

Research paper

Daily global solar radiation time series prediction using variational mode decomposition combined with multi-functional recurrent fuzzy neural network and quantile regression forests algorithm

Mohammed Abdallah^{a,b}, Babak Mohammadi^{c,*}, Hamid Nasiri^d, Okan Mert Katipoğlu^e,
Modawy Adam Ali Abdalla^{f,g}, Mohammad Mehdi Ebadzadeh^d

^a College of Hydrology and Water Resources, Hohai University, Nanjing, Jiangsu 210024, China

^b The Hydraulics Research Station, PO Box 318, Wad Medani, Sudan

^c Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden

^d Department of Computer Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

^e Department of Civil Engineering, Erzincan Binali Yıldırım University, Erzincan, Turkey

^f College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China

^g Department of Electrical and Electronic Engineering, College of Engineering Science, Nyala University, Nyala 63311, Sudan

ARTICLE INFO

Keywords:

Global solar radiation
Recurrent fuzzy neural network
Quantile regression forests
Meteorological factor

ABSTRACT

Global solar radiation (GSR) prediction capability with a reliable model and high accuracy is crucial for comprehending hydrological and meteorological systems. It is vital for the production of renewable and clean energy. This research aims to evaluate the performance of combined variational mode decomposition (VMD) with a multi-functional recurrent fuzzy neural network (MFRFNN) and quantile regression forests (QRF) models for GSR prediction in daily scales. The hybrid VMD-MFRFNN and QRF models were compared with standalone MFRFNN, random forest (RF), extreme gradient boosting (XGB), and M5 tree (M5T) models across the Lund and Växjö meteorological stations in Sweden. The meteorological data from 2008 to 2017 were used to train the models, while the prediction accuracy was verified by using the data from 2018 to 2021 under five different input combinations. The various meteorological-based scenarios (including the input are air temperatures (Tmin, Tmax, T), wind speed (WS), relative humidity (RH), sunshine duration (SSH), and maximum possible sunshine duration (N)) were considered as input of predictor models. The current study resulted that the M5T model exhibited higher accuracy than RF and XGB models, while the QRF model showed equivalent performance with the M5T model at both study sites. The MFRFNN model outperformed QRF and M5T models across all input combinations at both study sites. The hybrid VMD-MFRFNN model showed the best performance when fewer input variables (Tmin, Tmax, T, WS at Lund station and Tmin, Tmax, T, WS, SSH, RH at Växjö station) were used for GSR prediction. We conclude that the MFRFNN model best predicts average daily GSR when combining all meteorological variables (Tmin, Tmax, T, WS, SSH, RH, N).

1. Introduction

One of the primary sources of the clean and most cheap renewable energy that enters the earth's surface is GSR, which plays a significant part in maintaining the surface-atmosphere system. GSR is essential for many solar purposes, including photovoltaic devices, solar-powered water heaters, sun drying, and architectural design (Muneer et al., 2005). Furthermore, GSR is regarded as one of the most critical variables in hydrology applications (Yang et al., 2006; Mohammadi and

Mehdizadeh, 2020). For example, the method of predict GSR significantly produced different crop yield of wheat and maize yields (Wang et al., 2015). The solar radiation components control the accuracy of reference evapotranspiration estimation across Iran (Mehdizadeh et al., 2021).

The GSR on a horizontal surface is measured with a pyranometer and blended throughout time (e.g., hour, day). But because of the limits of costs, materials, technologies, and environment, GSR measurements are not easily accessible in most places around the world, which negatively impacts research activities that rely on it (Fan et al., 2019b; Abdalla

* Corresponding author.

E-mail address: babak.mohammadi@nateko.lu.se (B. Mohammadi).

<https://doi.org/10.1016/j.egy.2023.10.070>

Received 22 June 2023; Received in revised form 16 October 2023; Accepted 23 October 2023

Available online 7 November 2023

2352-4847/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature*Abbreviations*

ACO	Ant Colony Optimization
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
AOA	Archimedes Optimization Algorithm
BA	Bat Algorithm
BAT	Bat Optimization Algorithms
BMA	Bayesian Averaging Model
BO	Bayesian Optimization
CMAES	Covariance Matrix Adaptive Evolution Strategies
DA	Dragonfly Algorithm
DL	Deep Learning
DT	Decision Tree
DWT	Discrete Wavelet Transforms
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
GA	Genetic Algorithm
GAM	Generalized Additive Model
GMDH	Group Method of Data Handling
GOA	Grasshopper Optimization Algorithm
GPR	Gaussian Process Regression
GSR	Global Solar Radiation
GWO	Gray Wolf Optimizer
IMFs	Intrinsic Mode Functions
KNEA	Kernel-based Nonlinear Extension of the Arps decline
KNN	K-Nearest Neighbors
LM	Levenberg-Marquardt
LSSVM	Least Square Support Vector Machine
LSTM	Long Short-Term Memory network
M5T	M5 Tree
MARS	Multivariate Adaptive Regression Spline
MFRFNN	Multi-Functional Recurrent Fuzzy Neural Network
ML	Machine Learning
MLFFNN	Multilayer Feedforward Neural Network
MLP	Multilayer Perceptron
MLQR	Multiple Linear Quantile Regression
MPMR	Minimax Probability Machine Regression

MVO	Multi Verse Algorithm
PSO	Particle Swarm Optimization
QKNN	Quantile K-Nearest Neighbors
QR	Quantile Regression
QRF	Quantile Regression Forests
QRFs	Quantile Regression Forests
QRNN	Quantile Regression Neural Network
RF	Random Forest
RP	Resilient Back Propagation
SCG	Scaled Conjugate Gradient
SMHI	Swedish Meteorological and Hydrological Institute
SSA	Salp Swarm Algorithm
SVM	Support Vector Machine
VMD	Variational Mode Decomposition
XGB	Extreme Gradient Boosting

Parameters

ΔT	Temperature variations
BC	Bristow-Campbell model
CB	Cloud-based models
CC	Correlation coefficient
CV	Coefficient of variation
DYB	Day-of-the-year-based
MAE	Mean absolute error
N	Maximum possible sunshine duration
NSE	Nash–Sutcliffe efficiency
R^2	Coefficient of determination
RH	Relative humidity
SD	Standard deviation
SDB	Sunshine duration-based models
SSH	Sunshine duration
T	Mean air temperature
TB	Temperature-based models
$T_{\max}$	Maximum air temperature
$T_{\min}$	Minimum air temperature
WS	Wind speed
$X_{\max}$	Maximum value
X_{mean}	Average value
$X_{\min}$	Minimum value

et al., 2020; Mohammadi and Moazenzadeh, 2021). This makes measuring GSR more difficult than measuring other meteorological variables such as temperature, relative and specific humidity, number of sunshine hours, wind speed and direction, and precipitation amount (Yousuf and Umair, 2018). Therefore, a wide range of techniques, such as empirical methods, machine learning (ML) models, and satellite-based models, have been suggested for estimating GSR depending on other freely available meteorological data (Janjai et al., 2011; Ramirez Camargo and Dörner, 2016; Cornejo-Bueno et al., 2019; Ghimire et al., 2019). Amongst models, empirical approaches and ML models have been widely developed and implemented due to their extremely low cost and excellent accuracy in time series prediction, and they are still attracting attention from scholars.

Over the last few decades, researchers from many countries have developed numerous empirical models based on the correlation between GSR and other meteorological variables, namely sunshine duration-based (SDB) models (Prescott, 1940), temperature-based (TB) models (Hargreaves and Samani, 1982), cloud-based (CB) models (Black, 1956), and other models based on meteorological variables (Yorukoglu and Celik, 2006; Zhang et al., 2017). As a result, the most frequently employed empirical models are those based on SDB and TB, while SDB models were significantly more accurate than TB models. However, the sunshine duration data as input of SDB models are not easily obtainable

in many parts of the world as easily as temperature data. Thus, many TB models also received widespread recognition. Hargreaves developed an approach based on simple inputs using the relationship between GSR and minimum and maximum temperature (Hargreaves and Samani, 1982). The Bristow-Campbell (BC) model, which might account for 70–90% of GSR variability in America, was created by Bristow and Campbell (1984), after they evaluated an exponential relationship between GSR and temperature variations (ΔT). In another work, Khorasanizadeh et al. (2014) investigated the usefulness of the day-of-the-year-based (DYB) models to estimate GSR and compared them with the SDB and TB models. The results showed that the predictions from DYB models were greater compared to TB models, with just a slight improvement in the accurate prediction of the SDB model.

Several studies have concentrated on developing ways to estimate the GSR in areas where observed data are unavailable. The most popular estimation approach is using correlations between GSR and other available observed meteorological variables to determine their statistical relationships (Sun et al., 2015). This approach is easy to use since it is straightforward and efficient while also not necessitating the collection of any extra information. The common meteorological variables utilized to produce such statistical relationships with GSR are: SSH, air temperature, RH, cloudiness, and precipitation (P) (Chen et al., 2006; Ouali and Alkama, 2014; Bayrakçı et al., 2018; Fan et al., 2019a; Li et al.,

2020; Al-Ghezi et al., 2022). A complex model developed by Feng et al. (2020) based on RH, P, and air temperature was compared with 19 calibrated local empirical models using 96 ground stations across different climatic zones of China. In another relevant study, Mohammadi and Moazenzadeh (2021) proposed several empirical approaches by creating a relationship between P and RH using linear analysis regression across 13 Peruvian weather stations.

Empirical models are often used to calculate daily GSR, but they are less successful than ML models which the ML models are better able to address complex and nonlinear interactions among the variables (Fan et al., 2018). Recently, many ML models have been employed to predict GSR, such as RF, support vector machine (SVM), decision tree (DT), XGB, M5T, extreme learning machine (ELM), generalized additive model (GAM), and k-nearest neighbors (KNN) (Feng et al., 2017b; Huang et al., 2021; Mohammadi et al., 2022; Qiu et al., 2022; Solano et al., 2022; Zhao et al., 2022; Ledmaoui et al., 2023). Among these models, the artificial neural network (ANN) model has been widely used, while the SVM model exhibits higher accuracy and lower computational cost and speed than the ANN model (Meenal and Selvakumar, 2018). However, when calculating GSR in humid subtropical regions and various seasons, the XGB model outperforms the SVM model with equivalent accuracy, greater stability, and significantly lower computational costs (Fan et al., 2018; Solano et al., 2022).

Many researchers developed approaches by combining ordinary ML models with evolutionary algorithms to improve their accuracy for the prediction of GSR in various regions. For instance, a hybrid of the SVM model with Bat optimization algorithms (SVM-BAT) was proposed to estimate daily diffuse solar radiation across air-polluted areas, and it outperformed the standalone SVM, XGB, and multi-adaptive regression splines (MARS) (Fan et al., 2020). Ant colony optimization (ACO) was used by Prasad et al. (2019) to optimize RF and minimax probability machine regression (MPMR) models, which led to the discovery that the new optimal models resulted acceptable estimation accuracy compared ordinary models. Also, least square support vector machine (LSSVM) was hybrid with multi verse algorithm original version (MVO) to predict solar radiation across the Southeast region of China (Ikram et al., 2022). The hybrid long short-term memory network (LSTM) with Bayesian Optimization (BO) techniques improved the accuracy of prediction solar radiation with 8.23% (Zuo et al., 2022). However, the group method of data handling (GMDH) and multilayer feedforward neural network (MLFFNN) models showed the highest accuracy for predicting daily GSR across 12 stations in different climate zones in Iran compared to the standalone adaptive neuro-fuzzy inference system (ANFIS), and hybrid ANFIS with particle swarm optimization algorithm (ANFIS-PSO), genetic algorithm (ANFIS-GA), and ant colony (ANFIS-ACO) models (Khosravi et al., 2018). Tao et al. (2021) employed novel ANFIS-based optimization algorithms, including Salp Swarm Algorithm (SSA) (ANFIS-SSA) and Grasshopper Optimization Algorithm (GOA) (ANFIS-GOA), PSO (ANFIS-PSO), Gray Wolf Optimizer (GWO) (ANFIS-GWO), and Dragonfly Algorithm (DA) (ANFIS-DA), to predict GSR at various locations in North Dakota, USA. The maximum, mean, and minimum air temperature data were utilized in the model setup. The ANFIS-muSG hybrid algorithm yielded the most optimal results due to the analyses. Zhang et al. (2021a) used PSO, GA, Bayesian averaging model (BMA), Archimedes optimization algorithm (AOA), and BAT based on the ANFIS and MLP to predict monthly solar radiation. These models consider the monthly average temperature, wind speed, relative humidity, and sunshine duration as inputs. The analysis resulted in the BMA algorithm providing superior results compared to the others. The hourly average GSR at six meteorological stations situated in Tamil Nadu, India was estimated by Geetha et al. (2022) using the ANN-based Levenberg-Marquardt (LM), Resilient Back Propagation (RP), and Scaled Conjugate Gradient (SCG) algorithm. The ANN model, developed using the LM algorithm, produced outputs with minimal error and maximum efficiency, as evidenced by the comparison of statistical metrics. Transfer learning was implemented by Manandhar et al. (2023),

utilizing Total Sky-Imager (TSI) images of the Southern Great Plains (SGP). TSI images were extracted and solar radiation was predicted by applying deep neural network architectures. The analysis has demonstrated that the transfer learning technique is a highly effective way to reduce the time and resources required for modeling solar radiation. To predict solar radiation values in Indian cities, including Delhi, Chennai, Hyderabad, Nagpur, Patna, Trivandrum, and Bhubaneswar, Sivakumar et al. (2023) implemented a combination of VMD and discrete wavelet transform (DWT) with five deep learning (DL) and two ML models. The integration of DL-based VMD has yielded significantly promising outcomes compared to DWT integration. Recently, Goliatt and Yaseen (2023) developed a framework to improve the prediction of GSR according to the hybridization of the covariance matrix adaptive evolution strategies (CMAES) model with XGB and MARS models at four ground radiation stations in Burkina Faso. They resulted that the new framework reduced the prediction error (i.e., root mean square error-RMSE) during the testing phase by 4.9%, 17.2%, 39.8%, and 44.5%. Relative speaking, some nonparametric approaches developed and demonstrated their stability and potential to produce probabilistic forecasts of solar radiation and production, such as quantile regression (QR), quantile K-nearest neighbors (QKNN), multiple linear quantile regression (MLQR), quantile regression forests (QRFs), and quantile regression neural network (QRNN) (Bakker et al., 2019; Bracale et al., 2019; Bozorg et al., 2020; Liu et al., 2020).

VMD is considered as most recent signal decomposition technique, which decomposes the input signal into distinct band-limited intrinsic mode functions (IMFs). VMD offers advantages over empirical mode decomposition (EMD), wavelet transformation, and Hilbert-Huang transform, including the absence of modal aliasing and not being sensitive to noise (Chaitanya et al., 2021). The voltage signal is subjected to VMD, and an index is constructed to distinguish islanding from non-islanding circumstances (Salimi and Koochaki, 2019). Due to the ability of VMD analysis in solar radiation time series decomposition (Majumder et al., 2018; Ali et al., 2021; Korkmaz, 2021) and the importance of GSR prediction, therefore the current study seeks to evaluate and improve solar radiation time series prediction via the following aims: (1) proposing a novel combined VMD approach with MFRFNN (VMD-MFRFNN) to decompose and estimate global solar radiation time series data for the first time. VMD-MFRFNN is the first decomposition-based recurrent fuzzy neural network that identifies the system's state, learns a single function for each state, and models multiple functions at the same time. (2) Comparing proposed approach (VMD-MFRFNN) with the prediction of solar radiation time series by MFRFNN model, (3) the first application of QRF model for the solar radiation prediction, (4) evaluating and comparing the ability of classical ML models over VMD-MFRFNN, MFRFNN, and QRF models for daily global solar radiation in Sweden.

2. Materials and methods

2.1. Study sites

The objective of this study was to estimate GSR time series using newly developed ML techniques in two meteorological stations in southern Sweden. Based on the map (Fig. 1), both studied stations (Lund and Växjö) are prone to receive suitable solar energy that can be used as a clean energy source in these areas. Lund meteorological station is located at 55°71' N latitude, and 13°21' E longitude by an altitude of 90 m, and Växjö meteorological station is located at 56°92' N latitude and 14°73' E longitude by an altitude of 182 m.

2.2. Data used

Daily meteorological variables such as average daily GSR, air temperature (Tmin, Tmax, T), wind speed (WS) at 2 m above the ground, RH, and SSH were obtained from the Swedish Meteorological and



Fig. 1. Location of the study sites.

Hydrological Institute (SMHI) during 2008–2021 from Vaxjö and Lund meteorological stations. The maximum possible sunshine duration (N) was estimated using the solar constant, location latitude, solar declination, sunset hour angle, and year date (Allen et al., 1998). Table 1 lists the statistical characteristics of meteorological datasets, including maximum (Xmax), minimum (Xmin), average (Xmean), coefficient of variation (CV), and standard deviation (SD) of each measured data. The correlation matrix between measured meteorological variables is shown in Fig. 2.

Table 1
Statistical indices for meteorological variables at the Vaxjö and Lund stations during the whole period (2008–2021).

Stations	Parameter	Xmin	Xmean	Xmax	CV	SD
Vaxjö	Tmin (°C)	-16.2	7.5	25.6	1.0	7.4
	T (°C)	-22.0	4.0	19.8	1.7	6.6
	Tmax (°C)	-11.8	11.4	33.6	0.8	8.7
	WS (m/s)	0.1	3.1	9.8	0.4	1.2
	SSH (h)	0.0	4.5	16.5	1.0	4.6
	RH (%)	37.0	82.2	100.0	0.2	12.9
	N (h)	6.5	12.0	17.5	0.3	3.8
	GSR (W/m ²)	1.0	113.8	359.5	0.9	97.9
	Tmin (°C)	-19.3	6.0	21.8	1.1	6.5
	T (°C)	-11.8	9.4	27.0	0.8	7.1
Lund	Tmax (°C)	-8.9	13.2	34.4	0.6	8.3
	WS (m/s)	0.2	3.0	11.1	0.5	1.5
	SSH (h)	0.0	5.0	16.8	1.0	4.8
	RH (%)	42.3	81.0	100.0	0.1	10.6
	N (h)	6.8	12.0	17.3	0.3	3.6
	GSR (W/m ²)	1.8	118.7	361.7	0.8	100.0

2.2.1. Random forest (RF)

RF is a nonparametric supervised ML approach (Fatahi et al., 2023) and was first introduced by Breiman (2001). To build individual decision trees, samples for learning are chosen randomly. By using some regression techniques, these trees can make their predictions, and the mean value of those predictions serves as the outcome. Regression techniques can lessen the over-fitting issue, and the model can handle huge dimensions regression and difficult nonlinear issues. Based on Hosseinzadeh et al. (2012), the RF model only requires two user-defined parameters: the number of variables that can be used at each node and the number of trees. Several studies applied the RF model to predict the average daily GSR across different climate zones (Feng et al., 2019; Prasad et al., 2019).

2.2.2. Extreme gradient boosting (XGB)

XGB is a scalable tree-boosting method proposed by Chen and Guestrin (2016). XGB has gained popularity in ML due to its strong predictive power and computational efficiency. It uses the idea of boosting to create an ensemble of weak learners, each of which is trained on subsets of training samples, and a weighted combination of these individual models is used to make predictions (Pan and Iop, 2017; Zhang et al., 2021b). XGB smooths the learned weights and reduces the probability of overfitting using the regularization technique (Song et al., 2020). It employs a second-order Taylor expansion for the approximation of the loss function. The advantages of modeling by XGB include reducing the model's complexity, preventing overfitting, the capability to handle large datasets, being parallelizable, and being computationally inexpensive (Wang et al., 2018).

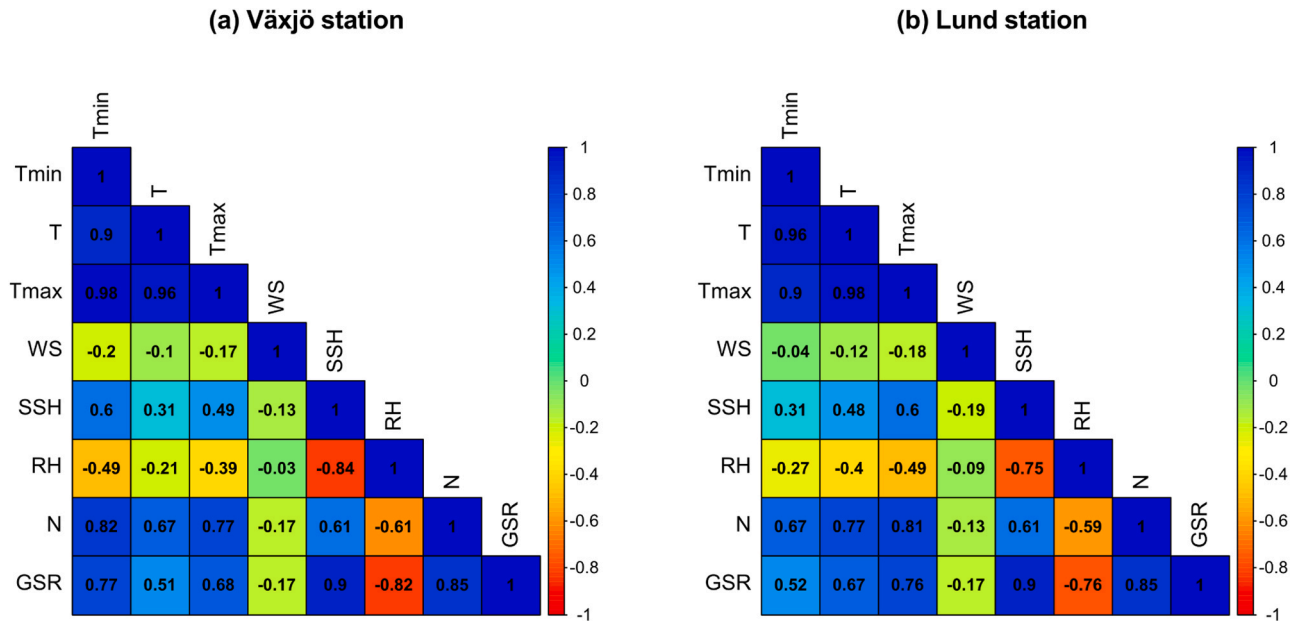


Fig. 2. The correlation coefficients between measured meteorological variables and GSR in both case studies.

2.2.3. M5 tree (M5T)

Quinlan (1992) created the original M5T model proposal, which Wang and Witten (1997) later adjusted and reconstructed. The M5Tree technique uses linear regression formulas at the current node to build a link between the inputs and output. The M5T model chooses the optimal split for the estimated error reduction after scanning all feasible splits. The M5T model usually includes two steps: (i) the raw data is divided into smaller groups to build a decision tree that considers the standard deviation based on the decision scores and the estimated error reduction. (ii) the outgrowth tree is decreased, and the linear regression formula controls these sub-trees.

2.3. Quantile regression forest (QRF)

The QRF approach was developed by Meinshausen and Ridgeway (2006). This model was built based on the RF technique using conditional quantiles for prediction instead of the conditional mean (Breiman, 2001; Din et al., 2018). The CART decision tree in RF is used as a weak trainer; nevertheless, for every tree produced, the necessary number of attributes will be randomly chosen from the raw dataset variables. Consequently, in a regression issue, the outcome of weak learners is combined to produce the final prediction model (Breiman, 1996). It is crucial to use the RF averaging strategy to minimize bias, variation, and correlation among trees (Hameed et al., 2021).

The QRF model uses QR rather than the average strategy to determine the final output, and it is regarded as an improved variation of RF (Gyamerah et al., 2020). Additionally, the QRF model is viewed as a nonparametric technique strengthened by a strong theoretical underpinning. The main principle of the QRF model is that the weighting of observed data used to estimate the conditional mean might also provide a reasonable approximation of the whole conditional distribution (Meinshausen and Ridgeway, 2006). Assume that Y is a real-valued observed or measured variable and X is a predictor or estimated variable with a potential high dimensional. A probability assumption that Y is relatively smaller than when $X = x$ (single tree), the conditional distributed function $F(y|X = x)$ is used in QRF definition as follows:

$$F(y|X = x) = P(Y \leq y|X = x) \quad (1)$$

For a continuously distributed function, the α -quantile $Q_\alpha(x)$ is therefore defined so that the possibility of Y being less than $Q_\alpha(x)$ is for a

given $X = x$, absolutely equivalent to α .

$$Q_\alpha(x) = \inf\{y : F(y|X = x) \geq \alpha\} \quad (2)$$

By taking into account a specific value of QR then, it could be utilized to construct prediction levels. For instance, consider a 95% prediction level for the value of:

$$I(x) = [Q_{0.025}(x), Q_{0.975}(x)] \quad (3)$$

That is an innovative observation of Y for $X = x$ occurs with the highest probability ($\alpha = 0.95$), in the interval $I(x)$.

The QRF model calculates the conditional distribution function of Y considering $X = x$ according to Eq. 1.

$$\hat{F}(y|X = x) = \sum_{i=1}^n w_i(x) I_{\{y \leq y_i\}} \quad (4)$$

Predictions of the conditioned quantiles $Q_\alpha(x)$ are produced from $\hat{F}(y|X = x)$, as shown in (Eq. (3)).

The QRF model can be used for the in-sampling prediction (using the whole data period) and was then considered for simulating GSR based on the whole data of the interdependence structures between variables. To reduce unnecessary computational costs, we limited our testing for simulation by using only five quantile levels (5th, 25th, 50th, 75th, and 95th). Eq. (2) was used to construct conditional quantiles of GSR with $\alpha = [0.05, 0.25, 0.50, 0.75, 0.95]$. The RMSE measure was employed as a main criterion to evaluate the accuracy of predicted GSR and analyze the uncertainty of the QRF model at various quantile levels. The framework used for implementing the QRF model is shown in Fig. 3.

2.4. Combining variational mode decomposition with multi-functional recurrent fuzzy neural network (VMD-MFRFNN)

MFRFNN was developed by Nasiri and Ebadzadeh (2022) for time series prediction in 2022. Two fuzzy neural networks make up MFRFNN: the output and state networks. The output network produces the model's output, while the state network determines the system's state (Fig. 4). MFRFNN employs the least square method to train the output network and particle swarm optimization algorithm to learn the state network's weights. PSO can handle optimization problems with several local optima reasonably well and has global search capabilities.

Having a structure with a feedback loop allows the MFRFNN to

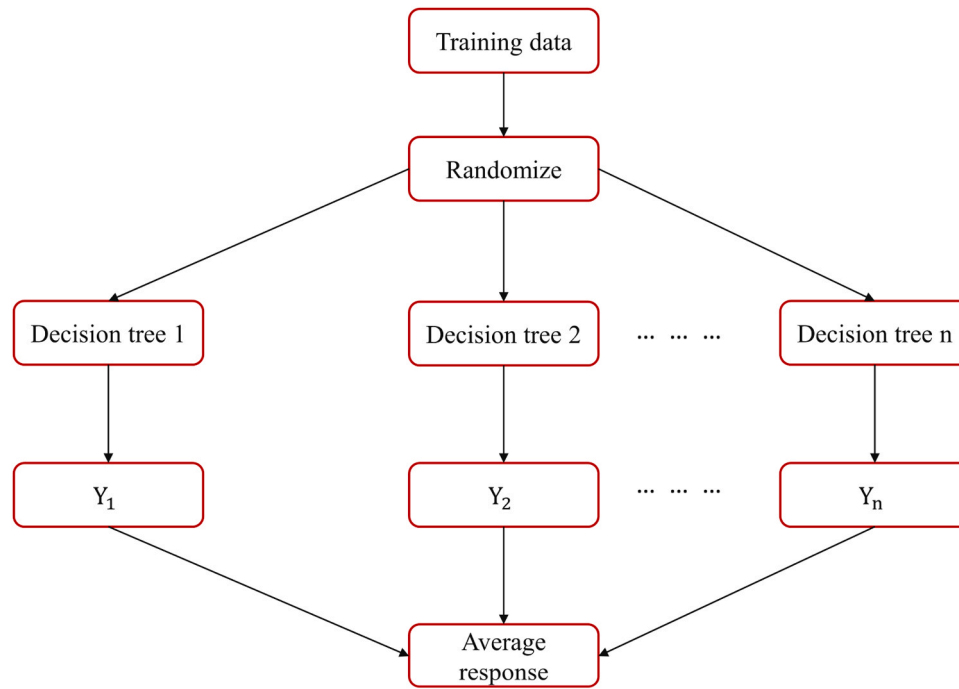


Fig. 3. The framework used for implementation of the QRF model in the current study.

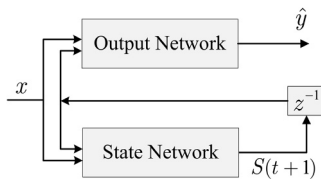


Fig. 4. Overview of MFRFNN structure (Nasiri and Ebadzadeh, 2022).

memorize historical information and capture the dynamic behavior of the time series. By integrating the states into its architecture, MFRFNN can approximate several functions simultaneously (Nasiri and Ebadzadeh, 2022). The VMD is a non-recursive adaptive signal decomposition technique based on the constrained variational model (Yang et al., 2021; Nasiri and Ebadzadeh, 2023). The VMD can decompose the original time series data into a finite set of IMFs and a residual signal (He et al., 2022). The IMFs ($u_k(t)$) and their respective center frequencies (ω_k) are computed by solving the following variational model:

$$\begin{aligned} & \min_{\{u_k\}_{\{\omega_k\}}} \sum_{k=1}^K \left\| \partial_t \left[\left(\partial(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & \text{subject to } \sum_{k=1}^K u_k = f \end{aligned} \quad (5)$$

where f represents the input time series, t is the time step, K denotes the number of IMFs, and $\delta(\bullet)$ is the delta function.

Recently, Nasiri and Ebadzadeh have proposed VMD-MFRFNN (Nasiri and Ebadzadeh, 2023), which combines VMD and MFRFNN. VMD-MFRFNN comprises two phases. In the first phase, i.e., the decomposition phase, VMD decomposes the input time series into several IMFs. In the second phase, i.e., the prediction phase, each IMF is given as an input to a distinct MFRFNN model. Each MFRFNN generates output and predicts the future values of its input signal. Finally, the predicted time series are aggregated to produce the final output (Nasiri and Ebadzadeh, 2023). By decomposing the time series into components of simpler structures, VMD-MFRFNN enables MFRFNN to more accurately predict stationary trends and increase the prediction accuracy.

2.5. Models' construction

In this study, different ML-based models were employed to predict average daily GSR, including RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models. All models were trained using their respective learning techniques and criteria. The average daily GSR observation was used as the target attribute in all models, and other meteorological variables were considered as input for the RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models. The optimal parameter setting for all applied models is listed in Table 2. Also, Fig. 5 shows the workflow of the current study, including connections between input, models, and output.

2.6. Scenario definition for global solar radiation (GSR) prediction

Table 3 shows proposed scenarios for various ML models based on seven input combinations to predict average daily GSR throughout the Lund and Vaxjo meteorological stations. The capabilities of MFRFNN,

Table 2

Hyperparameters and functions of the models used to predict the average daily GSR across the Vaxjo and Lund stations over the entire period (2008–2021).

Models	Parameters
RF	Ntree: 500 Mtry: 2
XGB	Nrounds: 500 Max_depth: 2 Learning rate: 0.3 Subsample: 1 Min_child_weight: 1
M5T	Committees: 100 Neighbors: 0
QRF	Num.trees: 2000 Quantiles: 0.5 Sample.fraction: 0.5 Min.node.size: 5
MFRFNN and VMD-MFRFNN	Number of output network's fuzzy sets: 2 Number of state network's fuzzy sets: 2 Number of states: 2 Number of IMFs: 9

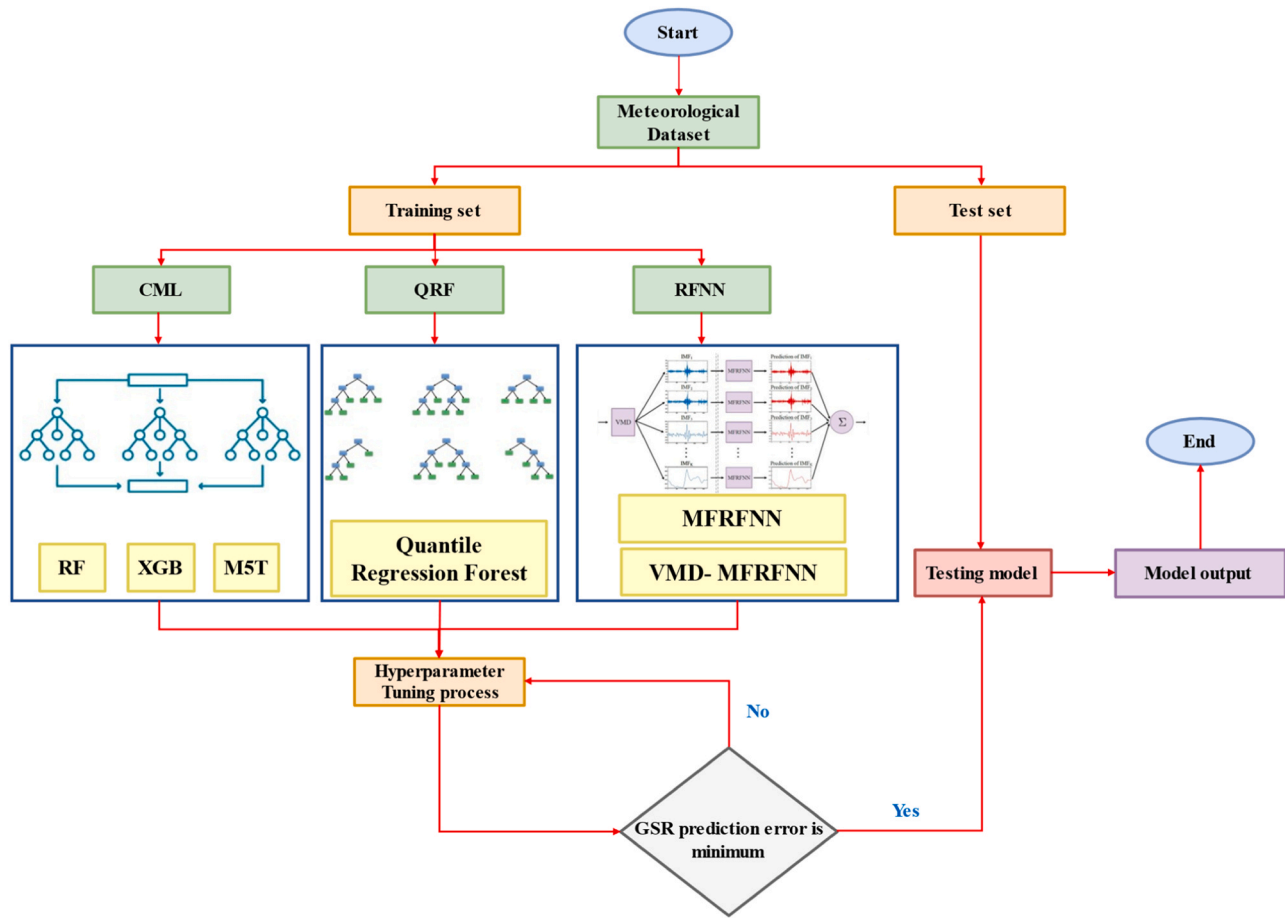


Fig. 5. A schematic flowchart of the current study.

Table 3

Input combinations of applied models to predict GSR using several meteorological variables.

Combination number	Models					Input variables		Output
C1	RF1	XGB1	M5T1	QRF1	MFRFNN1	VMD-MFRFNN1	Tmin, Tmax, T	GSR
C2	RF2	XGB2	M5T2	QRF2	MFRFNN2	VMD-MFRFNN2	Tmin, Tmax, T, WS	GSR
C3	RF3	XGB3	M5T3	QRF3	MFRFNN3	VMD-MFRFNN3	Tmin, Tmax, T, WS, SSH	GSR
C4	RF4	XGB4	M5T4	QRF4	MFRFNN4	VMD-MFRFNN4	Tmin, Tmax, T, WS, SSH, RH	GSR
C5	RF5	XGB5	M5T5	QRF5	MFRFNN5	VMD-MFRFNN5	Tmin, Tmax, T, WS, SSH, RH, N	GSR

VMD-MFRFNN, and QRF techniques were compared with ordinary ML-based tree (RF, XGB, and M5T) models according to the different combinations to assess the impact of specific meteorological variables on GSR prediction. In particular, six scenarios were defined based on meteorological variables (Tmin, Tmax, T, WS, SSH, RH, and N) were utilized as input data for the five proposed models to investigate the best-driving variables to predict the average daily GSR in the south of Sweden. The input data from 2008 to 2017 (75%) were employed to train models, and the data from 2018 to 2021 (25%) were used as a test set for the model prediction results.

2.7. Statistical metrics

This study employed the Coefficient of determination (R^2), the Nash–Sutcliffe efficiency (NSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the RMSE to evaluate the performance of six applied models based on the defined scenarios in Table 3.

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}, \quad (6)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (7)$$

$$MAE = \frac{\sum_{i=1}^n |(x_i - y_i)|}{n}, \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}, \quad (9)$$

$$MAPE = 100 \times \frac{1}{n} \times \sum_{i=1}^n \frac{|y_i - x_i|}{|y_i|}, \quad (10)$$

where x_i represents the predicted GSR with applied models (RF, XGB, M5T, QRF, MFRFNN, and VMD- MFRFNN) while y_i represents the ground truth GSR. The $\bar{x}$ and $\bar{y}$ donate the average value of predicted and observed GSR, respectively; the number of days in the dataset is denoted by n , and the subscript i denotes the number of samples. The R^2 indicates the correlation and agreement between predicted and observed GSR, the positive value close to 1 reflects the highest agreement and good correlation. The NSE reflects the relative magnitude of RMSE relative to observed GSR, the highest NSE is 1, which indicates the perfect prediction. The MAE, MAPE and RMSE metrics range from 0 to ∞ , and 0 is their optimal value.

3. Results

3.1. Performance of daily GSR prediction

This study compared the accuracy and capability of RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models of average daily GSR prediction across the Lund and Växjö stations. Table 4 presents the summary of statistical metrics for comparing the predicted GSR by classical ML (CML) (RF, XGB, and M5T), QR (QRF), and RFNN (MFRFNN and VMD-MFRFNN) models against observed GSR at the Växjö station. In terms of CML models, the RF model exhibits the highest accuracy during the training phase across all scenarios, followed by XGB and M5T models, with R^2 in 0.939–0.997, RMSE in 5.8–24.6 W/m², MAE in

4.1–17.9 W/m², MAPE in 8.2–41.3%, and NSE in 0.936–0.996. However, the M5T model showed better performance during the testing phase than RF and XGB models, with R^2 in 0.777–0.988, RMSE in 11.4–47.7 W/m², MAE in 7.81–34.0 W/m², MAPE in 15.1–69.3%, and NSE in 0.775–0.987. The QRF model is higher than the M5T model but lower than RF and XGB models during the training phase across all scenarios. While the QRF model showed equivalent performance with the M5T model and outperformed RF and XGB models during the testing phase across all scenarios. In addition, applying the RFNN models through the MFRFNN model for the first time to predict the average daily GSR, the accuracy during the training phase is less than in RF, XGB, M5T, and QRF models. However, the MFRFNN model outperformed the RF, XGB, M5T, and QRF models during the testing phase, and the R^2 in 0.789–0.988, RMSE in 11.2–46.7 W/m², MAE in 7.9–33.7 W/m², MAPE in 16.9–78.0%, and NSE in 0.784–0.988. When the GSR estimation potential of the VMD-MFRFNN model was evaluated according to other algorithms during the training phase, the VMD-MFRFNN hybrid model outperformed the M5T, QRF, and MFRFNN models in all input scenarios, except for the C5 combination, but showed poor performance than the RF model, with R^2 in 0.820–0.947, RMSE in 23.3–42.7 W/m², MAE in 17.4–31.2 W/m², MAPE in 33.9–67.8%, and NSE in 0.819–0.946. When the test performance of the VMD-MFRFNN hybrid model is evaluated, it is seen that it performs better than the other models, except for the C5 combination. In addition, it is noted that the VMD-MFRFNN model performs weaker than the other models in the C5 combination in the test phase.

Table 5 summarizes statistical metrics for comparing the predicted GSR by CML, QR, and RFNN models against observed GSR at the Lund station. Between CML models, the RF model exhibits the highest

Table 4

Statistical performance for RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models based on a different scenario for estimating average daily GSR compared to observed average daily GSR during the training and testing phase at the Växjö Station.

Models	Training (2008–2017)					Testing (2018–2021)				
	R^2	RMSE (W/m ²)	MAE (W/m ²)	MAPE (%)	NSE	R^2	RMSE (W/m ²)	MAE (W/m ²)	MAPE (%)	NSE
Tmin, Tmax, T										
RF1	0.939	24.6	17.9	41.3	0.936	0.768	48.8	35.2	82.8	0.764
XGB1	0.851	37.4	27.1	61.1	0.851	0.767	49.0	35.0	78.2	0.762
M5T1	0.797	43.7	31.6	64.2	0.796	0.777	47.7	34.0	69.3	0.775
QRF1	0.828	40.3	27.3	50.2	0.826	0.772	48.5	34.3	65.5	0.767
MFRFNN1	0.803	43.0	31.7	70.1	0.803	0.789	46.7	33.7	78.0	0.784
VMD-MFRFNN1	0.848	37.7	28.4	65.5	0.848	0.820	42.7	31.2	67.3	0.819
Tmin, Tmax, T, WS										
RF2	0.952	22.2	16.4	41.4	0.948	0.777	47.6	35.0	92.9	0.775
XGB2	0.871	34.9	25.5	60.6	0.870	0.779	47.8	34.6	82.8	0.773
M5T2	0.811	42.1	30.6	65.0	0.811	0.789	46.6	33.4	70.2	0.784
QRF2	0.839	39.1	26.5	49.2	0.837	0.779	47.7	33.8	66.4	0.775
MFRFNN2	0.818	41.3	30.6	69.1	0.818	0.805	44.9	32.6	73.9	0.800
VMD-MFRFNN2	0.855	36.9	27.7	64.2	0.855	0.845	39.8	29.6	67.8	0.843
Tmin, Tmax, T, WS, SSH										
RF3	0.980	14.2	10.8	27.0	0.978	0.906	31.1	23.5	58.9	0.904
XGB3	0.951	21.3	16.0	36.6	0.951	0.909	30.8	22.7	51.4	0.906
M5T3	0.928	26.0	19.1	34.8	0.928	0.919	28.8	20.9	41.7	0.918
QRF3	0.936	24.7	17.0	29.0	0.935	0.912	30.1	21.7	40.9	0.910
MFRFNN3	0.932	25.3	18.8	38.1	0.932	0.925	27.9	20.6	43.2	0.923
VMD-MFRFNN3	0.961	19.1	14.5	32.8	0.961	0.947	23.3	17.4	38.4	0.946
Tmin, Tmax, T, WS, SSH, RH										
RF4	0.988	10.9	8.0	17.7	0.987	0.937	26.1	19.3	45.3	0.933
XGB4	0.963	18.6	13.7	31.5	0.963	0.930	27.4	20.2	43.9	0.925
M5T4	0.944	23.0	16.8	31.2	0.943	0.940	25.5	18.6	37.2	0.936
QRF4	0.951	21.5	14.7	26.4	0.951	0.936	26.1	18.9	37.9	0.932
MFRFNN4	0.946	22.4	16.6	35.0	0.946	0.941	25.3	18.2	41.3	0.937
VMD-MFRFNN4	0.965	18.2	13.8	30.9	0.965	0.941	24.6	18.0	33.9	0.940
Tmin, Tmax, T, WS, SSH, RH, N										
RF5	0.997	5.8	4.1	8.2	0.996	0.983	13.4	9.4	19.8	0.982
XGB5	0.992	8.9	6.4	14.0	0.992	0.985	12.6	9.0	18.4	0.984
M5T5	0.988	10.8	7.5	14.7	0.988	0.988	11.4	7.8	15.1	0.987
QRF5	0.989	10.0	6.4	12.6	0.989	0.986	11.8	8.2	16.7	0.986
MFRFNN5	0.987	10.9	7.8	16.9	0.987	0.988	11.2	7.9	16.9	0.988
VMD-MFRFNN5	0.973	16.0	12.0	27.8	0.973	0.901	33.1	20.9	41.7	0.892

Note: The RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models that performed the best statistically are bolded during the training and testing phase

Table 5

Statistical performance for RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models based on a different scenario for estimating average daily GSR compared to observed average daily GSR during the training and testing phase at the Lund station.

Models	Training (2008–2017)					Testing (2018–2021)				
	R ²	RMSE (W/m ²)	MAE (W/m ²)	MAPE (%)	NSE	R ²	RMSE (W/m ²)	MAE (W/m ²)	MAPE (%)	NSE
Tmin, Tmax, T										
RF1	0.931	27.0	19.6	42.2	0.926	0.740	52.1	38.5	82.7	0.735
XGB1	0.824	41.7	30.1	62.2	0.824	0.743	51.7	38.1	75.7	0.739
M5T1	0.764	48.5	35.0	64.9	0.762	0.740	51.8	38.2	70.7	0.738
QRF1	0.803	44.4	30.0	52.1	0.801	0.752	50.9	36.8	63.9	0.748
MFRFNN1	0.782	46.5	34.0	69.4	0.782	0.768	49.2	36.3	75.7	0.764
VMD-MFRFNN1	0.795	45.1	33.3	67.2	0.795	0.815	43.8	31.8	62.9	0.813
Tmin, Tmax, T, WS										
RF2	0.944	24.8	18.4	45.6	0.938	0.754	50.3	37.5	91.3	0.753
XGB2	0.844	39.4	28.3	63.4	0.843	0.756	50.2	36.5	76.8	0.755
M5T2	0.787	46.0	33.1	66.2	0.786	0.766	49.1	35.6	65.9	0.765
QRF2	0.820	42.4	28.6	51.4	0.818	0.761	49.8	35.9	63.2	0.758
MFRFNN2	0.797	44.8	32.8	69.8	0.797	0.785	47.0	34.4	67.8	0.785
VMD-MFRFNN2	0.801	44.4	32.8	66.7	0.801	0.811	44.1	32.0	54.2	0.811
Tmin, Tmax, T, WS, SSH										
RF3	0.978	15.4	11.8	28.7	0.976	0.905	31.5	23.9	57.5	0.904
XGB3	0.941	24.1	18.0	37.6	0.941	0.908	31.0	23.2	47.2	0.906
M5T3	0.924	27.5	20.2	33.9	0.923	0.920	28.8	21.3	38.1	0.919
QRF3	0.932	26.1	18.3	29.2	0.931	0.912	30.3	22.3	39.3	0.910
MFRFNN3	0.925	27.2	20.3	39.1	0.925	0.923	28.3	21.0	43.2	0.922
VMD-MFRFNN3	0.881	34.6	24.7	49.7	0.879	0.893	33.3	23.7	46.1	0.892
Tmin, Tmax, T, WS, SSH, RH										
RF4	0.986	12.0	9.1	19.5	0.985	0.923	29.6	22.0	52.6	0.915
XGB4	0.955	21.2	16.0	33.8	0.955	0.916	30.8	22.9	50.4	0.907
M5T4	0.938	24.9	18.6	32.5	0.937	0.930	28.5	20.8	41.0	0.921
QRF4	0.945	23.4	16.5	27.5	0.945	0.923	29.0	21.0	39.1	0.918
MFRFNN4	0.935	25.4	19.0	38.3	0.935	0.933	28.5	20.9	42.9	0.921
VMD-MFRFNN4	0.876	35.3	25.0	49.0	0.874	0.881	35.6	24.9	48.8	0.876
Tmin, Tmax, T, WS, SSH, RH, N										
RF5	0.996	6.3	4.5	9.2	0.996	0.982	14.3	10.2	22.7	0.980
XGB5	0.990	9.9	7.1	14.6	0.990	0.983	13.5	9.5	19.3	0.982
M5T5	0.987	11.7	8.0	15.6	0.986	0.987	12.0	8.3	16.2	0.986
QRF5	0.988	10.8	6.9	13.4	0.988	0.986	12.2	8.5	16.7	0.986
MFRFNN5	0.987	11.5	8.1	17.3	0.987	0.987	12.0	8.3	18.4	0.986
VMD-MFRFNN5	0.890	33.2	23.5	45.8	0.889	0.817	43.4	30.2	59.6	0.816

Note: The RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models that performed the best statistically are bolded during the training and testing phase

accuracy during the training phase across all scenarios, followed by XGB and M5T models, with R² in 0.931–0.996, RMSE in 6.3 – 27.0 w/m², MAE in 4.5 – 19.6 w/m², MAPE in 9.2 – 42.2%, and NSE in 0.926–0.996. However, the M5T model showed better performance during the testing phase than RF and XGB models, with R² in 0.740–0.987, RMSE in 12.0 – 52.1 W/m², MAE in 8.3–38.5 W/m², MAPE in 16.7 – 70.7%, and NSE in 0.738–0.986. Regarding QR, the QRF model is higher than the M5T model and lower than RF and XGB models during the training phase. At the same time, it has equivalent performance with the M5T model during the testing phase across all scenarios. In addition, applying the RFNN models through the MFRFNN model for the first time to predict the average daily GSR, the accuracy during the training phase is less than in RF, XGB, M5T, and QRF models. However, the MFRFNN model outperformed the RF, XGB, M5T, and QRF models during the testing phase, and the R² in 0.768–0.987, RMSE in 12.0 – 49.2 W/m², MAE in 8.3–36.3 W/m², MAPE in 18.4 – 75.7%, and NSE in 0.764–0.986. When the GSR prediction success of the VMD-MFRFNN model is compared with other algorithms, the VMD-MFRFNN hybrid model in the training phase outperformed the M5T and MFRFNN models in C1 and C2 combinations but showed poor performance in other cases. In the test phase, while C1 and C2 combinations outperformed different algorithms, C3, C4, and C5 input combinations showed the weakest prediction success, with R² in 0.811–0.893, RMSE in 33.3–44.1 W/m², MAE in 23.7–32.0 W/m², MAPE in 46.1 – 62.9%, and NSE in 0.811–0.892.

The scatter density distribution diagrams of predicted daily average GSR based on RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models under five combinations were shown in Fig. 6 and Fig. 7 across the Växjö and Lund stations, respectively. It can be observed from two figures that

under the first (RF1, XGB1, M5T1, QRF1, MFRFNN1, and VMD-MFRFNN1) and second (RF2, XGB2, M5T2, QRF2, MFRFNN2, and VMD-MFRFNN2) input combinations of meteorological variables, the scatter density plots between the predicted outcomes and observed GSR values are equitably discrete. The agreement between predicted and observed GSR increased in the third and fourth input combinations across all Växjö and Lund stations. However, when it came to the last combination, the RF5, XGB5, M5T5, QRF5, and MFRFNN5 models were all close to the observed values, while the M5T5 and MFRFNN5 models outperforming the other models across two study sites. The VMD-MFRFNN5 model has the greatest dispersed distribution of average daily GSR, indicating that using original input meteorological data is preferable to transforming to signal bands.

Fig. 8 and Fig. 9 display the various distributions of predicted average daily GSR by the six models under different input combinations of meteorological variables and observed daily values across the Växjö and Lund stations, respectively. The findings in Fig. 8 and Fig. 9 showed that the boxplots of daily GSR fluctuated and that the models used improved in accuracy as more meteorological variables were added. In terms of the Växjö station, the median line of observed GSR was 116.9 w/m², whereas the average median lines of predicted GSR values by RF5, XGB5, M5T5, QRF5, MFRFNN5, and VMD-MFRFNN5 models were 119.9, 118.9, 119.6, 118.5, 118.6, and 117.3 W/m², respectively. In terms of Lund station, the median line of observed GSR was 120.6 W/m², whereas the average median lines of predicted GSR values by RF5, XGB5, M5T5, QRF5, MFRFNN5, and VMD-MFRFNN5 models were 125.4, 122.9, 123.8, 122.8, 122.9, and 120.0 W/m², respectively. This also highlighted that the QRF and MFRFNN models were the best models

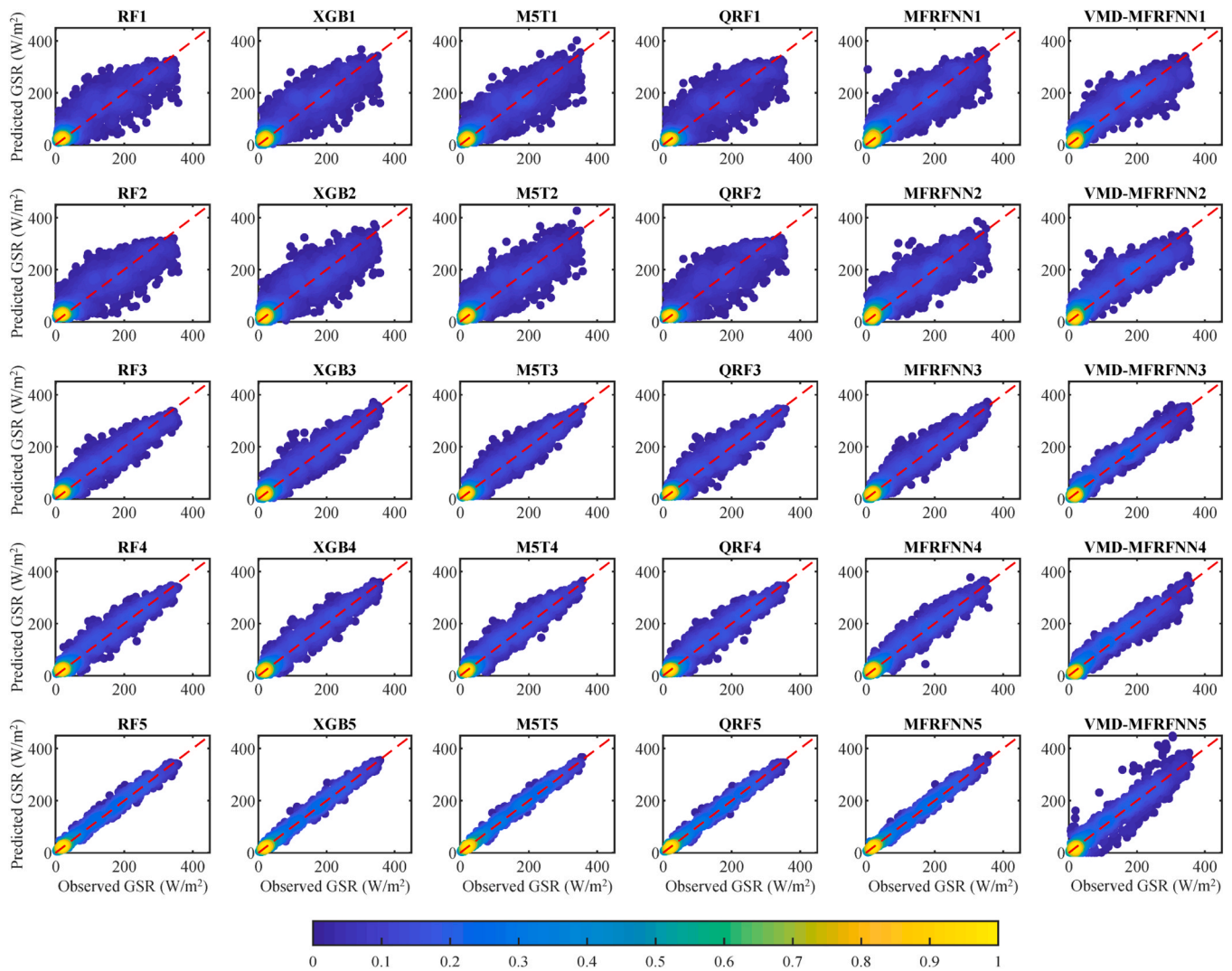


Fig. 6. Comparison between the performance of average daily GSR predicted by RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models based on a different scenario against the observed average daily GSR during the testing period at the Växjö Station.

for predicting daily GSR at both study sites.

3.2. Assessment of input meteorological variables combinations

Fig. 10 and Fig. 11 represent the Taylor diagram group that compares the predicted average daily GSR observed data at the Växjö and Lund stations, respectively. The Taylor diagram is a visualization tool for evaluating comprehensively the performance of models based on correlation coefficient (CC), centered RMSE, and standard deviation (SD) statistics (Taylor, 2001). The close point of model prediction to the observation point represents the best prediction accuracy. The Taylor diagram is super beneficial for comparing the effectiveness of various model predictions (Feng et al., 2017b; Abdallah et al., 2022). The benefits and disadvantages of the employed models are easier to understand by combining the three statistical metrics on the same graphic. It can be seen from Fig. 10 and Fig. 11 that the effectiveness of adding the specific meteorological variable improves the accuracy of GSR prediction at both study sites. The RMSE (red colored arc) values of the first and second input combinations (in Fig. 10a and b) ranged between 40 and 60 W/m^2 , while after adding the SSH reduced to 23–31 W/m^2 , whereas the RH reduced with 24–27 W/m^2 as shown in Fig. 10c and d. Especially, the SSH decreased the RMSE values by 36.3%, 37.2%, 39.6%, 37.9%, 40.2%, and 45.5%, while the RH decreased RMSE values by 46.6%, 44.0%, 46.6%, 46.2%, 45.9%, and 42.5% for RF, XGB, M5T,

QRF, MFRFNN, and VMD-MFRFNN models, respectively. In addition, by adding the maximum possible sunshine duration (N), the prediction of RF, XGB, M5T, QRF, and MFRFNN models increased to the highest level, and the RMSE values reduced by 72.5%, 74.2%, 76.1%, 75.6%, and 76.0%, respectively. However, coupling VMD techniques with the MFRFNN model under the last input combination of all meteorological variables showed a 22.6% increment for the RMSE. When the number of meteorological variables increases, the number of fuzzy rules also increases; as a result, the number of model parameters increases, and the training of the VMD-MFRFNN model becomes more difficult (note that the number of parameters for the VMD-MFRFNN model is nine times that of the MFRFNN model).

For the Lund station, the RMSE (red-colored arcs) values of the first and second input combinations (in Fig. 11a and b) ranged between 40 and 60 W/m^2 , while after adding the SSH reduced to 28–33 W/m^2 , whereas the RH reduced with 28–36 W/m^2 as shown in Fig. 11c and d. Especially, the SSH decreased the RMSE values by 39.9%, 40.1%, 44.5%, 40.4%, 42.5%, and 24.0%, while the RH decreased RMSE values by 43.2%, 40.4%, 45.1%, 43.1%, 42.1%, and 18.5% for RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models, respectively. In addition, by adding the maximum possible sunshine duration (N), the prediction of RF, XGB, M5T, QRF, and MFRFNN models increased to the highest level, and the RMSE values reduced by 72.5%, 73.9%, 76.8%, 76.1%, and 75.6%, respectively. However, coupling VMD techniques with the

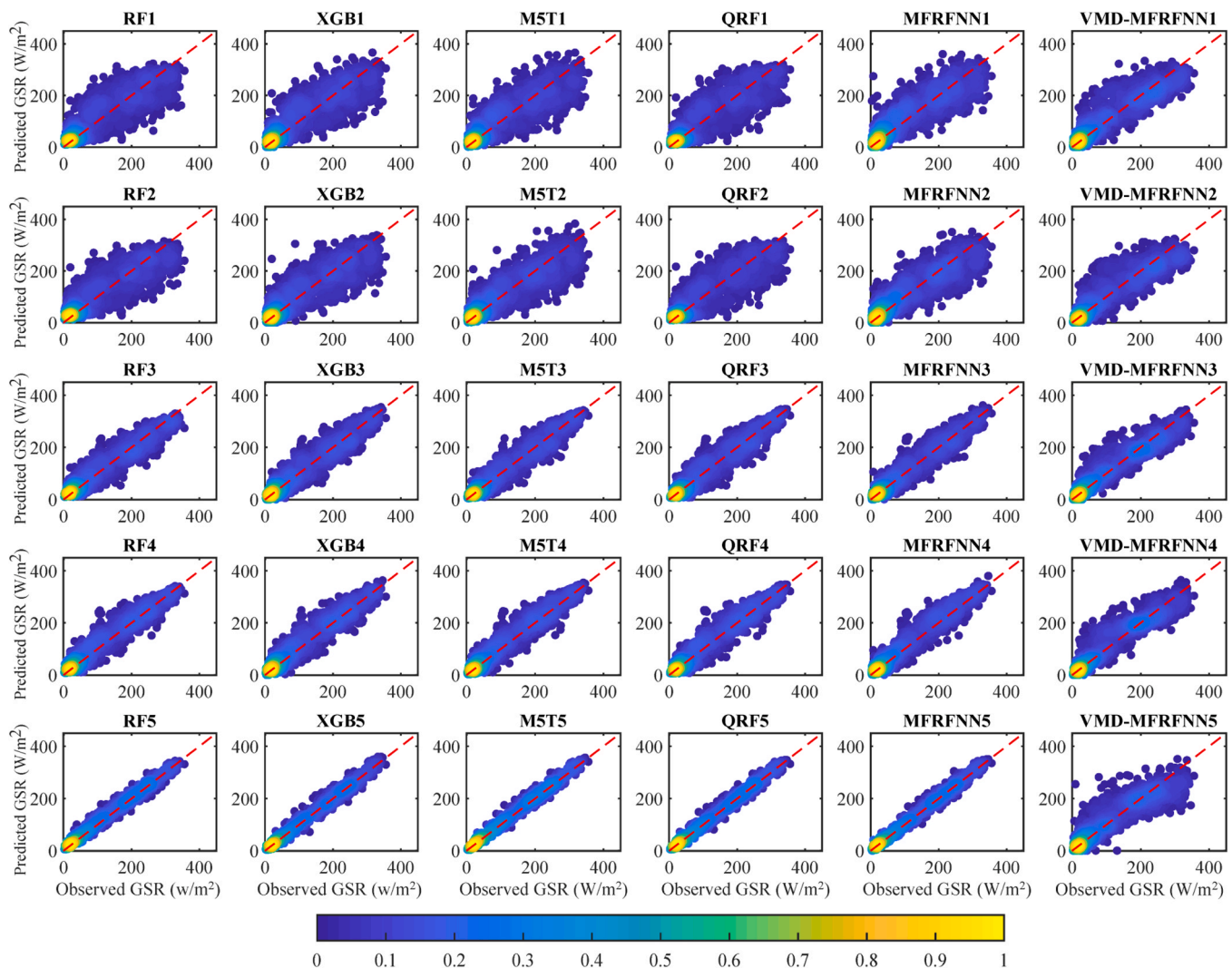


Fig. 7. Comparison between the performance of average daily GSR predicted by RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models based on a different scenario against the observed average daily GSR during the testing period at the Lund Station.

MFRFNN model under the last input combination of all meteorological variables proved a 0.8% increment for the RMSE. The MFRFNN model involves the training of recurrent fuzzy neural networks, which generally have a higher computational cost than ordinary models. We tried to reduce modeling complexity by designing the model with an optimized architecture to balance accuracy and computational efficiency. In the case of QRF model, which combines multiple decision trees, the overall complexity depends on the number of trees in the ensemble. Furthermore, QRF model is highly parallelizable, making it computationally efficient for large datasets.

3.3. Sensitivity analysis of daily GSR prediction

A sensitivity analysis was employed to assess the accuracy of predicted GSR using the QRF model based on different input combinations of meteorological variables and various quantile levels. Fig. 12a and b show the sensitivity in the accuracy of daily GSR prediction assessed against increasing quantile levels of testing phase to change of meteorological variables across the Lund and Växjö stations, respectively. Generally, both figures show similar patterns in which the GSR predicted at the median (GSR q0.50) had the maximum accuracy (lowest RMSE). In contrast, the driest (GSR q0.05) and wettest (GSR q0.95) conditions produced the largest error values throughout both study sites. The first (S1, red box) and second (S2, brown box) input combinations of

meteorological variables reflect the lowest accuracy indicating with highest RMSE, while the fifth input combinations (S5, green box) exhibit the highest accuracy indicating with lowest RMSE, across all quantile levels at both study sites. Overall, the criteria of selection input structure of meteorological variables and appropriate quantile level to predict daily GSR using the QRF model is important to reduce the uncertainty of the prediction.

Figs. 13 and 14 show the simulations of the low and high-frequency IMF components of the VMD-MFRFNN model. When these graphs are examined, it is seen that high-frequency IMF components are modeled with low performance, and low-frequency IMF components are modeled with high performance. The number of fuzzy rules is few to simulate high-frequency IMF components successfully. Thanks to the multi-function learning capacity and structure of the MFRFNN model, it can successfully determine the dynamic properties of the GSR data. The model demonstrates the ability to memorize historical information through the feedback loop. In addition, applying multiple states, the MFRFNN model allows it to store previous values and track system dynamics. Based on a feedback loop, its structure enables the model to memorize previous values and predict future ones. The successful results of the MFRFNN model in this study are related to the superiority of its structure and feedback loop. The good performance of the VMD-MFRFNN model depends on its structure.

Additionally, by combining multiple states and the VMD techniques,

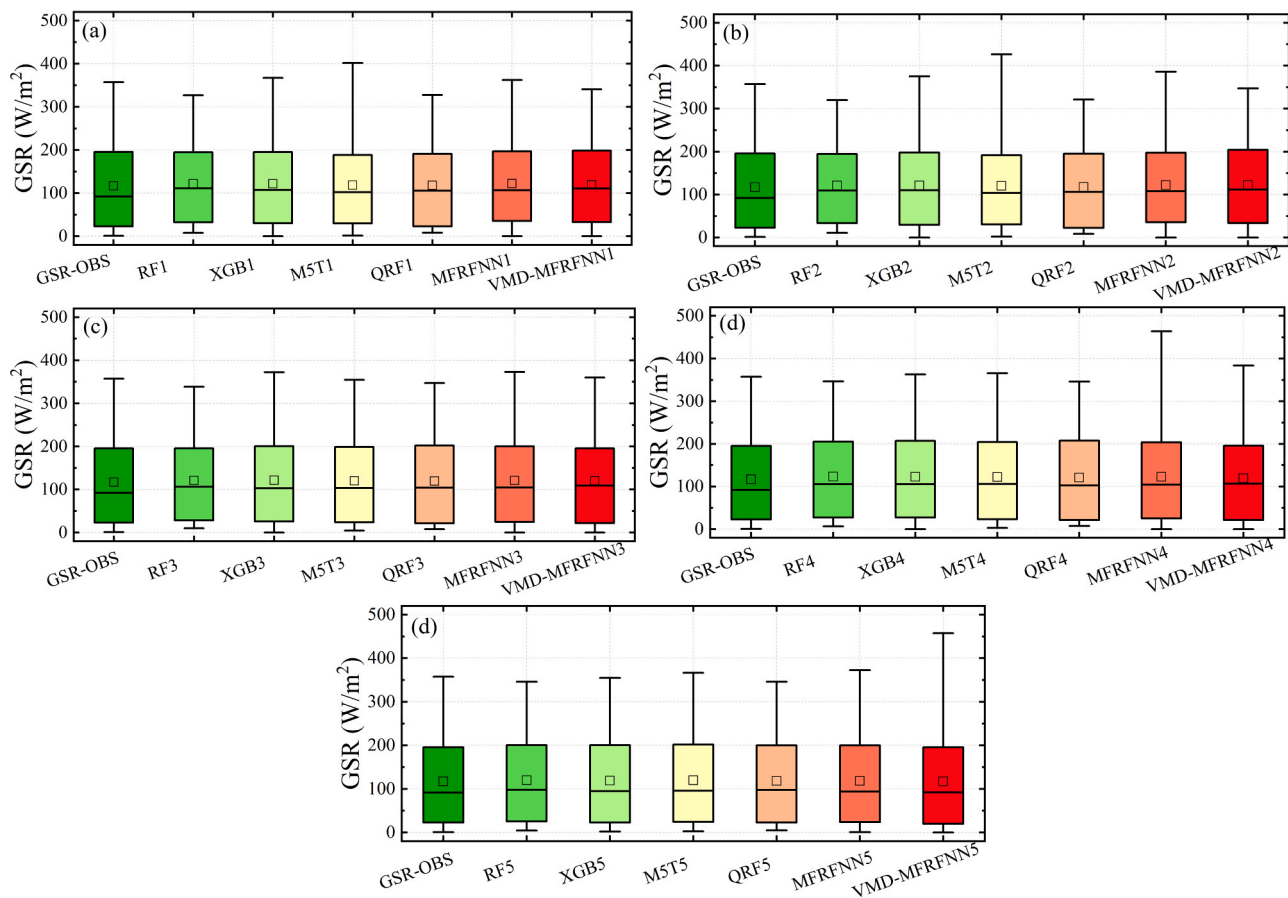


Fig. 8. Boxplots of average daily GSR for observed and predicted by different models; including RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models based on a different scenario at the Växjö Station. The dataset containing 1461 samples from 2018 to 2021 was used to assess the models.

the VMD-MFRFNN hybrid model is able to store past information and display the dynamic characteristics of the time series. Decomposition techniques increase the prediction achievement of ML models by decomposing complex time series into components with simpler structures and more stable tendencies. The VMD-MFRFNN hybrid model's success in estimating GSR is a result of its ability to learn multiple functions and divide the problem into more manageable sub-components. Reducing complexity makes time series estimation easier. The VMD-MFRFNN model's satisfactory performance can be attributed to its ability to learn multiple functions concurrently and employ a decomposition algorithm to facilitate the estimation task. Employing the two states allows the VMD-MFRFNN model to discover and monitor the dynamic behavior of the average daily GSR time series over time. In addition, the VMD-MFRFNN hybrid model eliminates the noise of the dataset with a fuzzy neural network. The VMD-MFRFNN model showed higher accuracy and standard deviation than the MFRFNN model. Due to a high standard deviation, the VMD-MFRFNN model's variance increases, creating a disadvantage. It can be said that the VMD techniques successfully performs and eliminates the mode mixing problem. The VMD-MFRFNN hybrid model has high estimation accuracy, as it decomposes the input data into IMFs. It is worth mentioning that in terms of computational complexity; the MFRFNN model has a complexity of $O(p)$ in the training of output network weight matrix, where p refers the number of training samples, while VMD-MFRFNN's model complexity is $O(Kp)$, where K represents the number of IMFs. In the training of the state network weight matrix, MFRFNN has a complexity of $O(n \times pop)$, where n is the maximum iteration number and pop denotes population size for the PSO algorithm. VMD-MFRFNN model has a complexity of $O(K \times n \times pop)$.

4. Discussion

The CML algorithms based on the tree concept, such as M5T, RF, and XGB models, have grown in popularity for nonlinear tasks because of their inherent advantages and robust prediction abilities (Feng et al., 2017a). RF model performed better than XGB and M5T models during the training phase at both study sites, as shown in Table 4 and Table 5. The results indicate that the employment of the RF model offers advantages consistent with previous studies (Gala et al., 2016; Prasad et al., 2019), which revealed that the RF model successfully predicts daily GSR with acceptable accuracy throughout the training phase. The XGB model exhibited higher accuracy than the RF model during the testing phase at both study sites. The scatter density plots in Fig. 6 and Fig. 7 reflect the performance of employed models and confirm the statistical metrics as shown in Table 4 and Table 5. Huang et al. (2021) compared the capability of 12 ML models to predict monthly solar radiation during 1980–2016 at the Ganzhou station in China. The results demonstrated that the XGB technique has the best prediction ability and the shortest consumption time. The XGB model performed better than the sunshine model based on generalized empirical when using 96 ground radiation stations across China (Qiu et al., 2022). Also, the XGB model outperformed LSTM, multiple linear regression (MLR), and DT regression (Bamisile et al., 2022), which matches the outcomes of our results. Relative speaking, The M5T model showed the highest potential to estimate daily GSR accurately when compared to three ML models at Adana station in Turkey (Keshtegar et al., 2018). The M5T model provided lower computation time to predict daily GSR than the multivariate adaptive regression spline (MARS) and Kernel-based nonlinear extension of the arps decline model (KNEA) (Wu et al., 2019). However, MARS and KNEA models exhibited the highest accuracy and potential

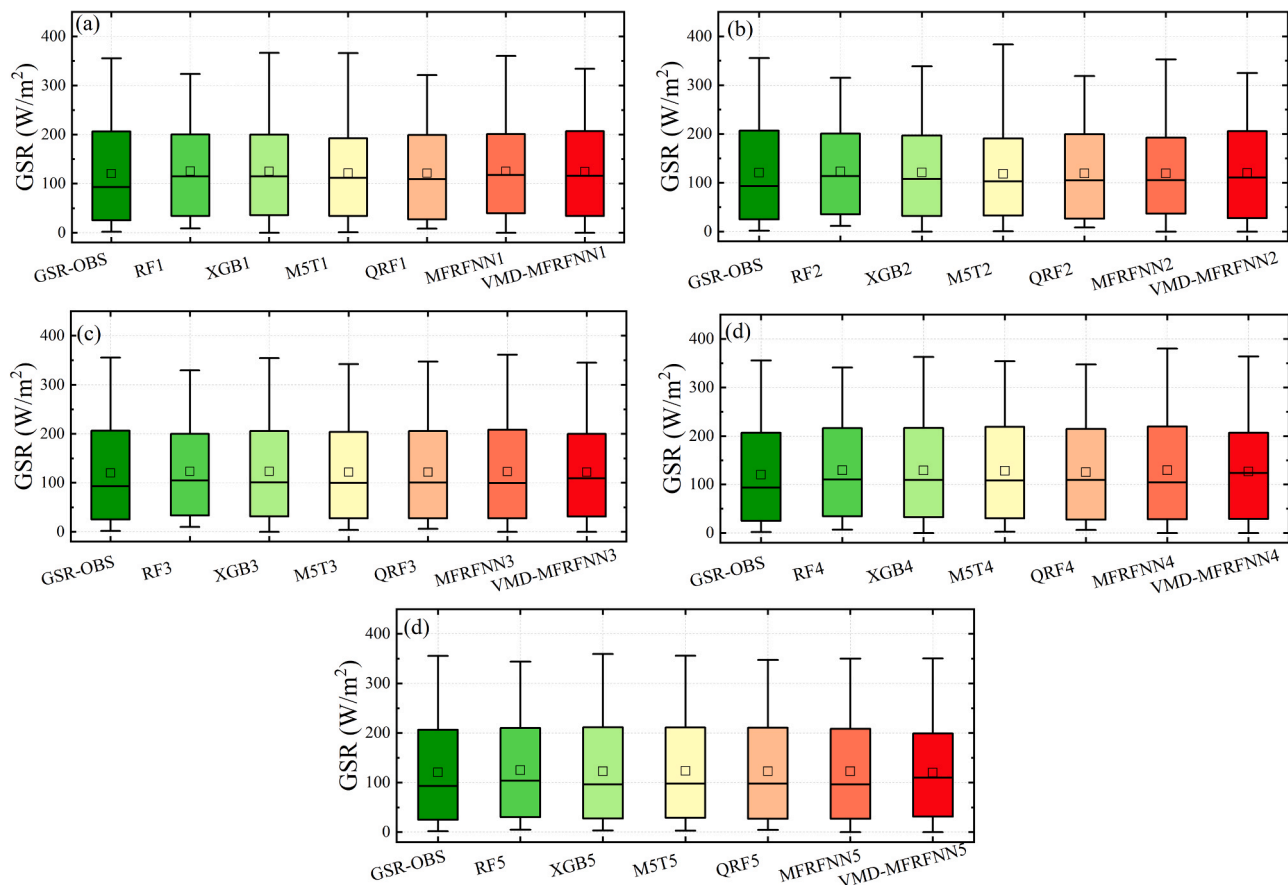


Fig. 9. Boxplots of average daily GSR for observed and predicted by different models; including RF, XGB, M5T, QRF, MFRFNN, and VMD-MFRFNN models based on a different scenario at the Lund station. The dataset containing 1461 samples from 2018 to 2021 was used to assess the models.

than M5T, RF, and multilayer perceptron (MLP) models regarding daily GSR prediction in humid regions. Our results showed that the performance (R^2 , RMSE, MAE, and NSE) of M5T model is better than XGB and RF models at both study sites.

The boxplots distribution was used to reflect the overestimation and underestimation of predicted daily GSR by employed models under different input combinations of meteorological variables at the Växjö and Lund stations, as shown in Fig. 8 and Fig. 9. The QRF model indicates with best performance than RF and XGB models. In contrast, it shows equivalent performance with the M5T model at both study sites. This confirmed that the conditional quantile techniques increase the prediction accuracy in terms of the QRF model more than the conditional mean in the RF model. Overall, the results are similar to the previous studies (Bakker et al., 2019; Bracale et al., 2019; Al Sudani and Salem, 2022), which demonstrated that the QRF model is more capable of predicting the target accurately than XGB, RF, quantile linear regression (QLR), quantile KNN, ELM, and gaussian process regression (GPR) models across different climate zones. This study demonstrates that the nonparametric techniques (QRF) are often more accurate than parametric models (RF and XGB) regarding GSR prediction.

The MFRFNN model performs better than traditional neural networks by using the fuzzy logic approach to overcome uncertainty. It also performs satisfactorily in modeling nonlinear and complex systems (Nasiri and Ebadzadeh, 2023). In the VMD-MFRFNN model, estimation success increases as various input variables are separated into IMFs and modeled. Because when the data is divided into sub-signals, it becomes easier to model, and more stable trends can be obtained. The signal decomposition process decomposes the inputs into several internal modes to be expressed more effectively. In addition, VMD techniques can predict multiple non-stationary components by performing

multi-resolution analysis. Thus, the most effective predictors can be identified, and unnecessary ones removed (Li et al., 2017; Zosso et al., 2017; Nasiri and Ebadzadeh, 2023). Ali et al. (2021) combined various ML algorithms with the VMD signal decomposition approach in solar radiation estimation. According to the results, it has been determined that VMD-ML techniques' training and test performances are better than single ML algorithms. The outputs of the present study overlap with the literature in the case of a limited number of inputs. The C3, C4, and C5 combinations contradict the literature in that the MFRFNN model shows more accurate results than the VMD-MFRFNN model.

This study indicated that the GSR substantially correlates with different meteorological variables, including air temperature (T_{min} , T_{max} , T), SSH, RH, and N at both study sites. The first input combinations based on air temperature were used as a benchmark to assess the impact of adding each meteorological variable on prediction accuracy. It was found that the SSH variable reduced the RMSE in ranging from 36.3% to 40.2% and from 39.9% to 44.5%, while the RH variable in ranging 44.0–46.6% and 40.4–45.1% at the Växjö and Lund stations, respectively. This study demonstrated that the SSH and RH meteorological variables are sensitive in GSR prediction. However, the N variable is the more sensitive meteorological variable which reduces the RMSE by more than 72.5% at both study sites, as shown in Fig. 10 and Fig. 11. This is due to the GSR level's close relationship and direct correlation with SSH, RH, and N (Shamshirband et al., 2016); which is characterized by correlation coefficient (CC) 0.90, -0.82 , and 0.85 across the Växjö station and 0.90 , -0.78 , and 0.85 across the Lund stations, respectively. In addition, the level of clouds and its associated weather conditions are the primary factors that restrict the quantity of GSR. GSR is emitted throughout the air or absorbs as long-wavelength radiation via the surrounding atmosphere. The Air temperature is

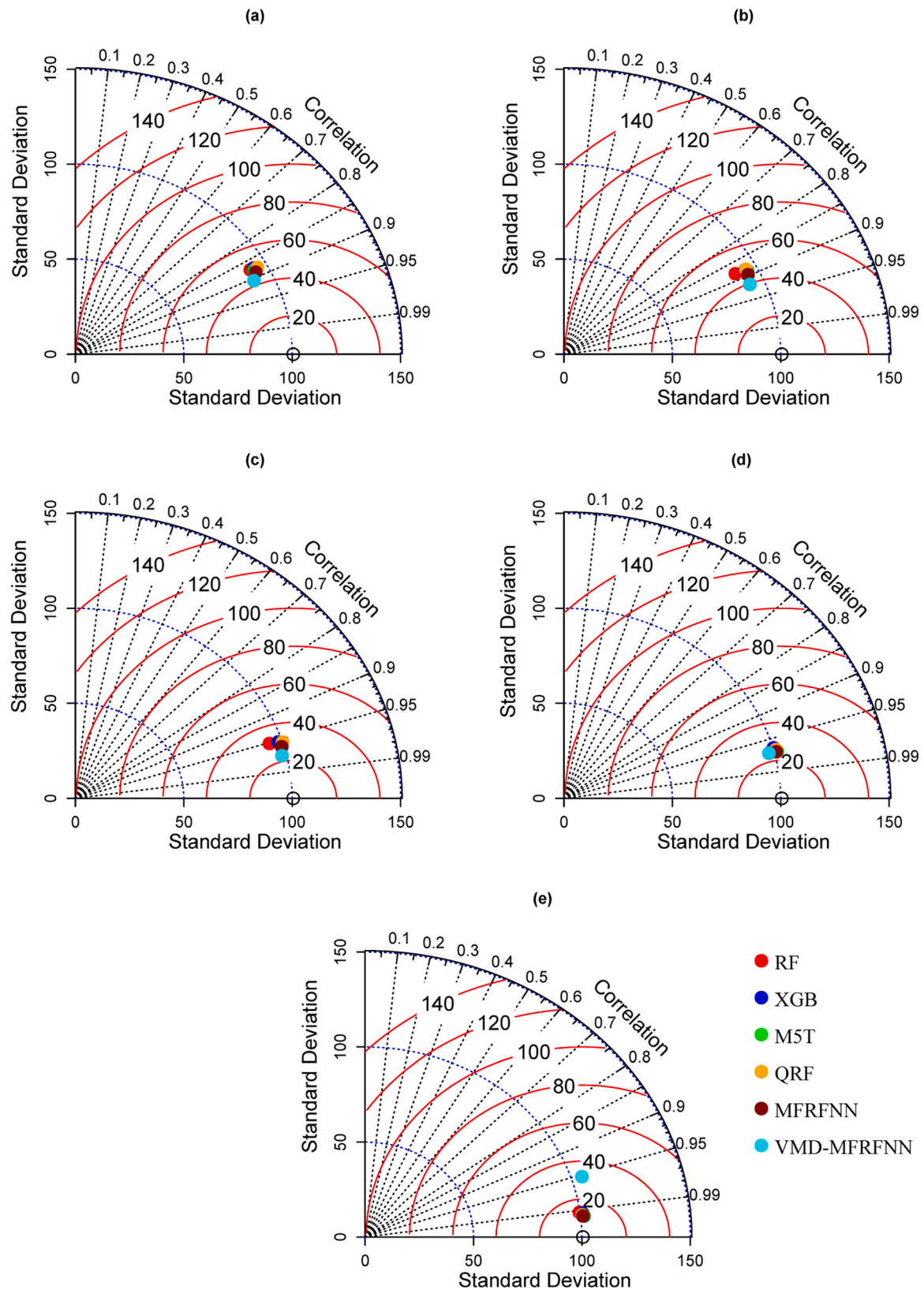


Fig. 10. Comparison of prediction accuracy of average daily GSR using Taylor diagram of different models including RF (red dot), XGB (blue dot), M5T (green dot), QRF (orange dot), MFRFNN (dark red dot), and VMD-MFRFNN (cyan dot) models based on a different scenario against observed GSR (open black circle) at the Växjö Station. The radial black dash line represents the CC, the red-colored arcs represent the RMSE, and the blue dot arcs represent the standard deviation.

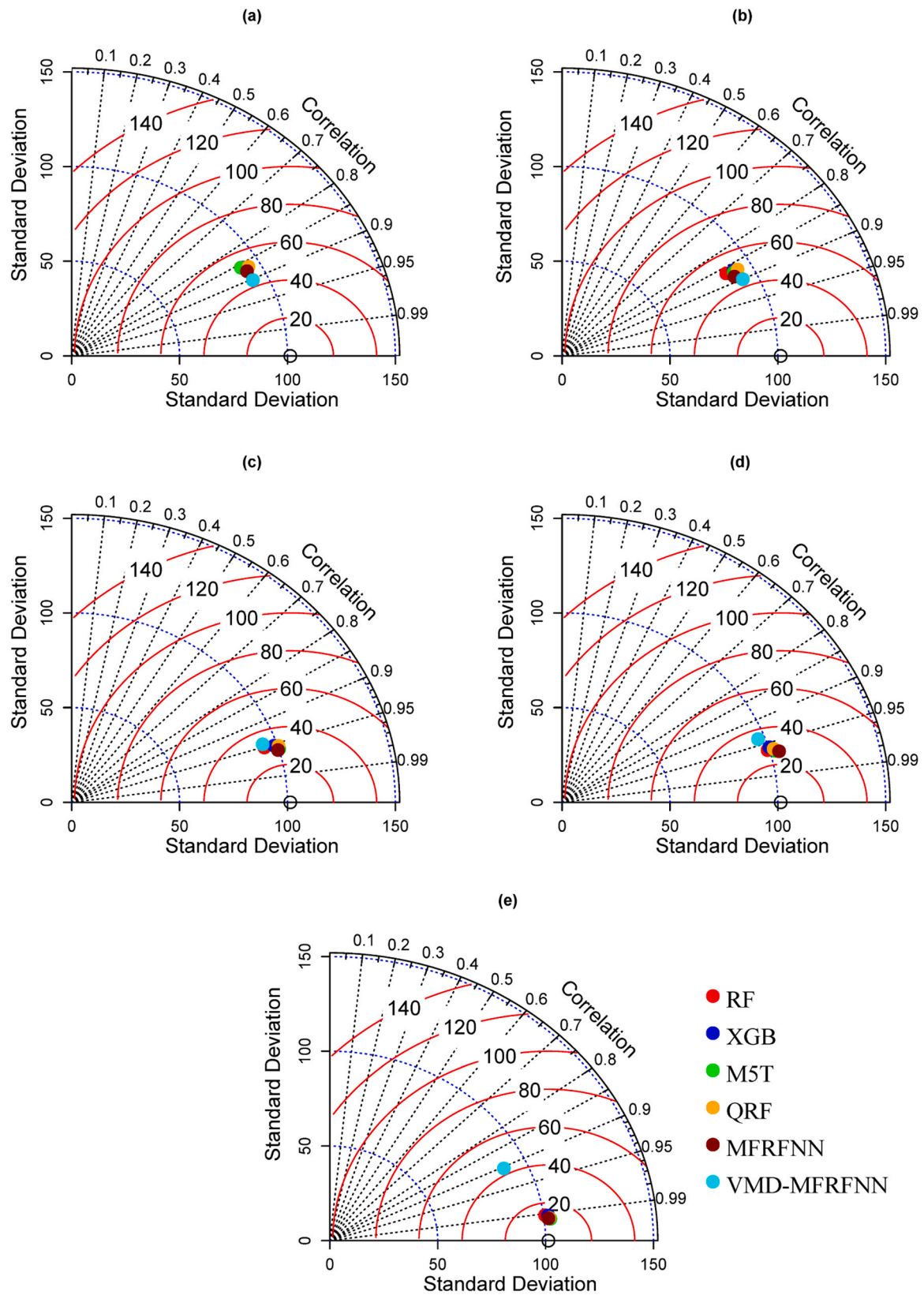


Fig. 11. Comparison of prediction accuracy of average daily GSR using Taylor diagram of different models including RF (red dot), XGB (blue dot), M5T (green dot), QRF (orange dot), MFRFNN (dark red dot), and VMD-MFRFNN (cyan dot) models based on a different scenario against observed GSR (open black circle) at the Lund station. The radial black dash line represents the CC, the red-colored arcs represent the RMSE, and the blue dot arcs represent the standard deviation.

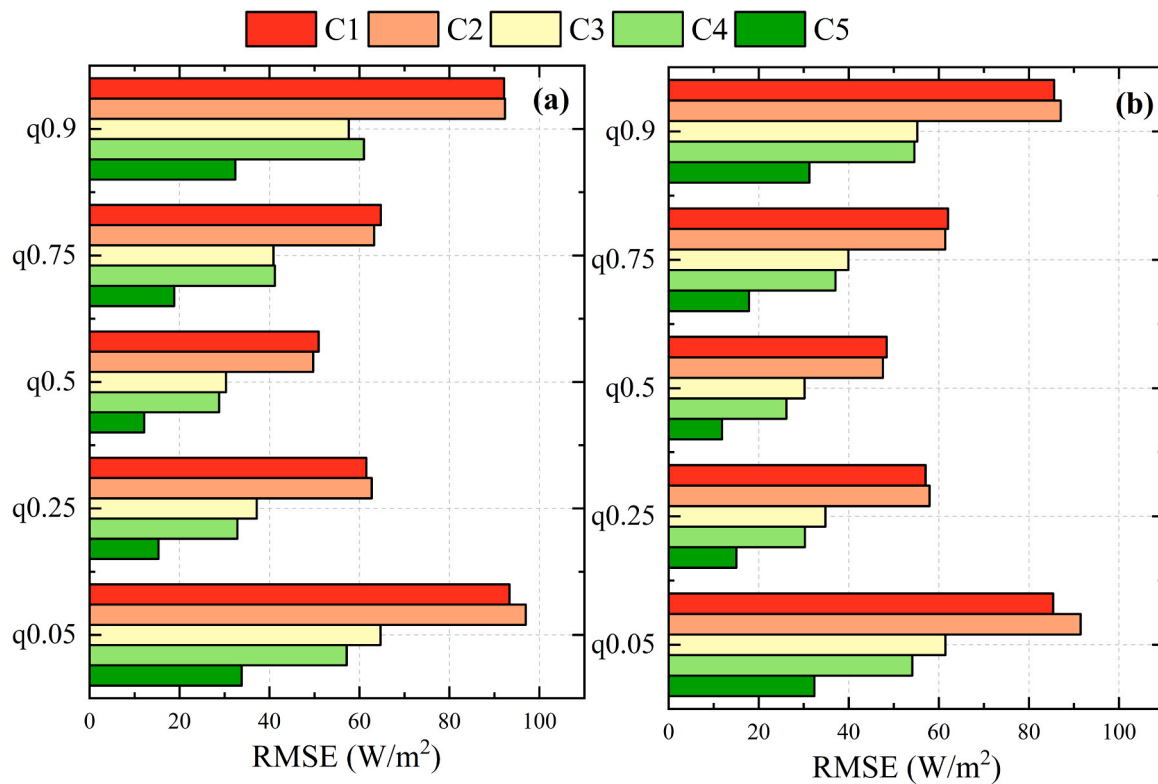


Fig. 12. Sensitivity analysis utilizing RMSE for predicted average daily GSR at various quantile levels of input data at the (a) Lund and (b) Växjö Stations under various scenarios. The input combinations of meteorological variables are represent by C1(Tmin, Tmax, T), C2(Tmin, Tmax, T, WS), C3(Tmin, Tmax, T, WS, SSH), C4 (Tmin, Tmax, T, WS, SSH, RH), and C5(Tmin, Tmax, T, WS, SSH, RH, N).

roughly connected to GSR, albeit not exactly, because the portion that gets taken in by the surrounding atmosphere might cause it to rise (Zhang et al., 2019). Because of this, combining of air temperature (Tmin, Tmax, and T) indicated with accuracy than those based on SSH, RH, and N. He et al. (2020) predicted the GSR at the variety of climate zones over the China mainland. They found that the N, SSH, air temperature and extraterrestrial radiation significantly affected estimating GSR. Similar studies reported that the meteorological variables, including SSH and N had a big effect on how accurate the models were in predicting GSR (Meenal and Selvakumar, 2018; Yao et al., 2018; Zhao et al., 2022). These studies support our finding that solar radiation and SSH/N have a significant positive relationship.

5. Conclusions

This study combines VMD with a MFRFNN model to predict daily GSR across the Lund and Växjö meteorological stations in Sweden. The hybrid VMD-MFRFNN model was assessed to standalone MFRFNN, QRF, RF, XGB, and M5 Tree (M5T) models. The employed models were trained using data from 2008 to 2017, and the prediction performance was validated using data from 2018 to 2021. The following are the most significant outcomes:

- (1) Among the CML models, the RF and M5T models exhibit the highest performance of other models during the training and testing phase across all input combinations at both study sites, respectively. The ranking of CML models during the testing phase is M5T, XGB, and RF models.
- (2) The QRF model resulted in lower performance than the RF model during the training phase, whereas it reflects equivalent performance with the M5T model and outperformed RF and XGB models during the testing phase across all scenarios.

- (3) The MFRFNN model outperformed the CML and QRF models during the testing phase with R^2 ranging 0.789–0.988, RMSE ranging 11.2–46.7 (W/m^2), MAE ranging from 7.9 to 33.7 (W/m^2), MAPE ranging 16.9–78.0%, and NSE ranging 0.768–0.988 at the Växjö station, while with R^2 ranging 0.789–0.987, RMSE ranging 12.0–49.2 (W/m^2), MAE ranging 8.3–36.3 (W/m^2), MAPE ranging from 18.4% to 75.7%, and NSE ranging 0.764–0.986 at the Lund station.
- (4) The hybrid VMD-MFRFNN model showed the best performance than standalone MFRFNN, QRF, and CML models when fewer input variables were used for GSR prediction (first and second input combinations at the Växjö station, while all input combinations except the last combination at the Lund station).
- (5) The most important and beneficial variables that affect the prediction of daily GSR variability in the two study sites are maximum possible sunshine duration (N), RH, and SSH. The SSH, RH, and N reduced the RMSE value of daily GSR prediction by more than 36.3%, 44.5%, and 72.5%, at the Växjö station and 39.9%, 40.4%, and 72.5%, at the Lund station, respectively.

For future studies, optimizing the ML algorithms with bio-inspired algorithms such as gray wolf optimizer, genetic algorithms, firefly algorithm, whale optimization algorithm, and artificial bee colony techniques to increase performance. In addition, it is thought that GSR prediction performances can be improved by combining newly developed signal decomposition techniques such as robust empirical mode decomposition, complete ensemble empirical mode decomposition with adaptive noise, empirical wavelet transforms, Cosine-Sine decomposition, and singular spectrum analysis with ML methods. On the other hand, hybrid VMD-MFRFNN model can be compared with different ensemble ML approaches, such as model averaging, stacking, and bagging to predict GSR across different topographic and climate conditions. The uncertainty analysis of hybrid VMD-MFRFNN model can

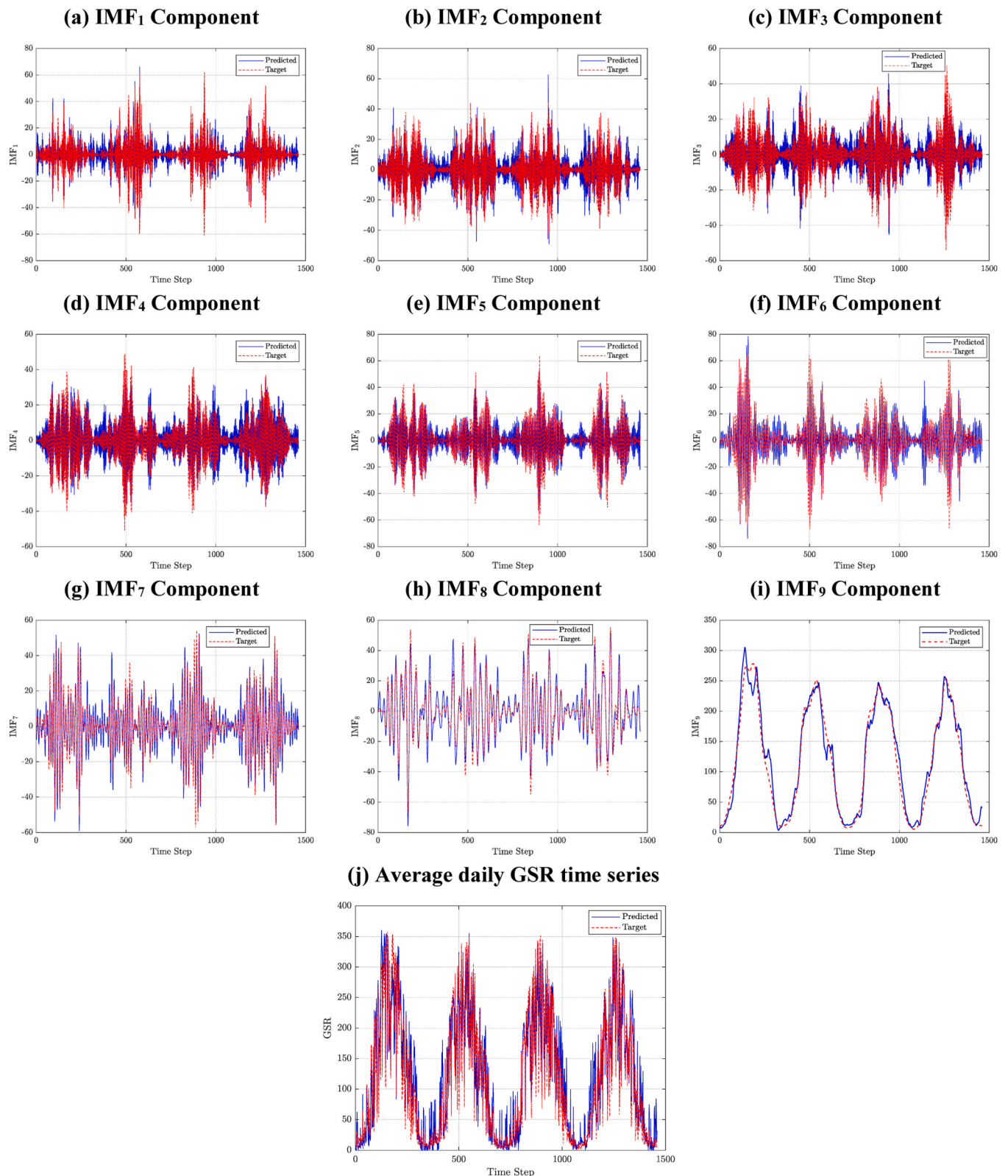


Fig. 13. Predictive curves of VMD decomposition elements for the average daily GSR time series from the meteorological station in Växjö.

take into consideration through various temporal and spatial scales. Understanding and reflection the contribution of input variables to predictions and provide insights into the decision-making process. Techniques like SHapley Additive exPlanations (SHAP) or Local Interpretable Model-Agnostic Explanations (LIME) can be used to convert the

black-box VMD-MFRFNN model to a logical human basis and interpretable system. In terms of improving the data normalization, prediction accuracy, stability, noise reduction, and the correlation between meteorological variables with GSR data the Box–Cox transformation (BC) techniques can hybrid with VMD-MFRFNN model. Furthermore,

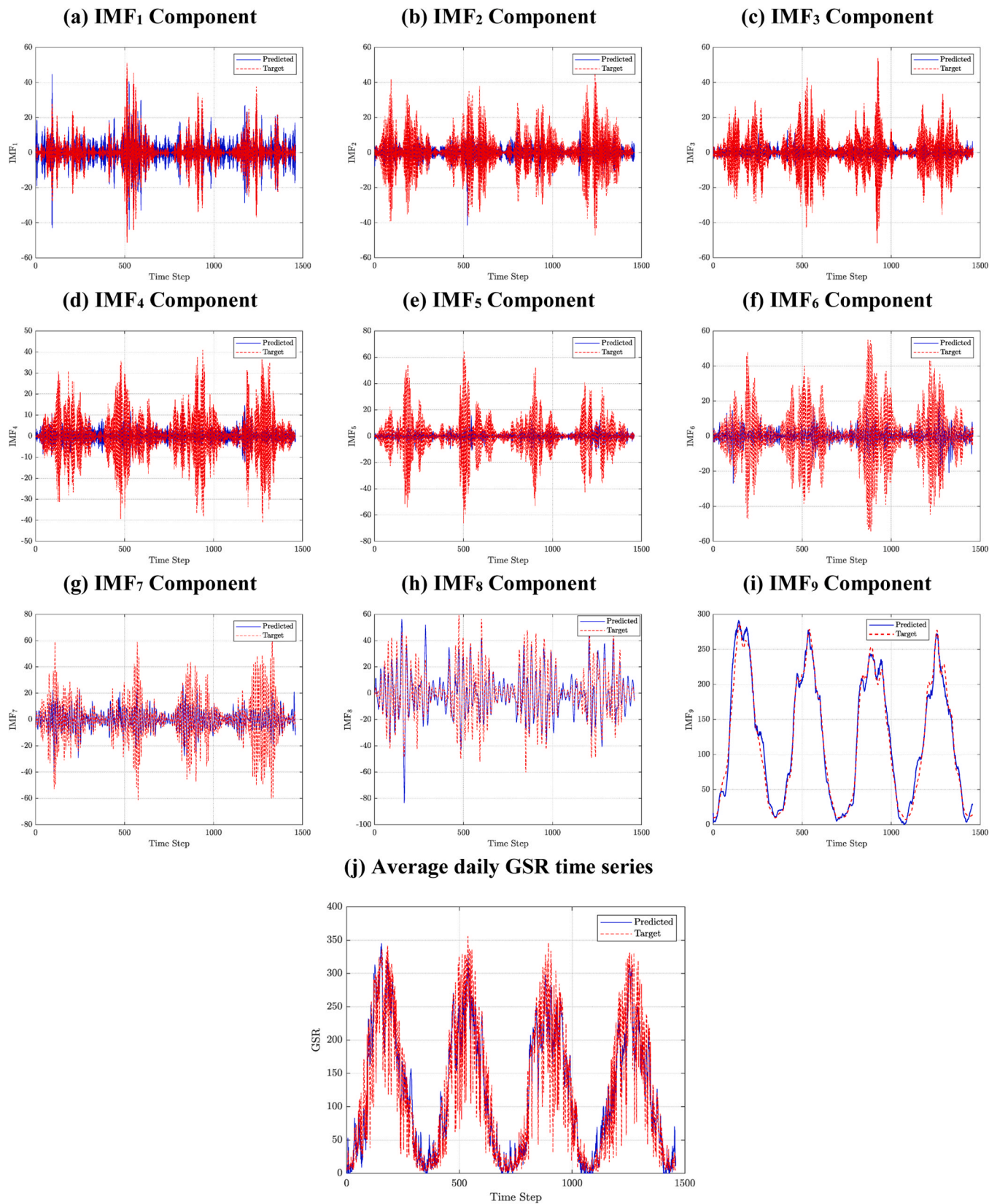


Fig. 14. Predictive curves of VMD decomposition elements for the average daily GSR time series from the meteorological station in Lund.

there is a prospect of investigating the use of remote sensing data and optimized machine learning methodologies in diverse climatic regions in the future. Also, it is recommended to use techniques such as the gamma test, mutual information, Laplacian score algorithm, Recursive Feature Elimination, and Lasso regression in the selection of input combinations to uncover the most effective parameters for global solar radiation prediction and enable more accurate predictions. By providing a reliable estimate of solar radiation, it becomes possible to assist government and policy makers in their planning of renewable energy systems and optimization of power generation sources.

CRedit authorship contribution statement

Mohammed Abdallah: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft. **Babak Mohammadi:** Conceptualization, Data Curation, Investigation, Methodology, Validation, Writing – original draft, Project administration. **Hamid Nasiri:** Conceptualization, Methodology, Software, Validation, Writing – original draft, Visualization. **Okan Mert Katipoğlu:** Investigation, Writing – review & editing. **Modawy Adam Ali Abdalla:** Validation, Writing – review & editing. **Mohammad Mehdi Ebadzadeh:** Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

References

- Abdalla, M.A.A., Min, W., Mohammed, O.A.A., 2020. Two-stage energy management strategy of EV and PV integrated smart home to minimize electricity cost and flatten power load profile. *Energies* 13 (23), 18. <https://doi.org/10.3390/en13236387>.
- Abdallah, M., Mohammadi, B., Zaroug, M.A.H., Omer, A., Cheraghizadeh, M., Eldow, M.E.E., Duan, Z., 2022. Reference evapotranspiration estimation in hyper-arid regions via D-vine copula based-quantile regression and comparison with empirical approaches and machine learning models. *J. Hydrol. -Reg. Stud.* 44, 26. <https://doi.org/10.1016/j.ejrh.2022.101259>.
- Al Sudani, Z.A., Salem, G.S.A., 2022. Evaporation rate prediction using advanced machine learning models: a comparative study. *Adv. Meteorol.* 2022.
- Al-Ghezi, M.K.S., Mahmoud, B.K., Alnasser, T., Chaichan, M.T., 2022. A comparative study of regression models and meteorological parameters to estimate the global solar radiation on a horizontal surface for Baghdad City, Iraq. *Int. J. Renew. Energy Dev.* 11 (1).
- Ali, M., Prasad, R., Xiang, Y., Khan, M., Farooque, A.A., Zong, T., Yaseen, Z.M., 2021. Variational mode decomposition based random forest model for solar radiation forecasting: new emerging machine learning technology. *Energy Rep.* 7, 6700–6717.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, 300. Fao, Rome, p. D05109.
- Bakker, K., Whan, K., Knap, W., Schmeits, M., 2019. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. *Sol. Energy* 191, 138–150. <https://doi.org/10.1016/j.solener.2019.08.044>.
- Bamisile, O., Ejayi, C.J., Osei-Mensah, E., Chikwendu, I.A., Li, J., Huang, Q., 2022. Long-term prediction of solar radiation using XGboost, LSTM, and machine learning algorithms. 2022 4th Asia Energy and Electrical Engineering Symposium (AEEES). IEEE, pp. 214–218.
- Bayrakçı, H.C., Demircan, C., Keçebaş, A., 2018. The development of empirical models for estimating global solar radiation on horizontal surface: a case study. *Renew. Sustain. Energy Rev.* 81, 2771–2782.
- Black, J.N., 1956. The distribution of solar radiation over the earth's surface. *Arch. für Meteorol., Geophys. und Bioklimatol., Ser. B* 7 (2), 165–189.
- Bozoglu, M., Bracale, A., Caramia, P., Carpinelli, G., Carpitia, M., De Falco, P., 2020. Bayesian bootstrap quantile regression for probabilistic photovoltaic power forecasting. *Prot. Control Mod. Power Syst.* 5 (1), 1–12.
- Bracale, A., Carpinelli, G., De Falco, P., 2019. Developing and comparing different strategies for combining probabilistic photovoltaic power forecasts in an ensemble method. *Energies* 12 (6), 16. <https://doi.org/10.3390/en12061011>.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24 (2), 123–140.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Bristow, K.L., Campbell, G.S., 1984. On the relationship between incoming solar-radiation and daily maximum and minimum temperature. *Agric. Meteorol.* 31 (2), 159–166. [https://doi.org/10.1016/0168-1923\(84\)90017-0](https://doi.org/10.1016/0168-1923(84)90017-0).
- Chaitanya, B.K., Yadav, A., Pazoki, M., Abdelaziz, A.Y., 2021. A comprehensive review of islanding detection methods. *Uncertainties Mod. Power Syst.* 211–256.
- Chen, R., Kang, E., Lu, S., Yang, J., Ji, X., Zhang, Z., Zhang, J., 2006. New methods to estimate global radiation based on meteorological data in China. *Energy Conv. Manag.* 47 (18–19), 2991–2998.
- Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. *Proc. 22nd acm sigkdd Int. Conf. Knowl. Discov. data Min.* 785–794.
- Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S., 2019. Machine learning regressors for solar radiation estimation from satellite data. *Sol. Energy* 183, 768–775. <https://doi.org/10.1016/j.solener.2019.03.079>.
- Din, Q., Donchev, T., Kolev, D., 2018. Stability, bifurcation analysis and chaos control in chlorine dioxide-iodine-malonic acid reaction. *MATCH Commun. Math. Comput. Chem.* 79 (3), 577–606.
- Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., Xiang, Y., 2018. Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China. *Energy Conv. Manag.* 164, 102–111.
- Fan, J., Wu, L., Zhang, F., Cai, H., Ma, X., Bai, H., 2019a. Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. *Renew. Sustain. Energy Rev.* 105, 168–186.
- Fan, J.L., Wu, L.F., Zhang, F.C., Cai, H.J., Zeng, W.Z., Wang, X.K., Zou, H.Y., 2019b. Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. *Renew. Sust. Energy. Rev.* 100, 186–212. <https://doi.org/10.1016/j.rser.2018.10.018>.
- Fan, J.L., Wu, L.F., Ma, X., Zhou, H.M., Zhang, F.C., 2020. Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. *Renew. Energy* 145, 2034–2045. <https://doi.org/10.1016/j.renene.2019.07.104>.
- Fatahi, R., Nasiri, H., Homafar, A., Khosravi, R., Siavoshi, H., Chelgani, S.C., 2023. Modeling operational cement rotary kiln variables with explainable artificial intelligence methods - a "conscious lab" development. *Part. Sci. Technol.* 41 (5), 715–724. <https://doi.org/10.1080/02726351.2022.2135470>.
- Feng, Y., Cui, N., Zhang, Q., Zhao, L., Gong, D., 2017a. Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain. *Int. J. Hydrog. Energy* 42 (21), 14418–14428.
- Feng, Y., Cui, N.B., Zhang, Q.W., Zhao, L., Gong, D.Z., 2017b. Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain. *Int. J. Hydrog. Energy* 42 (21), 14418–14428. <https://doi.org/10.1016/j.ijhydene.2017.04.084>.
- Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L., Cui, N., 2019. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Conv. Manag.* 198, 111780.
- Feng, Y., Gong, D., Jiang, S., Zhao, L., Cui, N., 2020. National-scale development and calibration of empirical models for predicting daily global solar radiation in China. *Energy Conv. Manag.* 203, 112236.
- Gala, Y., Fernández, A., Díaz, J., Dorronsoro, J.R., 2016. Hybrid machine learning forecasting of solar radiation values. *Neurocomputing* 176, 48–59.
- Geetha, A., Santhakumar, J., Sundaram, K.M., Usha, S., Theutral, T.M.T., Boopathi, C.S., Ramya, R., Sathyamurthy, R., 2022. Prediction of hourly solar radiation in Tamil Nadu using ANN model with different learning algorithms. *Energy Rep.* 8, 664–671. <https://doi.org/10.1016/j.egy.2021.11.190>.
- Ghimire, S., Deo, R.C., Raj, N., Mi, J.C., 2019. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Appl. Energy* 253, 20. <https://doi.org/10.1016/j.apenergy.2019.113541>.
- Goliatt, L., Yaseen, Z.M., 2023. Development of a hybrid computational intelligent model for daily global solar radiation prediction. *Expert Syst. Appl.* 212, 118295.
- Gyamerah, S.A., Ngare, P., Ikpe, D., 2020. Probabilistic forecasting of crop yields via quantile random forest and Epanechnikov Kernel function. *Agric. Meteorol.* 280, 107808.
- Hameed, M.M., AlOmar, M.K., Mohd Razali, S.F., Kareem Khalaf, M.A., Baniya, W.J., Sharafati, A., AlSaadi, M.A., 2021. Application of artificial intelligence models for evapotranspiration prediction along the southern coast of Turkey. *Complexity* 2021, 118295.
- Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. *J. Irrig. Drain. Div.* 108 (3), 225–230.
- He, C., Liu, J., Xu, F., Zhang, T., Chen, S., Sun, Z., Zheng, W., Wang, R., He, L., Feng, H., 2020. Improving solar radiation estimation in China based on regional optimal combination of meteorological factors with machine learning methods. *Energy Conv. Manag.* 220, 113111.
- He, D.Q., Liu, C.Y., Jin, Z.Z., Ma, R., Chen, Y.J., Shan, S., 2022. Fault diagnosis of flywheel bearing based on parameter optimization variational mode decomposition energy entropy and deep learning. *Energy* 239, 11. <https://doi.org/10.1016/j.energy.2021.122108>.
- Hosseinzadeh, F., Ebrahimi, M., Goliaei, B., Shamabadi, N., 2012. Classification of lung cancer tumors based on structural and physicochemical properties of proteins by bioinformatics models. *PLoS One* 7 (7), e40017.
- Huang, L.X., Kang, J.F., Wan, M.X., Fang, L., Zhang, C.Y., Zeng, Z.L., 2021. Solar radiation prediction using different machine learning algorithms and implications for extreme climate events. *Front. Earth Sci.* 9, 17. <https://doi.org/10.3389/feart.2021.596860>.
- Ikram, R.M.A., Dai, H.L., Ewees, A.A., Shiri, J., Kisi, O., Zounemat-Kermani, M., 2022. Application of improved version of multi verse optimizer algorithm for modeling

- solar radiation. *Energy Rep.* 8, 12063–12080. <https://doi.org/10.1016/j.egy.2022.09.015>.
- Janjai, S., Pankaew, P., Laksanaboonsong, J., Kitichantaropas, R., 2011. Estimation of solar radiation over Cambodia from long-term satellite data. *Renew. Energy* 36 (4), 1214–1220. <https://doi.org/10.1016/j.renene.2010.09.023>.
- Keshtegar, B., Mert, C., Kisi, O., 2018. Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree. *Renew. Sustain. Energy Rev.* 81, 330–341.
- Khorasanizadeh, H., Mohammadi, K., Jalilvand, M., 2014. A statistical comparative study to demonstrate the merit of day of the year-based models for estimation of horizontal global solar radiation. *Energy Conv. Manag.* 87, 37–47. <https://doi.org/10.1016/j.enconman.2014.06.086>.
- Khosravi, A., Nunes, R.O., Assad, M.E.H., Machado, L., 2018. Comparison of artificial intelligence methods in estimation of daily global solar radiation. *J. Clean. Prod.* 194, 342–358. <https://doi.org/10.1016/j.jclepro.2018.05.147>.
- Korkmaz, D., 2021. SolarNet: A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting. *Appl. Energy* 300, 117410.
- Ledmaoui, Y., El Maghraoui, A., El Aroussi, M., Saadane, R., Chebak, A., Chehri, A., 2023. Forecasting solar energy production: a comparative study of machine learning algorithms. *Energy Rep.* 10, 1004–1012.
- Li, F., Lin, Y., Guo, J., Wang, Y., Mao, L., Cui, Y., Bai, Y., 2020. Novel models to estimate hourly diffuse radiation fraction for global radiation based on weather type classification. *Renew. Energy* 157, 1222–1232.
- Li, Z., Chen, J., Zi, Y., Pan, J., 2017. Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive. *Mech. Syst. Signal Process.* 85, 512–529.
- Liu, Y., Zhou, Y., Chen, Y., Wang, D., Wang, Y., Zhu, Y., 2020. Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: A case study in China. *Renew. Energy* 146, 1101–1112.
- Majumder, I., Dash, P.K., Bisoi, R., 2018. Variational mode decomposition based low rank robust kernel extreme learning machine for solar irradiation forecasting. *Energy Conv. Manag.* 171, 787–806.
- Manandhar, P., Temimi, M., Aug, Z., 2023. Short-term solar radiation forecast using total sky imager via transfer learning. *Energy Rep.* 9, 819–828. <https://doi.org/10.1016/j.egy.2022.11.087>.
- Meenal, R., Selvakumar, A.I., 2018. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renew. Energy* 121, 324–343. <https://doi.org/10.1016/j.renene.2017.12.005>.
- Mehdizadeh, S., Mohammadi, B., Pham, Q.B., Duan, Z., 2021. Development of boosted machine learning models for estimating daily reference evapotranspiration and comparison with empirical approaches. *Water* 13 (24), 3489.
- Meinshausen, N., Ridgeway, G., 2006. Quantile regression forests. *J. Mach. Learn. Res.* 7 (6).
- Mohammadi, B., Mehdizadeh, S., 2020. Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. *Agric. Water Manag.* 237, 14. <https://doi.org/10.1016/j.agwat.2020.106145>.
- Mohammadi, B., Moazenzadeh, R., 2021. Performance analysis of daily global solar radiation models in Peru by regression analysis. *Atmosphere* 12 (3), 28. <https://doi.org/10.3390/atmos12030389>.
- Mohammadi, B., Moazenzadeh, R., Pham, Q.B., Al-Ansari, N., Rahman, K.U., Anh, D.T., Duan, Z., 2022. Application of ERA-Interim, empirical models, and an artificial intelligence-based model for estimating daily solar radiation. *Ain Shams Eng. J.* 13 (1), 12. <https://doi.org/10.1016/j.asej.2021.05.012>.
- Muneer, T., Grassie, T., Munawwar, S., 2005. Dedicated to Measurement And Modelling Of Solar Radiation And Daylight-challenges for the 21st Century - Edinburgh, UK - 15-16 September 2003 - editorial introduction. *Energy* 30 (9), 1499–1500. <https://doi.org/10.1016/j.energy.2004.04.030>.
- Nasiri, H., Ebadzadeh, M.M., 2022. MFRNN: multi-functional recurrent fuzzy neural network for chaotic time series prediction. *Neurocomputing* 507, 292–310. <https://doi.org/10.1016/j.neucom.2022.08.032>.
- Nasiri, H., Ebadzadeh, M.M., 2023. Multi-step-ahead stock price prediction using recurrent fuzzy neural network and variational mode decomposition. *Appl. Soft Comput.* 148, 110867. <https://doi.org/10.1016/j.asoc.2023.110867>.
- Ouali, K., Alkama, R., 2014. A new model of global solar radiation based on meteorological data in Bejaia City (Algeria). *Energy Procedia* 50, 670–676.
- Pan, B.Y., Iop, 2017. Application of XGBoost algorithm in hourly PM2.5 concentration prediction, 3rd International Conference on Advances in Energy Resources and Environment Engineering (ICAESSEE). IOP Conference Series-Earth and Environmental Science. IOP Publishing Ltd, Harbin, PEOPLES R CHINA. DOI: 10.1088/1755-1315/113/1/012127.
- Prasad, R., Ali, M., Kwan, P., Khan, H., 2019. Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. *Appl. Energy* 236, 778–792. <https://doi.org/10.1016/j.apenergy.2018.12.034>.
- Prescott, J.A., 1940. Evaporation from a water surface in relation to solar radiation. *Trans. Roy. Soc. S. Aust.* 46, 114–118.
- Qiu, R.J., Liu, C.W., Cui, N.B., Gao, Y., Li, L.A., Wu, Z.J., Jiang, S.Z., Hu, M., 2022. Generalized extreme gradient boosting model for predicting daily global solar radiation for locations without historical data. *Energy Conv. Manag.* 258, 9. <https://doi.org/10.1016/j.enconman.2022.115488>.
- Quinlan, J.R., 1992. Learning With Continuous Classes. 5th Australian joint conference on artificial intelligence. World Scientific, pp. 343–348.
- Ramirez Camargo, L., Dorner, W., 2016. Comparison of satellite imagery based data, reanalysis data and statistical methods for mapping global solar radiation in the Lerma Valley (Salta, Argentina). *Renew. Energy* 99, 57–68. <https://doi.org/10.1016/j.renene.2016.06.042>.
- Salimi, S., Koochaki, A., 2019. An effective method for islanding detection based on variational mode decomposition. *Electrica* 19 (2), 135–145.
- Shamshirband, S., Mohammadi, K., Khorasanizadeh, H., Yee, P.L., Lee, M., Petkovic, D., Zalnezhad, E., 2016. Estimating the diffuse solar radiation using a coupled support vector machine-wavelet transform model. *Renew. Sust. Energy Rev.* 56, 428–435. <https://doi.org/10.1016/j.rser.2015.11.055>.
- Sivakumar, M., Priya, S.J., George, S.T., Subathra, M.S.P., Leebanon, R., Kumar, N.M., 2023. Nine novel ensemble models for solar radiation forecasting in Indian cities based on VMD and DWT integration with the machine and deep learning algorithms. *Comput. Electr. Eng.* 108, 15. <https://doi.org/10.1016/j.compeleceng.2023.108691>.
- Solano, E.S., Dehghanian, P., Affonso, C.M., 2022. Solar radiation forecasting using machine learning and ensemble feature selection. *Energies* 15 (19), 18. <https://doi.org/10.3390/en15197049>.
- Song, K., Yan, F., Ding, T., Gao, L., Lu, S.B., 2020. A steel property optimization model based on the XGBoost algorithm and improved PSO. *Comput. Mater. Sci.* 174, 13. <https://doi.org/10.1016/j.commatsci.2019.109472>.
- Sun, H., Zhao, N., Zeng, X., Yan, D., 2015. Study of solar radiation prediction and modeling of relationships between solar radiation and meteorological variables. *Energy Conv. Manag.* 105, 880–890.
- Tao, H., Ewees, A.A., Al-Sulttani, A.O., Beyaztas, U., Hameed, M.M., Salih, S.Q., Armanuos, A.M., Al-Ansari, N., Voyant, C., Shahid, S., Yaseen, Z.M., 2021. Global solar radiation prediction over North Dakota using air temperature: development of novel hybrid intelligence model. *Energy Rep.* 7, 136–157. <https://doi.org/10.1016/j.egy.2020.11.033>.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.: Atmospheres* 106 (D7), 7183–7192.
- Wang, H., Liu, C.Y., Deng, L., 2018. Enhanced prediction of hot spots at protein-protein interfaces using extreme gradient boosting. *Sci. Rep.* 8, 13. <https://doi.org/10.1038/s41598-018-32511-1>.
- Wang, J., Wang, E.L., Yin, H., Feng, L.P., Zhao, Y.X., 2015. Differences between observed and calculated solar radiations and their impact on simulated crop yields. *Field Crop. Res.* 176, 1–10. <https://doi.org/10.1016/j.fcr.2015.02.014>.
- Wang, Y., Witten, I.H., 1997. Inducing model trees for continuous classes. *Proc. ninth Eur. Conf. Mach. Learn. Citeseer* 128–137.
- Wu, L., Huang, G., Fan, J., Zhang, F., Wang, X., Zeng, W., 2019. Potential of kernel-based nonlinear extension of Arps decline model and gradient boosting with categorical features support for predicting daily global solar radiation in humid regions. *Energy Conv. Manag.* 183, 280–295.
- Yang, K., Koike, T., Ye, B.S., 2006. Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets. *Agric. Meteorol.* 137 (1–2), 43–55. <https://doi.org/10.1016/j.agrformet.2006.02.001>.
- Yang, Y.J., Yang, Y.M., Zhou, W., 2021. Research on a hybrid prediction model for stock price based on long short-term memory and variational mode decomposition. *Soft Comput.* 25 (21), 13513–13531. <https://doi.org/10.1007/s00500-021-06122-4>.
- Yao, W., Zhang, C., Hao, H., Wang, X., Li, X., 2018. A support vector machine approach to estimate global solar radiation with the influence of fog and haze. *Renew. Energy* 128, 155–162.
- Yorukoglu, M., Celik, A.N., 2006. A critical review on the estimation of daily global solar radiation from sunshine duration. *Energy Conv. Manag.* 47 (15–16), 2441–2450. <https://doi.org/10.1016/j.enconman.2005.11.002>.
- Yousuf, M.U., Umair, M., 2018. Development of diffuse solar radiation models using measured data. *Int. J. Green. Energy* 15 (11), 651–662. <https://doi.org/10.1080/15435075.2018.1525738>.
- Zhang, G.D., Band, S.S., Jun, C., Bateni, S.M., Chuang, H.M., Turabieh, H., Mafarja, M., Mosavi, A., Moshlehpour, M., 2021a. Solar radiation estimation in different climates with meteorological variables using Bayesian model averaging and new soft computing models. *Energy Rep.* 7, 8973–8996. <https://doi.org/10.1016/j.egy.2021.10.117>.
- Zhang, J.Y., Zhao, L., Deng, S., Xu, W.C., Zhang, Y., 2017. A critical review of the models used to estimate solar radiation. *Renew. Sust. Energy Rev.* 70, 314–329. <https://doi.org/10.1016/j.rser.2016.11.124>.
- Zhang, W.G., Wu, C.Z., Zhong, H.Y., Li, Y.Q., Wang, L., 2021b. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geosci. Front.* 12 (1), 469–477. <https://doi.org/10.1016/j.gsf.2020.03.007>.
- Zhang, Y., Cui, N.B., Feng, Y., Gong, D.Z., Hu, X.T., 2019. Comparison of BP, PSO-BP and statistical models for predicting daily global solar radiation in arid Northwest China. *Comput. Electron. Agric.* 164, 12. <https://doi.org/10.1016/j.compag.2019.104905>.
- Zhao, S.T., Wu, L.F., Xiang, Y.Z., Dong, J.H., Li, Z., Liu, X.Q., Tang, Z.J., Wang, H., Wang, X., An, J.Q., Zhang, F.C., Li, Z.J., 2022. Coupling meteorological stations data and satellite data for prediction of global solar radiation with machine learning models. *Renew. Energy* 198, 1049–1064. <https://doi.org/10.1016/j.renene.2022.08.111>.
- Zosso, D., Dragomiretskiy, K., Bertozzi, A.L., Weiss, P.S., 2017. Two-dimensional compact variational mode decomposition: spatially compact and spectrally sparse image decomposition and segmentation. *J. Math. Imaging Vis.* 58, 294–320.
- Zuo, H.M., Qiu, J., Jia, Y.H., Wang, Q., Li, F.F., 2022. Ten-minute prediction of solar irradiance based on cloud detection and a long short-term memory (LSTM) model. *Energy Rep.* 8, 5146–5157. <https://doi.org/10.1016/j.egy.2022.03.182>.