cribb, M., Chen, H., & Zhao, Y. (2008). Analysis of photosynthetic photon flux density and its parameterization in Northern China. *Agricultural and Forest Meteorology*, 148, 1101–1108. <https://doi.org/10.1016/j.agrformet.2008.02.008>
- Yaseen, Z. M., Ghareb, M. I., Ebtehaj, I., Bonakdari, H., Siddique, R., Heddam, S., ... Deo, R. (2018). Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA. *Water Resource Management*, 32, 105–122. <https://doi.org/10.1007/s11269-017-1797-0>
- Yu, X., & Guo, X. (2016). Hourly photosynthetically active radiation estimation in Midwestern United States from artificial neural networks and conventional regressions models. *International Journal of Biometeorology*, 60, 1247–1259. <https://doi.org/10.1007/s00484-015-1120-9>
- L. A. Zadeh, (1965). “Fuzzy sets,” *Information Control*, 32, 338–353
- Zaher, H., Kandil, A. E., & Fahmy, R. (2014). Comparison of Mamdani and Sugeno Fuzzy Inference Systems for prediction (with application to prices of Fund in Egypt). *British Journal of Mathematics & Computer Science*, 4, 3014–3022. <https://doi.org/10.9734/BJMCS>

How to cite this article: Deo RC, Downs NJ, Adamowski JF, Parisi AV. Adaptive Neuro-Fuzzy Inference System integrated with solar zenith angle for forecasting sub-tropical Photosynthetically Active Radiation. *Food Energy Secur.* 2018;e151. <https://doi.org/10.1002/fes3.151>