Expert Systems With Applications 222 (2023) 119811

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A novel global solar exposure forecasting model based on air temperature: Designing a new multi-processing ensemble deep learning paradigm

Mehdi Jamei [a,*], Masoud Karbasi [b], Mumtaz Ali [c,h], Anurag Malik [d], Xuefeng Chu [e], Zaher Mundher Yaseen [f,g]

[a] Faculty of Engineering, Shohadaye Hoveizeh Campus of Technology, Shahid Chamran University of Ahvaz, Dashte Azadegan, Iran
[b] Water Engineering Department, Faculty of Agriculture, University of Zanjan, Zanjan, Iran
[c] UniSQ College, Univrsity of Southern Queensland, QLD 4350, Australia
[d] Punjab Agricultural University, Regional Research Station, Bathinda, Punjab, India
[e] Department of Civil, Construction & Environmental Engineering, North Dakota State University, Fargo, ND, USA
[f] Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia
[g] Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabi
[h] School of Climate Change and Adaptation, University of Prince Edward Island, University Ave, Charlottetown PE, Canada

ARTICLE INFO

Keywords:
Solar exposure
Ensemble deep random vector functional link
Deep learning
Empirical mode decomposition
Extra tree feature selection
Time-varying filtering based empirical mode decomposition

ABSTRACT

The total quantity of solar energy falling on a horizontal plane surface is the global solar exposure (GSE, i.e., total solar energy). Precise forecasting of GSE is important in many fields such as renewable energy, agriculture, and public health, particularly by the limited hydro-meteorological time series information. This research aims to develop an advanced multi-processing deep learning (DL) paradigm to forecast weekly GSE based on maximum (T_{max}) and minimum (T_{min}) air temperatures as the drivers at Brisbane and Perth airport stations in eastern and western Australia during 2000 to 2022. The proposed model was comprised of an extra tree feature selection (FS) integrated with two novel decomposition techniques, namely time-varying filtering-based empirical mode decomposition (TVF-EMD), and empirical wavelet transform (EWT), and a powerful ensemble deep random vector functional link (ED-RVFL) approach. To validate the main model, the RVFL, bidirectional long-short term memory (Bi-LSTM), and bagged regression tree (Bagging) machine learning (ML) models were examined in hybrid and standalone counterpart frameworks. First, the extra-tree FS determined the significant lags of the predictors based on an importance benchmark criterion. Then, by applying the optimal gained lags to the feeding models, all of the original predictors were decomposed using TVF-EMD and EWT univariate feature extraction. The final forecast was computed by aggregating all the individual forecasts of the intrinsic mode functions (IMFs) and residual components. In addition, eight statistical indicators (including coefficient of correlation: R, root mean square error: RMSE, Kling-Gupta efficiency: KGE, index of agreement: I_A, uncertainty coefficient with 95% confidence level: $U_{95\%}$, mean absolute percent error: MAPE, Nash-Sutcliffe efficiency: NSE, and mean absolute error: MAE) and several graphical methods were utilized to evaluate the performances of the models. The modeling results indicated that the ED-R_{VFL}-$T_{VF\text{-}EMD}$ (R = 0.9665, RMSE = 1.9193 MJ/m^2, and KGE = 0.9565 for the Perth airport station) and ED-R_{VFL}-E_{WT} (R = 0.9218, RMSE = 1.9708 MJ/m^2, and KGE = 0.8552 for the Brisbane station) outperformed all other models, followed by RVFL, Bi-LSTM, and Bagging in a hybrid format. With the high predictive robustness, both decomposition-based frameworks can be useful for solving energy forecasting problems. The new modeling approach developed in this study can provide more precise forecasts for decision-makers to better address climate change, agriculture, and energy crises.

* Corresponding author.
E-mail address: M.jamei@shhut.ac.ir (M. Jamei).

https://doi.org/10.1016/j.eswa.2023.119811
Received 21 October 2022; Received in revised form 29 January 2023; Accepted 5 March 2023
Available online 8 March 2023

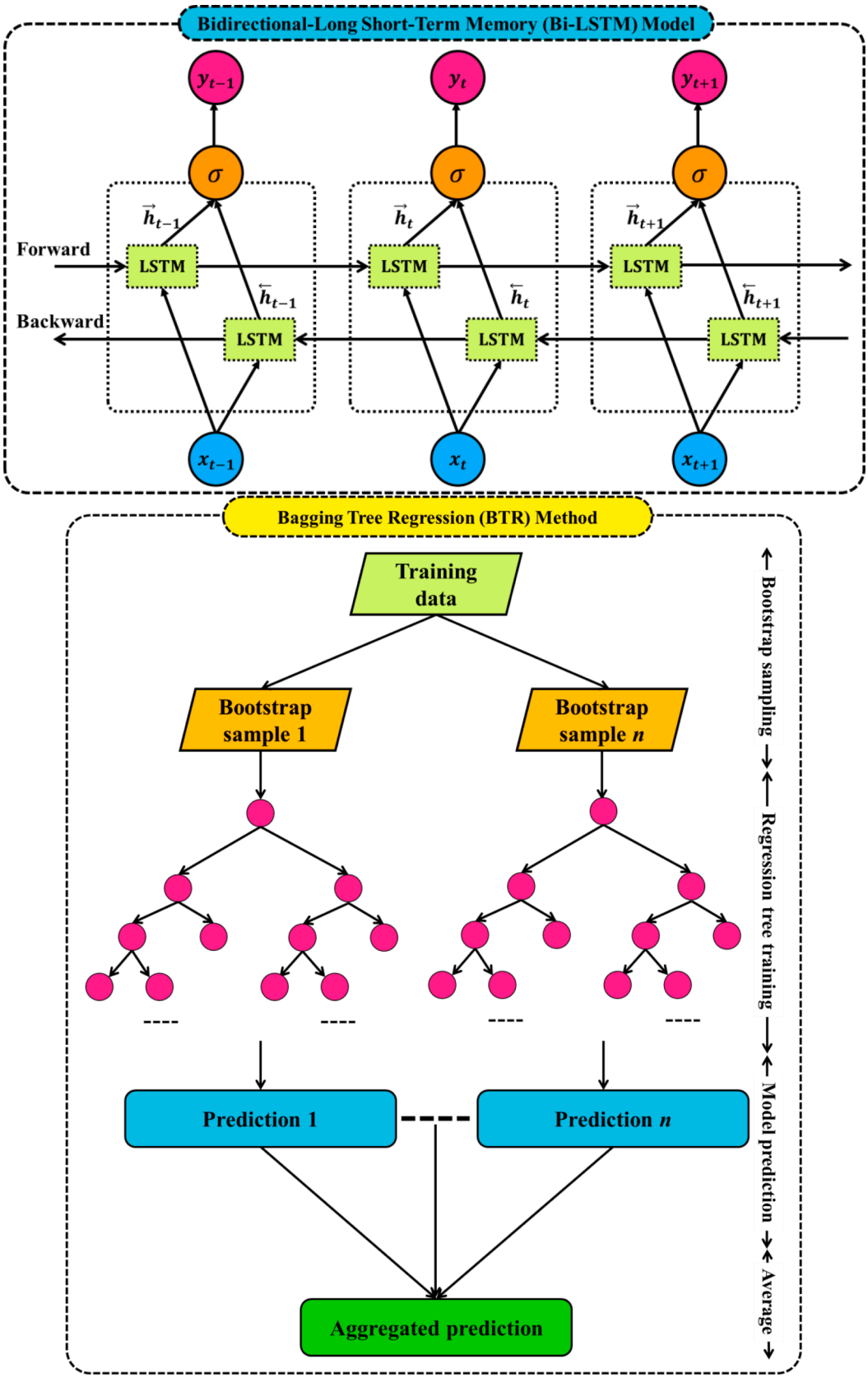


Fig. 1. Structure of the Bi-LSTM network (upper panel) and systematic illustration of the ensemble bagging tree regression method (lower panel).

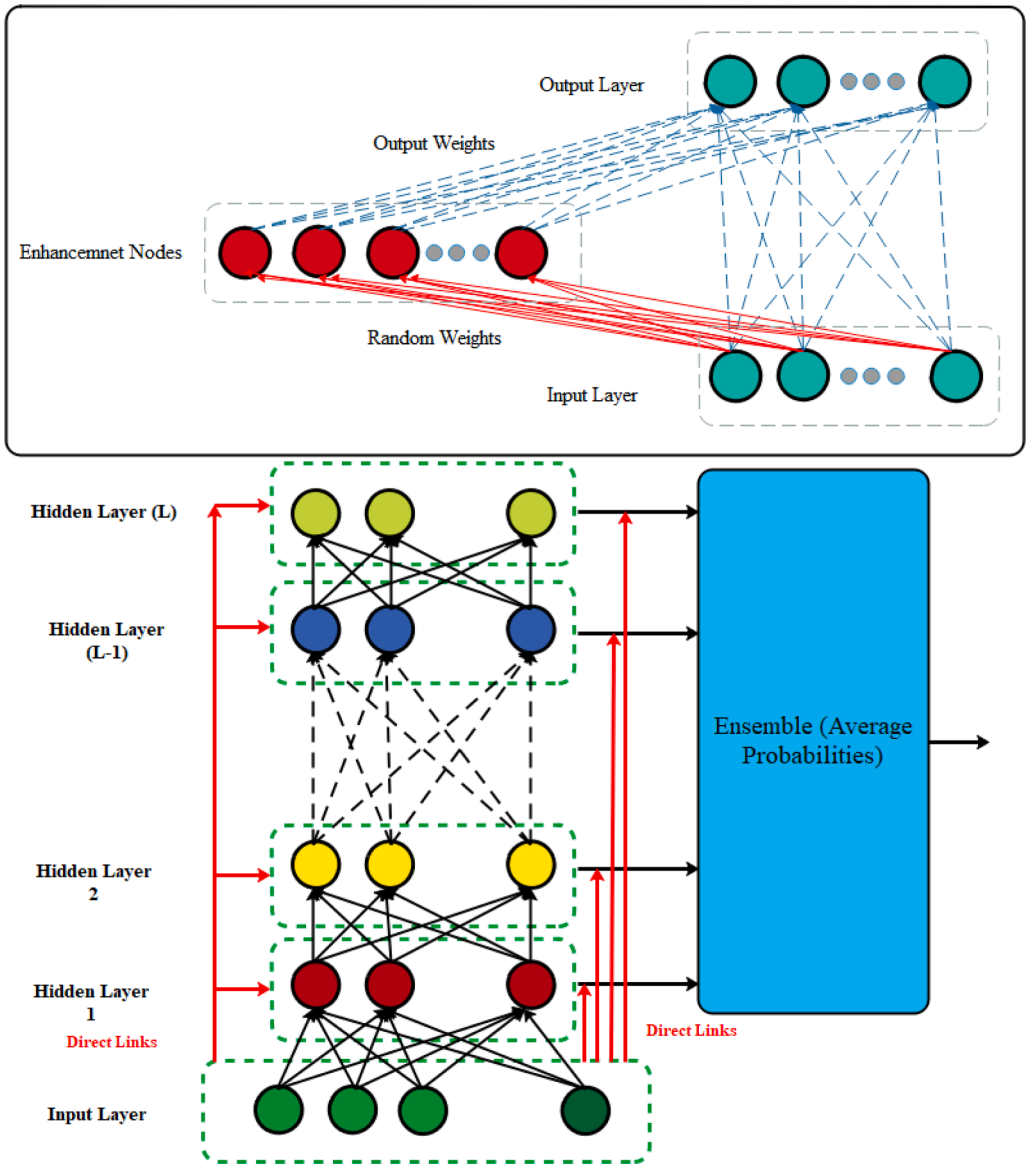


Fig. 2. Schematic of the modeling structures of RVFL (upper panel) and ED-RVFL (lower panel).

Table 1
Descriptive statistics of the temperature data used for GSE forecast for Brisbane and Perth airport stations.

Station	Brisbane			Perth airport		
Indices	T_{max} (°C)	T_{min} (°C)	Observed GSE (MJ/m^2)	T_{max} (°C)	T_{min} (°C)	Observed GSE (MJ/m^2)
Minimum	18.01	5.057	9.757	15.57	2.129	6.329
25% Percentile	23.74	12.73	14.76	20.33	9.029	11.81
Median	27.01	16.99	18.6	24.91	11.91	18.16
75% Percentile	29.26	20.23	22.56	30.16	15.82	25.6
Maximum	34.77	24.59	30.59	39.41	22.33	33.41
Range	16.76	19.53	20.83	23.84	20.2	27.09
Mean	26.62	16.45	18.85	25.38	12.4	18.92
Std. Deviation	3.318	4.299	4.861	5.518	4.175	7.45
Std. Error of Mean	0.09685	0.1255	0.1419	0.161	0.1218	0.2174
Coefficient of variation	12.46%	26.14%	25.79%	21.74%	33.66%	39.37%
Skewness	−0.1651	−0.2501	0.2257	0.2787	0.215	0.2033
Kurtosis	−0.8635	−1.062	−0.8963	−1.118	−0.863	−1.278
Coordinate	Latitude: −27.4705 S Longitude:153.0260 E			Latitude: 31.9385° S Longitude:115.9672° E		

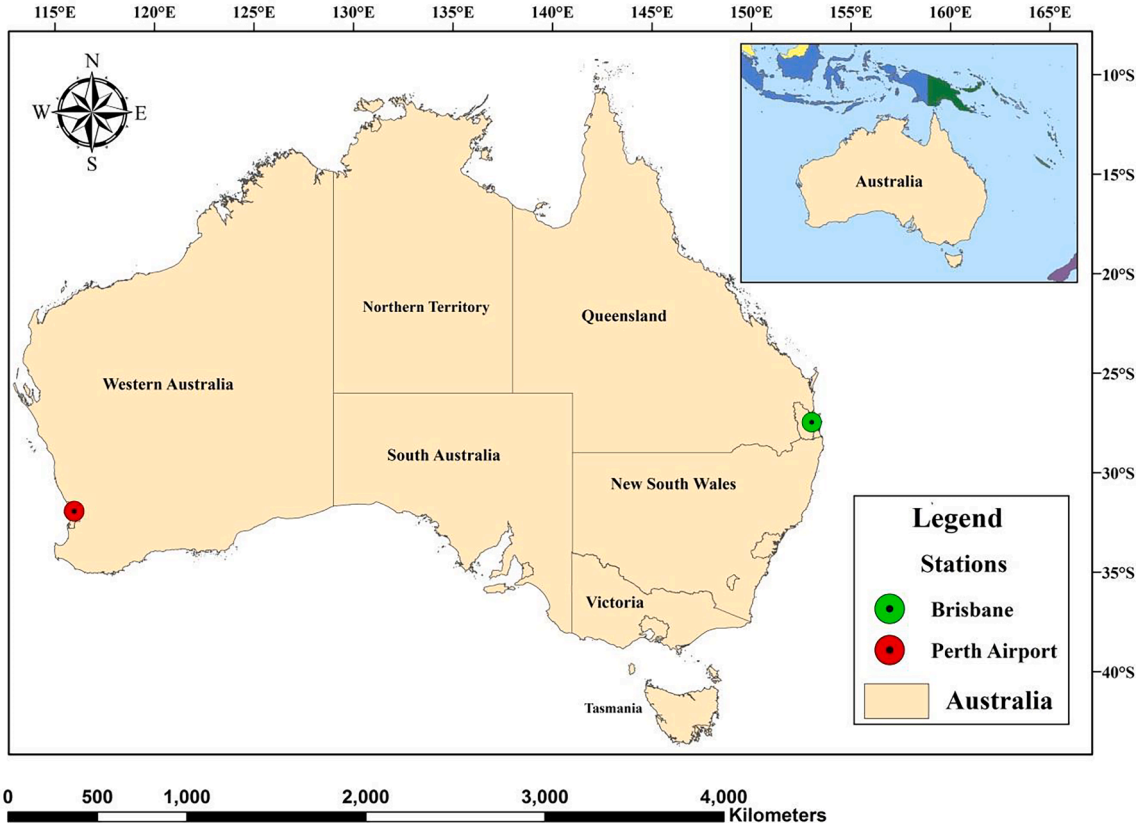


Fig. 3. Study area and the locations of Perth airport and Brisbane stations.

1. Introduction

1.1. Research background

The ongoing use of fossil fuels will continuously cause serious environmental pollution (Wu, Qin, & Liu, 2022). Energy crisis is becoming a crucial issue (Manieniyan, Thambidurai, & Selvakumar, 2009; Tracker, 2022) and hence solutions are needed to address this pressing issue. It is also noteworthy that research has been conducted to use renewable energy to address the related problems (Hai et al., 2020; Voyant, Notton, Duchaud, Almorox, & Yaseen, 2020). Solar energy, as a clean and renewable source of energy, meets the needs for steady and dependable electricity (Kannan & Vakeesan, 2016). However, solar radiation $S_{(r)}$ varies erratically, depending on many factors such as clouds and air pollution. The safe and steady operation of an electricity system can be significantly impacted (Ali & Mohamed, 2022). Thus, accurate forecast of $S_{(r)}$ can help maximize the utilization of solar energy resources, while also enhancing the safety and stability of power system operation (Sharafati et al., 2019). This will lead to the realization of the social and economic benefits, while also encouraging the healthy growth of renewable energy (Bokde, Yaseen, & Andersen, 2020). However, due to the high investment and maintenance costs, the global $S_{(r)}$ data are lacking or incomplete in many regions of the world. Hence, the assessment and prediction of global $S_{(r)}$ have been important tasks over the past decade (El Mghouchi, Chham, Zemmouri, & El Bouardi, 2019). Applications of global $S_{(r)}$ contribute remarkably to diverse engineering fields.

1.2. Literature review

Various models have been developed and used for $S_{(r)}$ modeling, such as empirical models (Fan et al., 2019b; Jahani, Dinpashoh, & Nafchi, 2017), remote sensing methods (Olpenda, Stereńczak, & Będkowski, 2018; Şenkal, 2010), physical models (Cogliani, Ricchiazzi, & Maccari, 2007; Gautier, Diak, & Masse, 1980), and advanced computer-aid models (Hai et al., 2020; Kisi, Heddam, & Yaseen, 2019; Sada, Faskari, Ilyasu, & Abba, 2022). The ultimate goal of the modeling efforts is to develop more robust and reliable methodologies and attain more accurate predictions. The physical models involve direct calculation of the reflection of $S_{(r)}$ from the ground by combining the altitude angle of the sun and the azimuth angle of the ground, which is challenging and time-consuming (Kaaya, Ascencio-Vásquez, Weiss, & Topič, 2021). In contrast, the remote sensing technique predicts $S_{(r)}$ using high-resolution images and geographic data (Huang et al., 2022). The empirical methods are the widely-used methods for $S_{(r)}$ prediction because of their low computing cost and data demand. Several meteorological variables have been employed to build different empirical models (Fan et al., 2019a). Such models include hybrid models, models of sunshine duration, and models of $S_{(r)}$ (Bailek et al., 2018). The comparatively high volatility of $S_{(r)}$ has a significant impact on the performances of empirical models. Hence, this research suggests the exploration of some advanced versions of machine learning (ML) models that can more accurately predict $S_{(r)}$ while addressing the drawbacks of the three other types of models.

ML models have been widely used for forecasting climate-related variables such as $S_{(r)}$ (Ghimire, Deo, Downs, & Raj, 2019; Khosravi, Koury, Machado, & Pabon, 2018; Kumari & Toshniwal, 2021a; Kumari & Toshniwal, 2021b; Qazi et al., 2015). The newly enhanced ML models have also been applied for modeling $S_{(r)}$. Acikgoz (2022) developed a model for $S_{(r)}$ prediction, which consisted of a feature selection phase that relied on continuous wavelet transform, an extreme learning model for forecasting, complete ensemble empirical model decomposition with adaptive noise (CEEMDAN), and feature extraction networks. Bamisile et al. (2022) predicted hourly global and diffuse $S_{(r)}$ in Nigeria using a DL model and compared with many benchmark ML models. The DL models have been proven to be more reliable than other ML models.

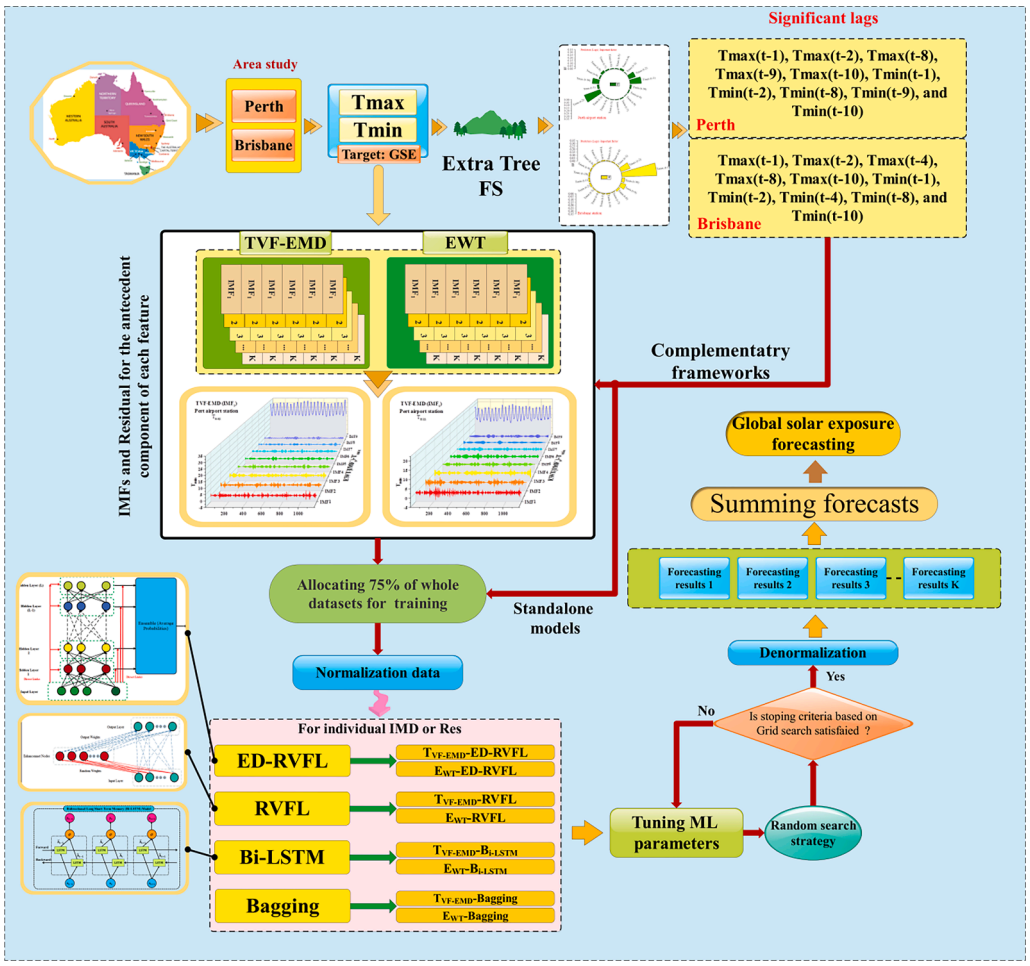


Fig. 4. Flowchart of the proposed multi-stage ML-based models for forecasting the weekly GSE at two stations in Australia using meteorological drivers.

Ghimire et al. (2022a) proposed a 4-phase modeling approach for simulating global $S_{(r)}$, which combined DL, enhanced random forest with extreme gradient, and the Harris Hawks algorithm (HHA). Abubakr et al. (2022) utilized the transfer learning technology to develop a better $S_{(r)}$ prediction model based on the capabilities of DL neural networks. Ghimire et al. (2022b) developed a hybrid DL model via the combination of two-stage stacked regression models for simulating daily global $S_{(r)}$, which demonstrated the potential to function as an effective tool for contemporary solar energy monitoring. Optimization of the prediction feature matrix is another essential component in developing ML models for $S_{(r)}$ prediction. Goliatt and Yaseen (2022) developed the covariance matrix adaptive evolution strategy with extreme gradient boosting and demonstrated its potential for modeling $S_{(r)}$ in an arid area.

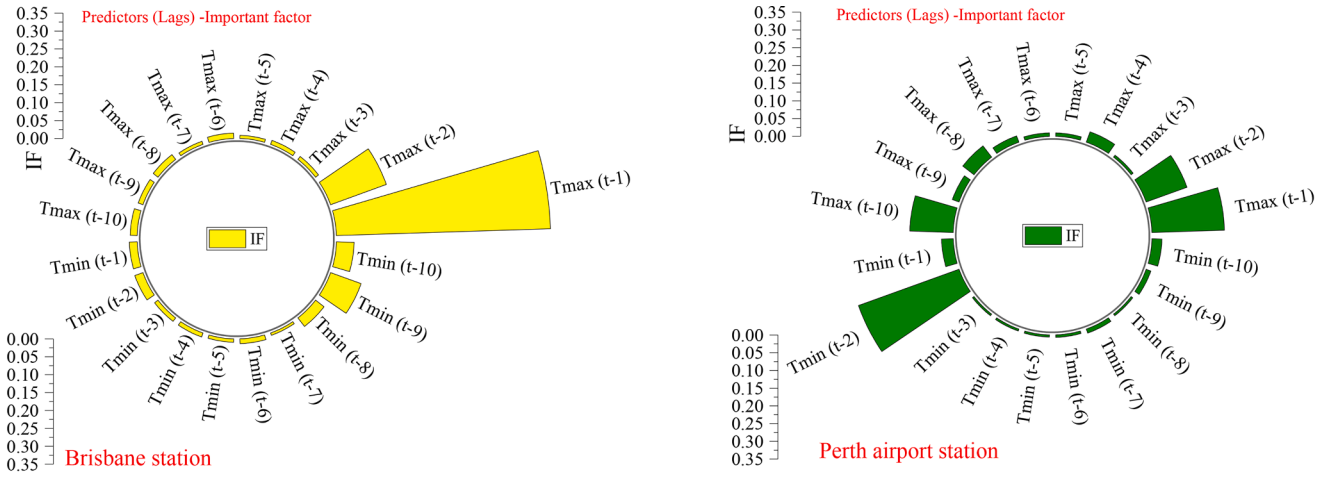
1.3. Research gap and motivation

A variety of ML models such as extra tree, support vector machine, evolutionary programming, fuzzy logic, and even the latest version of DL have been applied for $S_{(r)}$ modeling. The implementation of ML models for such highly stochastic natural problems can vary from one case to another, depending on the type of data, learning process, learning mechanism, optimization of input parameters, and other conditions. Hence, there are research gaps to be filled and room for further exploration.

Pre-processing the original signals or data has been performed to strengthen the reliability of a model and to boost the accuracy of the $S_{(r)}$ prediction. Non-stationary data can be modeled by using the wavelet transform (WT) to compress the size of the original data (Sharma, Yang, Walsh, & Reindl, 2016). Peng et al. (2021) used WT to perform threshold denoising to increase the prediction accuracy. However, there is a great restriction to the use of WT for high-dimensional data processing. Hence, the empirical mode decomposition (EMD) method has been developed to overcome this restriction. Monjoly et al. (2017) compared WT, EMD, and ensemble empirical mode decomposition (EEMD) in a study. The prediction accuracy of a single ML model can also be improved by combining intelligent algorithms and ML models. To forecast $S_{(r)}$, for instance, Eseye et al. (2018) merged the WT technique with a hybrid support vector machine model. Mostafavi et al. (2013) combined genetic programming (GP) and a simulated annealing (SA) algorithm to achieve accurate prediction of global $S_{(r)}$. Peng et al. (2021) developed a DL model for $S_{(r)}$ prediction based on a sine cosine algorithm (SCA), bi-directional long short-term memory (LSTM), and CEEMDAN, and compared with seven other models to demonstrate the superiority of their model.

1.4. Research objectives

The goal of the current research is to develop a reliable and robust forecasting approach by hybridizing different ML algorithms. It is worth mentioning that this is the first investigation using such a new approach to forecast the global solar exposure (GSE). Most of the previous studies focused on the forecasting/prediction of solar radiation. GSE is more representable for the solar energy harvesting component. The specific objectives of this study include:



T_{max} (t-1)	T_{max} (t-2)	T_{max} (t-3)	T_{max} (t-4)	T_{max} (t-5)
0.60161	0.1666	0.01072	0.01167	0.00884
T_{max} (t-6)	T_{max} (t-7)	T_{max} (t-8)	T_{max} (t-9)	T_{max} (t-10)
0.02103	0.00938	0.01427	0.01434	0.01232
T_{min} (t-1)	T_{min} (t-2)	T_{min} (t-3)	T_{min} (t-4)	T_{min} (t-5)
0.02193	0.02466	0.0192	0.01194	0.00958
T_{min} (t-6)	T_{min} (t-7)	T_{min} (t-8)	T_{min} (t-9)	T_{min} (t-10)
0.01345	0.00702	0.03047	0.09104	0.04986

T_{max} (t-1)	T_{max} (t-2)	T_{max} (t-3)	T_{max} (t-4)	T_{max} (t-5)
0.2044	0.1223	0.0071	0.0310	0.0099
T_{max} (t-6)	T_{max} (t-7)	T_{max} (t-8)	T_{max} (t-9)	T_{max} (t-10)
0.0098	0.0188	0.0401	0.0302	0.1224
T_{min} (t-1)	T_{min} (t-2)	T_{min} (t-3)	T_{min} (t-4)	T_{min} (t-5)
0.0132	0.3014	0.0056	0.0059	0.0061
T_{min} (t-6)	T_{min} (t-7)	T_{min} (t-8)	T_{min} (t-9)	T_{min} (t-10)
0.0071	0.0121	0.0058	0.0384	0.0085

Fig. 5. Outcomes of extra tree feature selection to specify the most important lagged-time components of each predictor (T_{max} and T_{min}) in two study cases including importance factors of each lag in form of two sun plots.

Table 2
Summarized pre-processing information for the two stations: Decomposition tuning via TVF-EMD and EWT applied in all the hybrid ML-based models; and optimal antecedent information (lagged-time).

Station	Tuning and regularization	Brisbane	Perth airport
TVF-EMD	Decomposition parameters	Max_IMF = 9; b-spline order:26; Bandwidth threshold:0.1; end_flag = 0;	Max_IMF = 9; b-spline order:26; Bandwidth threshold:0.1; end_flag = 0;
	IMFs (T_{max})	9	9
	IMFs (T_{min})	9	9
	Total number of IMFs	5Lags@(9 + 9) = 90	5Lags@(9 + 9) = 90
EWT	Decomposition parameters	Signal = 'sig2', Global trend removal= 'plaw'; Polynomial interpolation degree: 6	Signal = 'sig2'; Global trend removal= 'none'; maximum number of bands:3; Polynomial interpolation degree: 6
	IMFs (T_{max})	7	10
	IMFs (T_{min})	7	13
	Total number of IMFs	5Lags@(7 + 7) = 70	5Lags@(10 + 13) = 115
Significant antecedent information		Tmax(t-1), Tmax(t-2), Tmax(t-8), Tmax(t-9), Tmax(t-10), Tmin(t-1), Tmin(t-2), Tmin(t-8), Tmin(t-9), and Tmin(t-10)	Tmax(t-1), Tmax(t-2), Tmax(t-4), Tmax(t-8), Tmax(t-10), Tmin(t-1), Tmin(t-2), Tmin(t-4), Tmin(t-8), and Tmin(t-10)

- Developing a novel expert DL-based framework to accurately forecast weekly GSE based on the minimum and maximum air temperatures at the Perth airport and Brisbane stations in Australia.
- Coupling extra tree feature selection, two decomposition schemes (i.e., time-varying filtering based empirical mode decomposition (TVF-EMD) and empirical wavelet transform (EWT)), and ensemble deep random vector functional link (ED-RVFL).
- Estimating the significant antecedent information among existing lags of the two predictors using the extra tree feature selection.
- Using the TVF-EMD and EWT schemes to decompose the raw time series of the predictors imposed by the pre-estimated lags to feed four advanced ML approaches.
- Validating the main modeling frameworks (i.e., ED-RVFL-TVF-EMD and ED-RVFL-EWT) with RVFL, Bi-LSTM, and bagging in hybrid and standalone counterpart formats.

2. Methodology and materials

2.1. Empirical wavelet transform (EWT)

The empirical wavelet transform (EWT) (Gilles, 2013) is a segmentation technique for signals. To remove amplitude and frequency modulation (AM–FM) from the Fourier spectrum, appropriate orthogonal wavelet filters are created during the segmentation process. Then, the AM–FM signal is transformed using the Hilbert transform to get instantaneous frequency amplitude. This transformation is comparable to the construction of bandpass filters from the standpoint of the Fourier transform (Gu, Chang, Xiong, & Chen, 2021).

After determining the boundary frequencies ω_{BN} (where $n = 0,1,2,\ldots, N$; $\omega_{B0} = 0$, $\omega_{BN} = \pi$), it is possible to define the empirical scaling function and the empirical wavelets, as in the creation of Littlewood-Paley and Meyer's wavelets (Daubechies, 1992; M. Wang & Sheng,

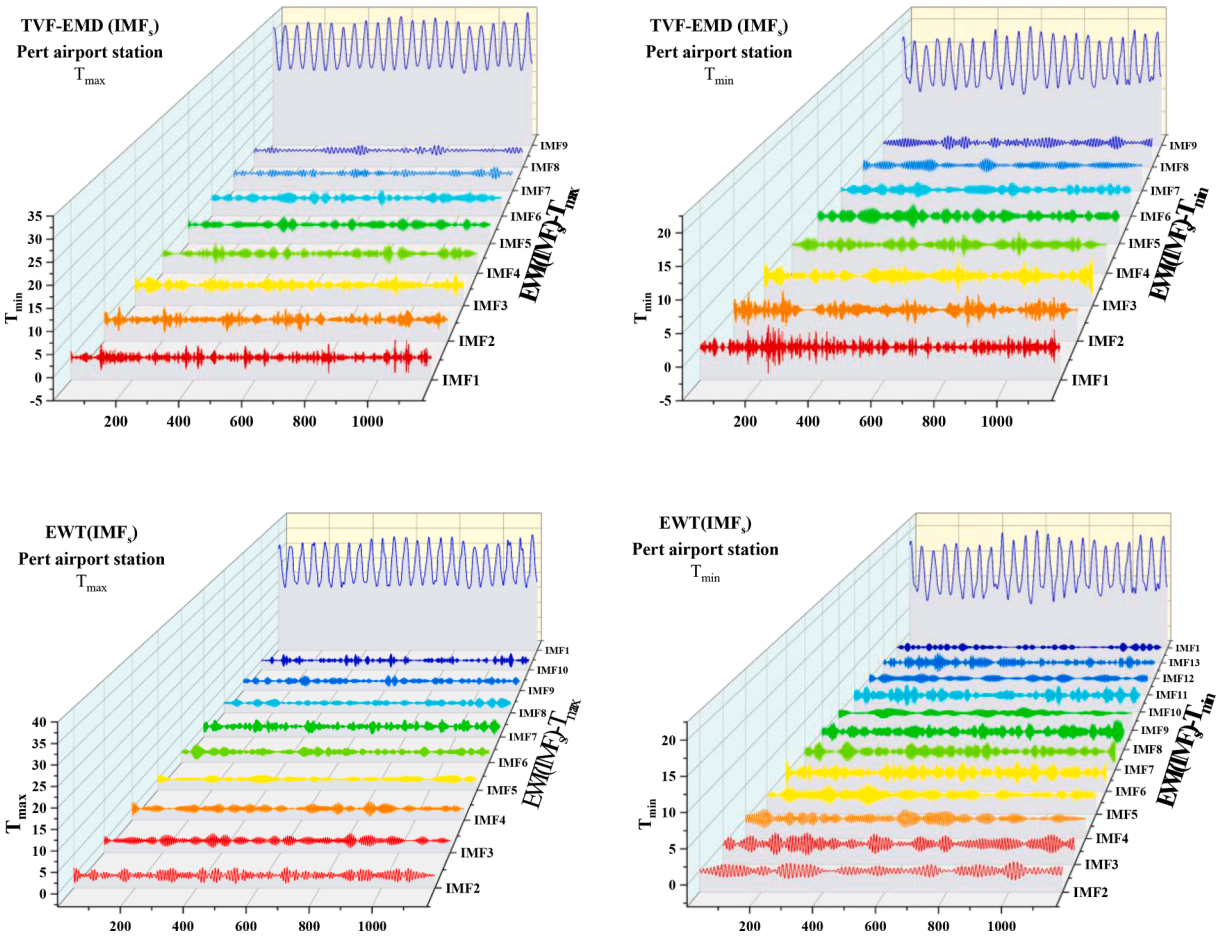


Fig. 6. Decomposed sub-sequences of predictors using TVF-EMD (upper panel) and EWT (lower panel) for Perth airport station.

2022).

$$\widehat{\phi}_n(\omega)=\begin{cases}1 & ,|\omega|\leqslant\omega_{Bn}-\tau_n\\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega|-\omega_{Bn}+\tau_n)\right)\right] & ,\omega_{Bn}-\tau_n\leqslant|\omega|\leqslant\omega_{Bn}+\tau_n\\ 0\end{cases} \tag{1}$$

$\min\left(\left(\omega_{B,n+1}-\omega_{Bn}\right)/\left(\omega_{B,n+1}+\omega_{Bn}\right)\right)$. The set $\left\{\varphi_1(t),\{\psi_n(t)\}_{n=1}^{N-1}\right\}$ is now known as a tight frame.

The wavelet approximation and the detailed coefficients are given by the following inner products (Gilles, 2013):

$$W_x^e(0,t)=\langle x,\phi_1\rangle=\int x(\tau)\phi_1^*(\tau-t)d\tau=F^{-1}[X(\omega)\widehat{\phi}_1(\omega)] \tag{3}$$

$$\widehat{\psi}_n(\omega)=\begin{cases}1 & ,\omega_{Bn}+\tau_n\leqslant|\omega|\leqslant\omega_{B,n+1}-\tau_{n+1n}\\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}\left(|\omega|-\omega_{B,n+1}+\tau_{n+1}\right)\right)\right] & ,\omega_{B,n+1}-\tau_{n+1}\leqslant|\omega|\leqslant\omega_{B,n+1}+\tau_{n+1}\\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega|-\omega_{Bn}+\tau_n)\right)\right] & ,\omega_{Bn}-\tau_n\leqslant|\omega|\leqslant\omega_{Bn}+\tau_n\\ 0\end{cases} \tag{2}$$

where $\beta(x)$ is a polynomial function having values in the range of $[0,1]$ and satisfies $\beta(x)+\beta(1-x)=1,\forall x\in[0,1]$. Many polynomial functions meet this criterion. $\beta(x)=x^4\left(35-84x+70x^2+20x^3\right)$ has been widely used for this purpose. $\tau_n=\gamma\omega_{Bn}$ defines the transient phase with a width of $2\tau_n$ around each ω_{Bn}. To ensure that there is no overlap between two successive transitions, the parameter must fulfill $0<\gamma<$

$$W_x^e(n,t)=\langle x,\psi_n\rangle=\int x(\tau)\psi_n^*(\tau-t)d\tau=F^{-1}[X(\omega)\widehat{\psi}_n(\omega)],\ (n=1,2,\ldots,N-1) \tag{4}$$

where * means "conjugate," $\widehat{\phi}_1(\omega)$ and $\widehat{\psi}_n(\omega)$ are defined in Equations

Table 3
Tuning parameters of each machine learning approach in hybrid and the corresponding standalone counterparts gained by the random search.

Hyperparameters used in the ML models		
Model	Perth airport	Brisbane
ED-R_{VFL}-T_{VF-EMD}	num_nodes: 5, regular_para = 0.05, num_layer = 3, weight_random_range = [-1, 1]	num_nodes: 5, regular_para = 0.0005, num_layer = 2, weight_random_range = [-1, 1]
ED-R_{VFL}-E_{WT}	num_nodes: 8, regular_para = 0.6, num_layer = 3, weight_random_range = [-1, 1]	num_nodes: 6, regular_para = 0.0005, num_layer = 2, weight_random_range = [-1, 1]
ED-RVFL	num_nodes: 200, regular_para = 0.00055, num_layer = 4, weight_random_range = [-1, 1]	num_nodes: 5, regular_para = 0.0005, num_layer = 2, weight_random_range = [-1, 1]
$B_{i\text{-}LSTM}$-T_{VF-EMD}	Neurons number: 40, Number of Layers: 2, activation='relu', Learning Rate: 0.0004, Epochs: 90, Training Algorithm: Adam, Batch Size: 256	Neurons number: 40, Number of Layers: 2, activation='relu', Learning Rate: 0.0005, Epochs: 90, Training Algorithm: Adam, Batch Size: 256
$B_{i\text{-}LSTM}$-E_{WT}	Neurons number: 60, Number of Layers: 2, activation='relu', Learning Rate: 0.0008, Epochs: 90, Training Algorithm: Adam, Batch Size: 256	Neurons number: 40, Number of Layers: 2, activation='relu', Learning Rate: 0.0004, Epochs: 80, Training Algorithm: Adam, Batch Size: 256
Bi-LSTM	Neurons number: 40, Number of Layers: 2, activation='relu', Learning Rate: 0.0004, Epochs: 90, Training Algorithm: Adam, Batch Size: 256	Neurons number: 40, Number of Layers: 2, activation='relu', Learning Rate: 0.0005, Epochs: 80, Training Algorithm: Adam, Batch Size: 256
R_{VFL}-T_{VF-EMD}	num_nodes: 60, regular_para = 0.05, bias_random_range = [0, 1], weight_random_range = [-1, 1]	num_nodes: 70, regular_para = 0.1, bias_random_range = [0, 1], weight_random_range = [-1, 1]
R_{VFL}-E_{WT}	num_nodes: 50, regular_para = 0.1, bias_random_range = [0, 1], weight_random_range = [-1, 1]	num_nodes: 60, regular_para = 0.1, bias_random_range = [0, 1], weight_random_range = [-1, 1]
RVFL	num_nodes: 40, regular_para = 0.1, bias_random_range = [0, 1], weight_random_range = [-1, 1]	num_nodes: 100, regular_para = 0.1, bias_random_range = [0, 1], weight_random_range = [-1, 1]
B_{agging}-T_{VF-EMD}	Function: "Fitrensemble", Mode: "bag", 'MaxNumSplits': 10, 'NumVariablesToSample': All	Function: "Fitrensemble", Mode: "bag", 'MaxNumSplits': 10, 'NumVariablesToSample': All
B_{agging}-E_{WT}	Function: "Fitrensemble", Mode: "bag", 'MaxNumSplits': 10, 'NumVariablesToSample': All	Function: "Fitrensemble", Mode: "bag", 'MaxNumSplits': 10, 'NumVariablesToSample': All
Bagging	Function: "Fitrensemble", Mode: "bag", 'MaxNumSplits': 10, 'NumVariablesToSample': All	Function: "Fitrensemble", Mode: "bag", 'MaxNumSplits': 10, 'NumVariablesToSample': All

(1) and (2), and $\phi_1(t)$ is the inverse Fourier transform of $\widehat{\phi}_1(\omega)$.

2.2. Time varying filter-based empirical mode decomposition (TVF-EMD)

To solve the mode mixing problem, a modified version of empirical mode decomposition (EMD) by applying a time-varying filter (TVF) to complete the shifting process was proposed by Li et al. (2017). For time-frequency analysis, the TVF-EMD has been demonstrated as an effective technique. The implementation process of the TVF-EMD includes the following (H. Li et al., 2017; Song, Chen, Wu, & Jin, 2021):

(i) Shifting process of the empirical mode decomposition.
(ii) Estimate the local cut-off frequency (LCOF) and the realignment, which is defined as:

$$g_m^n(t) = \sum_{k=-\infty}^{\infty} c(k)\beta^n(t/m - k) \tag{5}$$

where $\beta^n(t)$ and $c(k)$ are the B-spline function and coefficient. The approximation is determined by n, m, and $c(k)$ to minimize the approximation error (ε_m^2), which is defined as:

$$\varepsilon_m^2 = \sum_{t=-\infty}^{+\infty} \left(x(t) - [c]_{\uparrow m} * b_m^n(t)\right)^2 \tag{6}$$

in which $b_m^n(t)$ is the $\beta^n(t/m)$, $[\bullet]_{\uparrow m}$ is the up-sampling operation by m, and * is the convolution operator. $c(k)$ is given by:

$$c(k) = \left[p_m^n * x\right]_{\downarrow m}(k) \tag{7}$$

where, $[\bullet]_{\downarrow m}$ is the down-sampling operation by m, and p_m^n is the pre-filter. Thus, Eq. (1) can be rewritten as:

$$g_m^n(t) = \left[p_m^n * x\right]_{\downarrow m} * b_m^n(t) \tag{8}$$

In Eq. (4), the B-spline approximation is a special form of low pass filtering. Thus, the LCOF is expected from the input signal and used to create the TVF. This process is carried out to obtain the LCOF, $\varphi_{bis}^{'}(t) = \varphi_1^{'}(t) + \varphi_2^{'}(t)/2$ (where, $\varphi_1^{'}(t)$, and $\varphi_2^{'}(t)$ = slow varying components). Realign the $\varphi_{bis}^{'}(t)$ to solve the issue of intermittences and attain the final LCOF by interpolating among the peaks.

(iii) Shifting process based on TVF, which filters the input signal into local higher and lower frequency components and obtains the extracted local narrowband signal. Check the stopping criterion for such a shifting process using Eq. (5) (K. Wang et al., 2020):

$$\theta(t) = \frac{B_{Loughlin}(t)}{\varphi_{avg}(t)} \tag{9}$$

where, $\varphi_{avg}(t)$ and $B_{Loughlin}(t)$ are the weighted average instantaneous frequency and Loughlin instantaneous bandwidth of separate components.

2.3. Extra tree feature selection

Feature selection is frequently required prior to the use of a learning algorithm. In the current study, the extra tree ensemble algorithm is used to determine the importance of input data. ET is an ensemble learning strategy that averages the predictions of decision trees (DTs) to boost accuracy and decrease computational complexity (Geurts, Ernst, & Wehenkel, 2006). The ET method generates a random ensemble of trees, then aggregates their predictions using a suitable method, such as mathematical averaging for the case of regression or voting majority for classification tasks (Seyyedattar, Ghiasi, Zendehboudi, & Butt, 2020). Unlike the traditional tree-based ensemble techniques, ETs construct trees using the entire learning sample rather than a bootstrap replica, and node splitting is performed by a totally random selection of cut points (Wehenkel, Ernst, & Geurts, 2006). The two basic parameters of the ETR are K (the number of features randomly nominated at each node) and n_{min} (the minimum sample size for splitting a node), which prevent overfitting and improve model prediction accuracy. K represents the number of randomly picked characteristics at each node throughout the process of tree growth, and it defines the intensity of the variable selection process. ETs are expected to successfully minimize the variance by the combination of explicit randomization of input features

Table 4
Goodness-of-fit metrics for assessing the robustness of the hybrid and standalone models to forecast the weekly GSE at the Perth airport station.

Models	Phase	R	RMSE	MAE	MAPE	NSE	KGE	I_A	$U_{95\%}$
ED-R_{VFL}-$T_{VF\text{-}EMD}$	Training	0.9682	1.8838	1.4526	8.6470	0.9374	0.9521	0.9835	5.2232
	Testing	0.9665	1.9193	1.4808	8.5879	0.9288	0.9565	0.9823	5.2803
ED-R_{VFL}-E_{WT}	Training	0.9623	2.0501	1.6085	9.5965	0.9259	0.9403	0.9802	5.6843
	Testing	0.9634	1.9832	1.5345	9.0832	0.9240	0.9583	0.9809	5.4685
ED-RVFL	Training	0.9283	2.8034	2.1939	12.9547	0.8614	0.8842	0.9607	7.7728
	Testing	0.8957	3.2408	2.4869	14.5794	0.7971	0.8884	0.9453	8.9815
$B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$	Training	0.9428	2.5127	1.9628	11.4506	0.8887	0.9099	0.9693	6.9652
	Testing	0.9490	2.3270	1.7433	9.7423	0.8954	0.9459	0.9736	6.4367
$B_{i\text{-}LSTM}$-E_{WT}	Training	0.9856	1.2720	0.8868	5.0420	0.9715	0.9797	0.9927	3.5267
	Testing	0.9413	2.6933	2.0390	11.9120	0.8599	0.8901	0.9672	7.4150
Bi-LSTM	Training	0.9053	3.2087	2.4754	14.4944	0.8184	0.8604	0.9476	8.8837
	Testing	0.9162	2.8940	2.1842	12.4839	0.8382	0.9030	0.9562	8.0286
R_{VFL}-$T_{VF\text{-}EMD}$	Training	0.9713	1.7915	1.3983	8.3708	0.9434	0.9572	0.9852	4.9672
	Testing	0.9608	2.0394	1.5360	8.7936	0.9204	0.9484	0.9804	5.5988
R_{VFL}-E_{WT}	Training	0.9659	1.9501	1.5371	9.2302	0.9329	0.9500	0.9823	5.4068
	Testing	0.9606	2.0718	1.5957	9.5140	0.9171	0.9514	0.9794	5.7209
RVFL	Training	0.8858	3.4949	2.8074	17.1343	0.7846	0.8383	0.9362	9.6903
	Testing	0.8729	3.5711	2.8427	17.0501	0.7537	0.8639	0.9321	9.8900
B_{agging}-$T_{VF\text{-}EMD}$	Training	0.9864	1.2796	0.9671	5.6477	0.9711	0.9410	0.9923	3.5479
	Testing	0.9468	2.3217	1.6924	9.8705	0.8959	0.9383	0.9726	6.4390
B_{agging}-E_{WT}	Training	0.9856	1.3187	1.0043	5.8938	0.9693	0.9387	0.9918	3.6564
	Testing	0.9408	2.4422	1.8213	10.3860	0.8848	0.9245	0.9692	6.7718
Bagging	Training	0.9627	2.1196	1.6079	9.4698	0.9208	0.8799	0.9774	5.8762
	Testing	0.9028	3.1000	2.3004	13.2733	0.8144	0.8612	0.9470	8.5932

and cut points with ensemble averaging, and the use of the whole original learning sample aids in the reduction of bias. The number of trees created by the ET model determines the variance reduction of the ensemble model aggregation (Geurts et al., 2006). The ETR model was created in this study by using the Python's scikit-library.

2.4. Bidirectional- LSTM (Bi-LSTM)

Graves and Schmidhuber (2005) proposed the concept of the bidirectional long short-term memory (Bi-LSTM) to solve the gradient explosion and information morphing drawbacks during the backpropagation, a particular type of recurrent neural network (Schuster & Paliwal, 1997). The Bi-LSTM model processes the sequential data into bidirectional (i.e., forward (past to future) and backward (future to past) directions) with two distinct hidden layers (or LSTM layers), which are connected to the same output layer (Ihianle et al., 2020). In recent years, DL-based models (e.g., Bi-LSTM) have been applied in different areas (Apaydin et al., 2020; Dey, Dey, & Mall, 2021; Kang, Yang, Huang, & Oh, 2020; Le, Nguyen, Jung, Yeon, & Lee, 2021; LI, MA, CHEN, & HUANG, 2021; Liang et al., 2021; Maddu, Vanga, Sajja, Basha, & Shaik, 2021; Park, Jung, Kim, & Park, 2020; Yin, Deng, Ines, Wu, & Rasu, 2020). The upper panel of Fig. 1 shows the structure of the Bi-LSTM network with a forward hidden state $\left(\overrightarrow{h_t}\right)$ and backward hidden state $\left(\overleftarrow{h_t}\right)$, and updated based on input layer (x_t). Both hidden states are calculated for forward and backward directions. The Bi-LSTM output is obtained by integrating the outputs of the forward and backward layers using Eq. (10) (Y.-H. Li, Harfiya, Purwandari, & Lin, 2020; Minh-Tuan & Kim, 2019):

$$y_t = \sigma\left(\overrightarrow{h_t}, \overleftarrow{h_t}\right) \tag{10}$$

in which, σ = a function used to combine the two output sequences. It can be a summation, average, concatenating, or multiplication function. The final output of the Bi-LSTM layer is characterized by a vector, $Y_T = [y_{T-n}, \cdots, y_{T-1}]$, in which the least element, y_{T-1}, is the forecasted GSE for the next iteration.

2.5. Bagging tree regression

Breiman (1996) proposed the concept of the bagging tree regression (BTR or bootstrap aggregation), and prediction is made based on ensemble trees (Y. Zhang, Chiew, Li, & Post, 2018). It is an effective method, compared to other ensemble-based decision tree algorithms like boosting (C. Zhang & Ma, 2012). The lower panel of Fig. 1 shows the prediction procedure of the bagging algorithm, which can be used in many classifications and regression to reduce the variance and thereby enhance the robustness of the prediction process (Basilio & Goliatt, 2022; Sutton, 2005). The practical implementation of the bagging algorithm includes the following steps (Avand et al., 2020; Zounemat-Kermani, Batelaan, Fadaee, & Hinkelmann, 2021):

Table 5
Goodness-of-fit metrics for assessing the robustness of the hybrid and standalone models to forecast the weekly GSE at the Brisbane station.

Models	Mode	R	RMSE	MAE	MAPE	NSE	KGE	I_A	$U_{95\%}$
ED-R_{VFL}-$T_{VF\text{-}EMD}$	Training	0.9151	1.9361	1.4958	8.0260	0.8374	0.8788	0.9538	5.3683
	Testing	0.9181	2.0313	1.5912	8.5692	0.8381	0.8601	0.9531	5.6017
ED-R_{VFL}-E_{WT}	Training	0.9204	1.8771	1.4465	7.7212	0.8472	0.8870	0.9570	5.2044
	Testing	0.9218	1.9708	1.5230	8.1847	0.8476	0.8552	0.9555	5.4655
ED-RVFL	Training	0.8284	2.6894	2.1487	11.7613	0.6863	0.7555	0.8987	7.4568
	Testing	0.8473	2.7043	2.0735	11.4236	0.7131	0.7453	0.9061	7.4873
$B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$	Training	0.8662	2.4034	1.8983	10.1830	0.7495	0.7985	0.9223	6.6602
	Testing	0.8928	2.3026	1.7430	9.4168	0.7920	0.8054	0.9359	6.3716
$B_{i\text{-}LSTM}$-E_{WT}	Training	0.8771	2.3061	1.8207	9.8176	0.7694	0.8292	0.9312	6.3940
	Testing	0.8780	2.4241	1.8934	10.0611	0.7694	0.8104	0.9298	6.7185
Bi-LSTM	Training	0.8422	2.5911	2.0369	11.0130	0.7088	0.7847	0.9095	7.1825
	Testing	0.8579	2.5994	1.9513	10.5992	0.7349	0.7865	0.9177	7.2065
R_{VFL}-$T_{VF\text{-}EMD}$	Training	0.9209	1.8720	1.4667	7.9018	0.8480	0.8850	0.9571	5.1905
	Testing	0.9077	2.1448	1.7065	9.3616	0.8195	0.8570	0.9478	5.9158
R_{VFL}-E_{WT}	Training	0.9276	1.7937	1.3987	7.4910	0.8605	0.8969	0.9611	4.9734
	Testing	0.9075	2.1252	1.6719	8.9160	0.8228	0.8512	0.9480	5.8947
RVFL	Training	0.7409	3.2854	2.6658	14.7741	0.5319	0.7107	0.8528	9.1092
	Testing	0.8383	2.7579	2.0882	11.3463	0.7016	0.7743	0.9063	7.6435
B_{agging}-$T_{VF\text{-}EMD}$	Training	0.9735	1.1577	0.8817	4.7705	0.9419	0.8939	0.9837	3.2100
	Testing	0.8832	2.3936	1.8106	9.7111	0.7752	0.7805	0.9287	6.6405
B_{agging}-E_{WT}	Training	0.9727	1.1736	0.8734	4.7175	0.9403	0.8925	0.9832	3.2541
	Testing	0.8960	2.2817	1.7343	9.0892	0.7957	0.7948	0.9359	6.3165
Bagging	Training	0.9488	1.5801	1.2242	6.6456	0.8917	0.8477	0.9679	4.3811
	Testing	0.8531	2.6622	2.0436	11.0874	0.7219	0.7419	0.9084	7.3746

(i) Build random subsets from the training dataset using the bootstrap sampling approach with replacement.
(ii) Apply some prediction models to each bootstrap sample.
(iii) Obtain the overall prediction by combining the results using averaging for regression and voting for classification.

2.6. Random vector functional link

RVFL is a feed-forward neural network with a functional link between the input and output layers (Pao, Park, & Sobajic, 1994). The initial random improvement layer transforms the input data into enhanced features. The concatenation of the improved features and input data serves as the input for the output layer. Instead of repeated training, the output weights are trained using linear regression methods (Adnan et al., 2021). This connection prevents the typical issue of overfitting in the traditional SLFNN. Fig. 2 (upper panel) depicts the architecture of the RVFL network.

RVFL collects N_s sample data (X), denoted by the pair (x_i, y_i), where y_i is the desired variable. The inputs are then processed by the mid (hidden) nodes. These mid nodes are known as enhancement nodes. The output of each mid node is calculated as follows (Zayed et al., 2021):

$$H = O_j\left(a_j x_i + b_j\right) = \frac{1}{1 + e^{\left(a_j x_i + b_j\right)}}, b_j \in [0, S], a_j \in [-S, S] \tag{11}$$

where a_j represents the weights between the input layer nodes and the middle (enhancement) layer nodes; and b_j is the bias. Scale factor S is examined during the optimization process. The following formula computes the final result of RVFL:

$$Z = Bw \tag{12}$$

in which, B comprises input data (B1) and enhancement node output (B2), and w is the weight of the output.

$$B_1 = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nn} \end{bmatrix}, B_2 = \begin{bmatrix} O(a_1 x_1 + b_1) & \cdots & O(a_p x_1 + b_p) \\ \vdots & \ddots & \vdots \\ O(a_1 x_1 + b_1) & \cdots & O(a_p x_N + b_p) \end{bmatrix} \tag{13}$$

The output weight is updated using the following equations, which represent the Moore–Penrose pseudo-inverse and ridge regression, respectively:

$$w = B^{\dagger} Z \tag{14}$$

$$w = \left(B^T B + \frac{1}{C}\right)^{-1} B^T Z \tag{15}$$

where † stands for the Moore-Penrose pseudo-inverse and $\frac{1}{C} = \lambda$ is the regulation parameter.

2.7. Ensemble Deep-RVFL (ED-RVFL)

Shi et al. (2021) developed the ED-RVFL and improved the original RVFL by building further hidden layers after the initial hidden layer. ED-

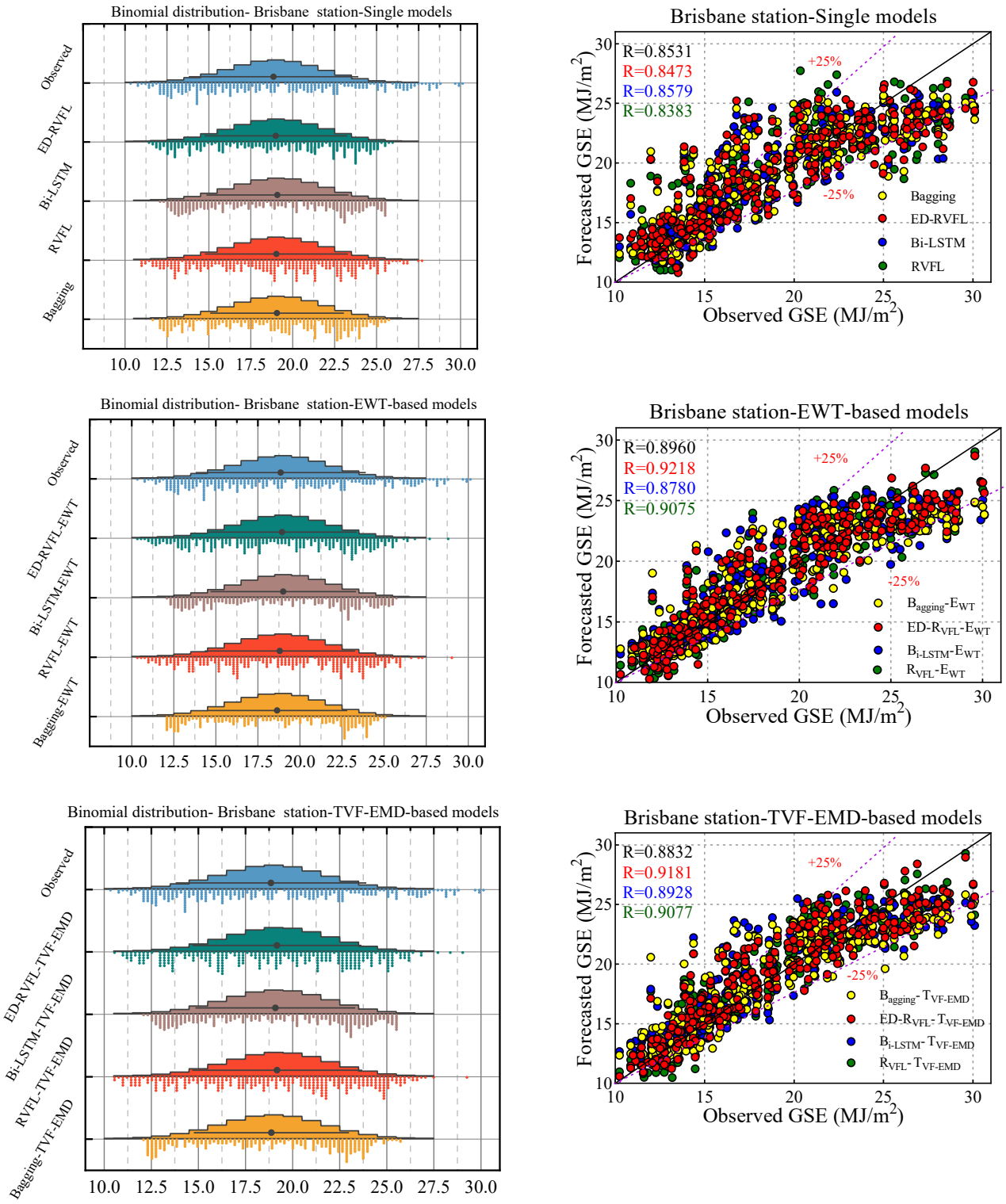


Fig. 7. Comparative assessment of the model efficiency based on the binomial distribution function (left) and scatter plots (right) in three categories of TVF-EMD-based models, EWT-based models, and their counterparts for the Brisbane and Perth airport stations in the testing period of GSE modeling.

RVFL randomizes and fixes the weights and bias between the hidden layers in the same manner as the original RVFL algorithm. Each output layer takes features from the output of its associated hidden layer, direct connections from the preceding hidden layer (if any), and the input layer. The architecture of the RVFL is shown in Fig. 2 (lower panel). In the first layer ($l = 1$), the output can be calculated by (Shi et al., 2021):

$$H^{(1)} = f\left(Xw^{(1)}\right) \tag{16}$$

where w is the weight connecting the input layer to the first hidden layer; and $f(.)$ is a non-linear activation function.

In subsequent layers ($l > 1$), direct linkages from the initial input X are added. The input feature space of this layer is the concatenation of

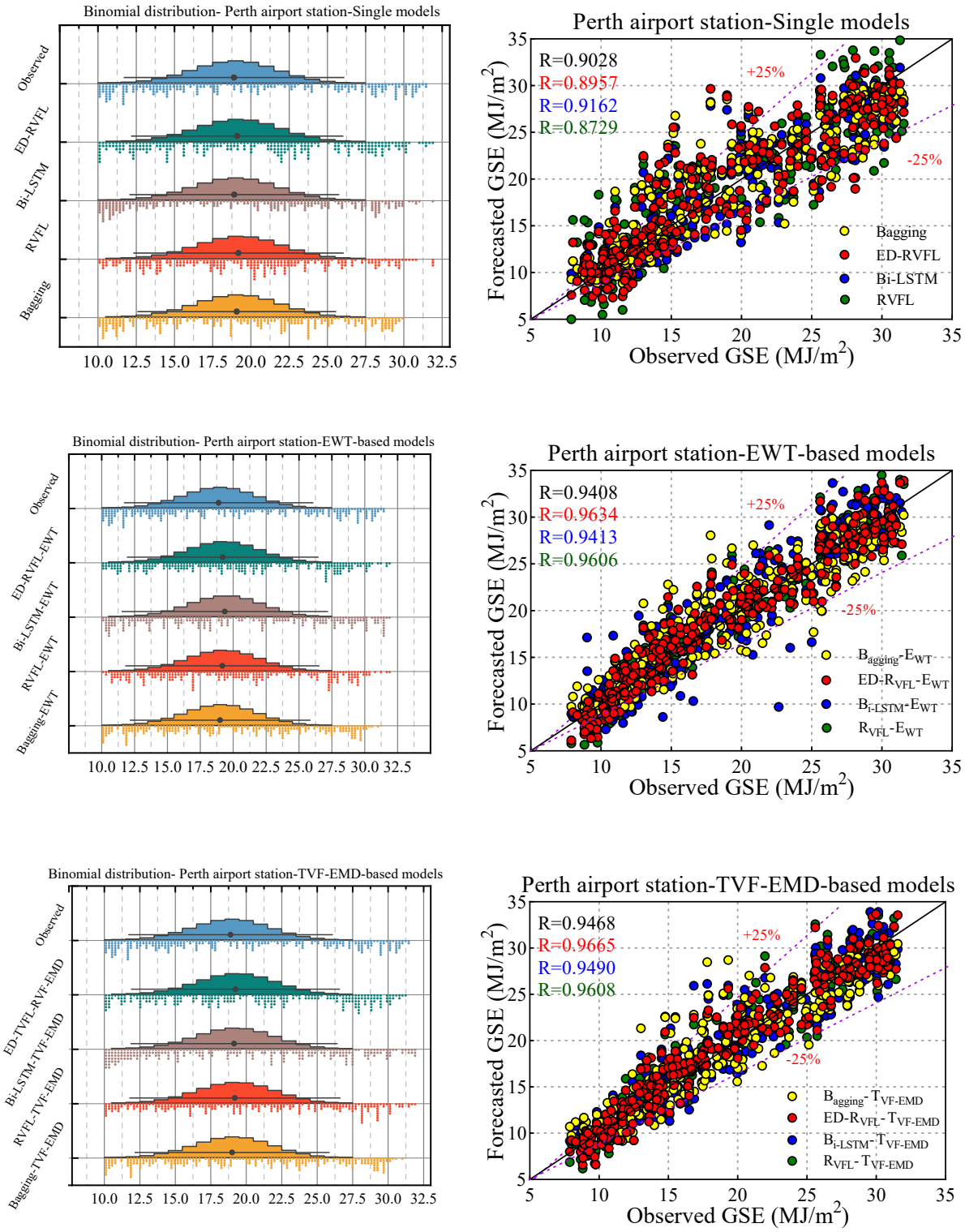


Fig. 7. (*continued*).

the output features of the previous layer ($H^{(l-1)}$) and the original input features (X):

$$H^{(l)} = f\left(\left[H^{(l-1)}X\right]w^{(l)}\right) \tag{17}$$

where $f(.)$ is a nonlinear activation function, and w represents the weight between the previous and current layers. L softmax layers act as classifiers for the ensemble in an ED-RVFL with L layers. The input feature space of the first softmax layer ($D_e^{(1)}$) has a direct connection to the original input features (X), as shown in the following equation (Cheng, Suganthan, & Katuwal, 2021):

$$D_e^{(1)} = \left[H^{(1)}X\right] \tag{18}$$

The input feature space of the next softmax layer ($D_e^{(1)}$) has direct linkages between the original input features (X) and the previous layer, as shown in the following equation:

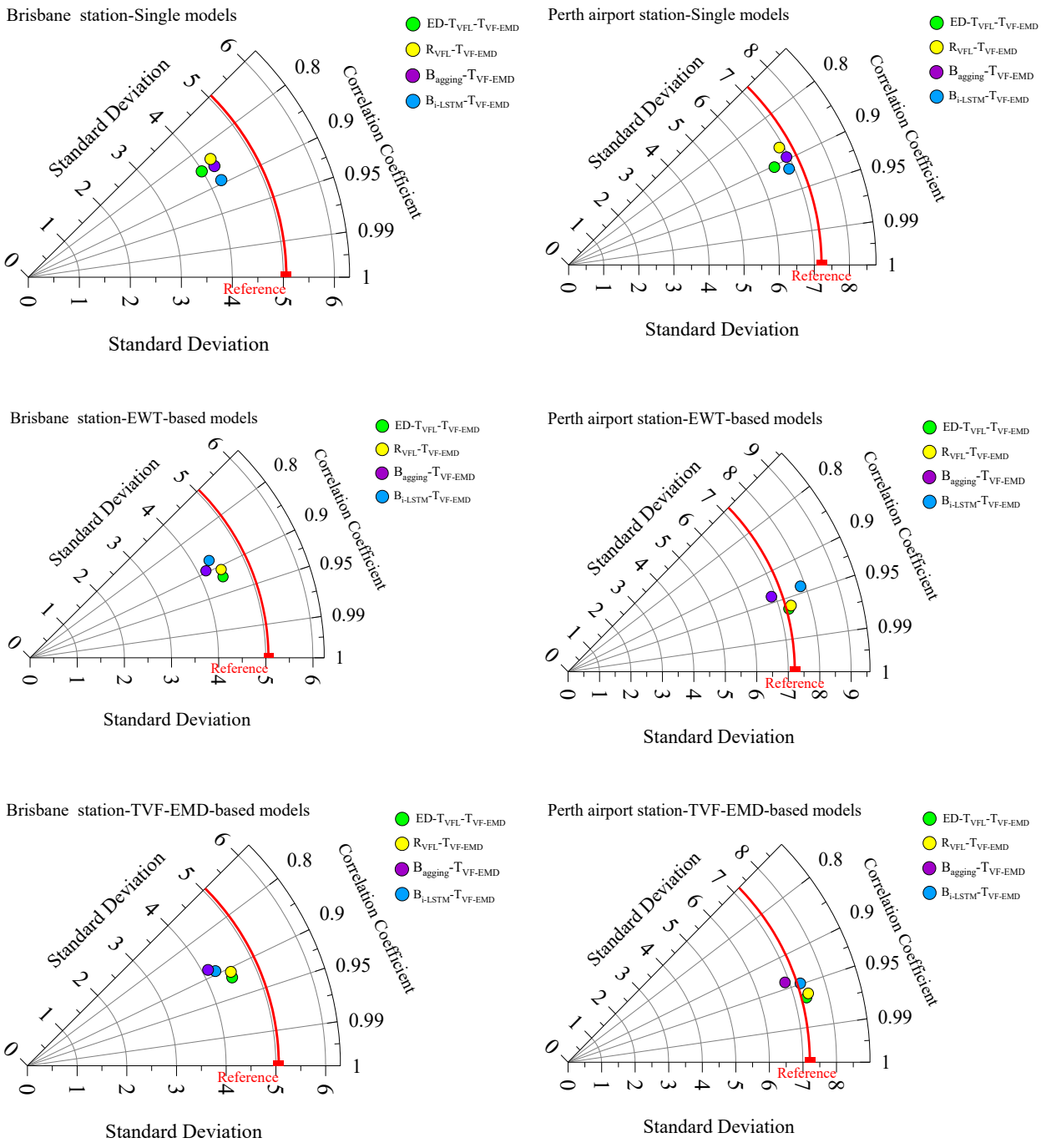


Fig. 8. Taylor diagrams of SGE forecasted by the hybrid models for the testing period at Brisbane station (left) and Perth airport station (right).

$$D_e^{(1)} = \left[H^{(l)} H^{(l-1)} X\right] \tag{19}$$

All weights between the input and hidden layers are randomized and held constant during the training process. The only parameters requiring optimization are the output weights. In this situation, each output layer can be viewed as a distinct classifier, and each output weight can be calculated independently. Using majority voting or average scoring, one can determine the final output label from an ED-RVFL ensemble, which is a collection of separate output layers. In this study, as a regression problem, averaging of the results is used (Cheng et al., 2021).

2.8. Evaluation indicators

The model performances are evaluated by using eight statistical metrics, including coefficient of correlation (R), root mean square error (RMSE), Kling-Gupta efficiency (KGE), index of agreement (I_A), uncertainty coefficient with 95% confidence level ($U_{95\%}$), mean absolute percent error (MAPE), Nash-Sutcliffe efficiency (NSE), and mean absolute error (MAE). The models used for forecasting weekly GSE at the Brisbane & Pert sites include ED-R$_{VFL}$-T$_{VF-EMD}$, ED-R$_{VFL}$-E$_{WT}$, ED-RVFL, B$_{i-LSTM}$-T$_{VF-EMD}$, B$_{i-LSTM}$-E$_{WT}$, Bi-LSTM, R$_{VFL}$-T$_{VF-EMD}$, R$_{VFL}$-E$_{WT}$, RVFL, B$_{agging}$-T$_{VF-EMD}$, B$_{agging}$-E$_{WT}$, and Bagging. Mathematically, R (Moriasi et al., 2015), RMSE (Jamei et al., 2020a), KGE (Gupta, Kling, Yilmaz, & Martinez, 2009; Jamei et al., 2022b), I_A (Willmott, 1981), $U_{95\%}$ (Patino & Ferreira, 2015; Rehamnia, Benlaoukli, Jamei, Karbasi, & Malik, 2021), MAPE (Singh, Jamei, Karbasi, Malik, & Pandey, 2022), NSE (Nash & Sutcliffe, 1970), and MAE can be respectively written as:

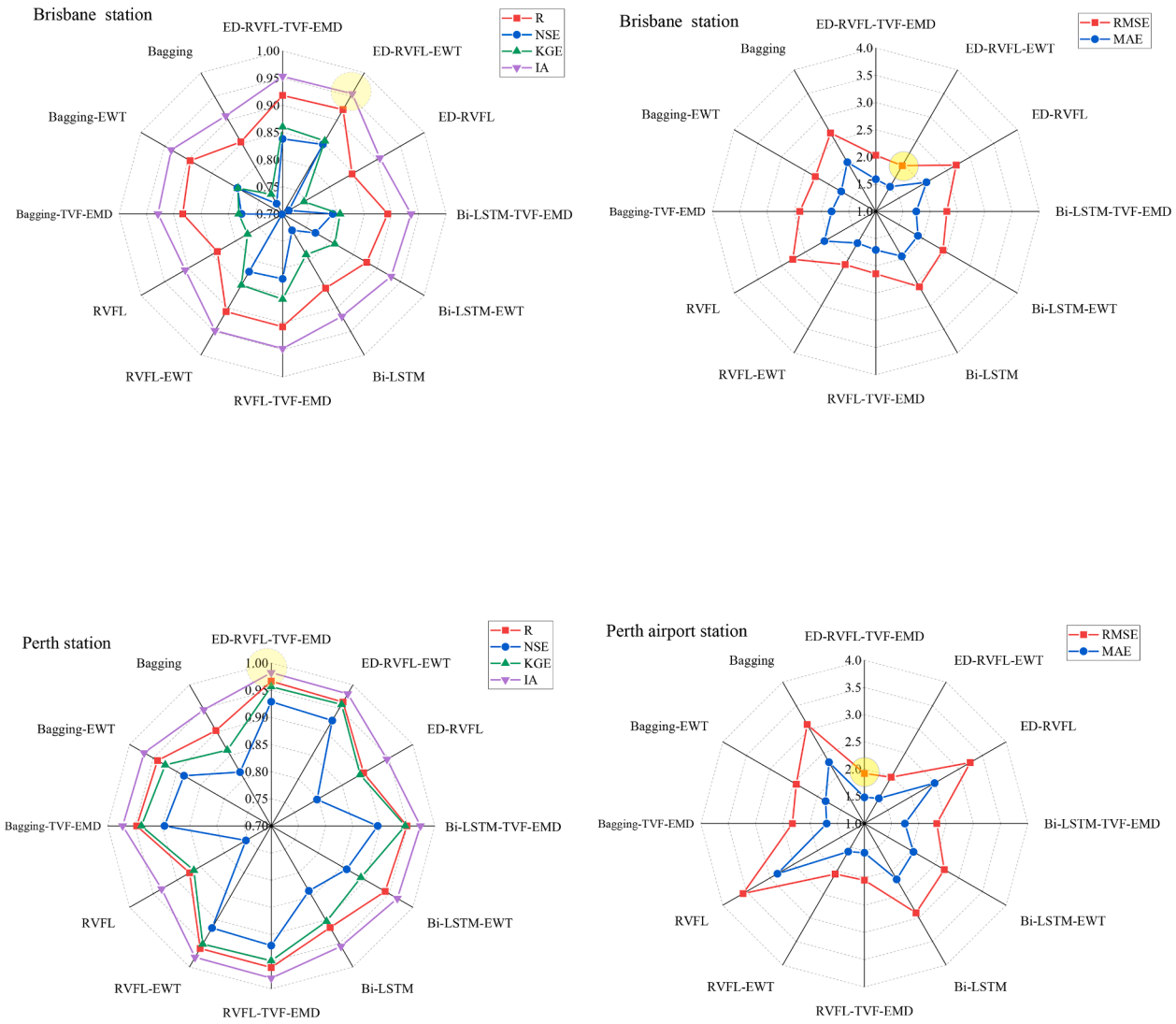


Fig. 9. Spider plots of the statistical metrics for different models for Brisbane station (upper panel) and Perth airport station (lower panel).

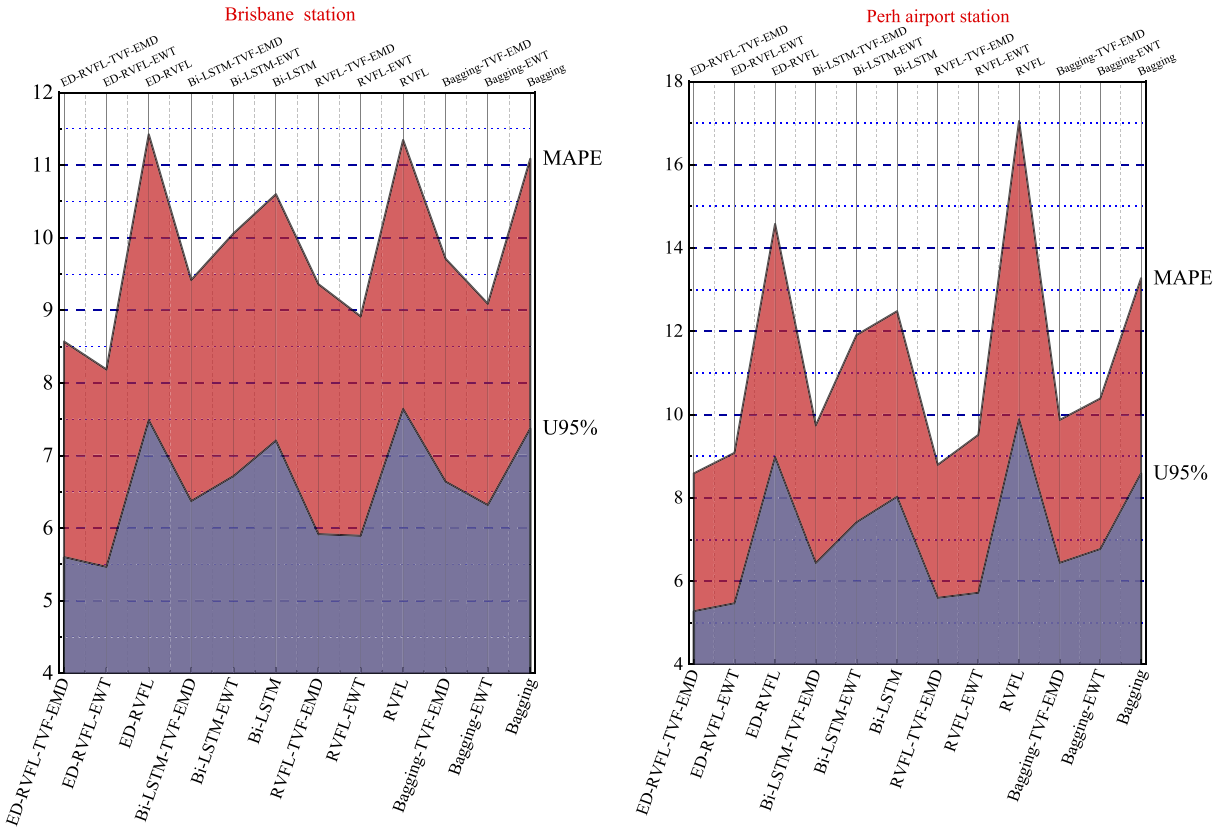


Fig. 10. Uncertainty and MAPE variation for the GSE forecasted by different models for Brisbane station (upper panel) and Perth airport station (lower panel).

$$R = \frac{\sum_{i=1}^{N}\left(GSE_{obs,i} - \overline{GSE_{obs}}\right)\left(GSE_{for,i} - \overline{GSE_{for}}\right)}{\sqrt{\sum_{i=1}^{N}\left(GSE_{obs,i} - \overline{GSE_{obs}}\right)^2 \sum_{i=1}^{N}\left(GSE_{for,i} - \overline{GSE_{for}}\right)^2}} \tag{20}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(GSE_{obs,i} - GSE_{for,i}\right)^2} \tag{21}$$

$$KGE = 1 - \sqrt{(R-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{22}$$

$$I_A = 1 - \left[\frac{\sum_{i=1}^{N}\left(GSE_{for,i} - GSE_{obs,i}\right)^2}{\sum_{i=1}^{N}\left(\left|GSE_{for,i} - \overline{GSE_{obs}}\right| + \left|GSE_{obs,i} - \overline{GSE_{obs}}\right|\right)^2}\right] \tag{23}$$

$$U_{95\%} = 1.96\sqrt{STD_e{}^2 + RMSE^2} \tag{24}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{GSE_{obs,i} - GSE_{for,i}}{GSE_{obs,i}}\right| \tag{25}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{N}\left(GSE_{obs,i} - GSE_{for,i}\right)^2}{\sum_{i=1}^{N}\left(GSE_{obs,i} - \overline{GSE_{obs}}\right)^2}\right] \tag{26}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|GSE_{for,i} - GSE_{obs,i}\right| \tag{27}$$

where, $GSE_{obs,i}$ and $GSE_{for,i}$ are the i^{th} observed and forecasted GSE values; $\overline{GSE_{obs}}$ and $\overline{GSE_{for}}$ are the averages of the observed and forecasted GSE values; α is the relative variability in the forecasted and observed GSE values; β is the ratio of the forecasted GSE average to the observed GSE average; STD_e denotes the standard deviation of error, and N is the total number of data points.

2.9. Study area and datasets

The GSE values range from 1 to 35 MJm^{-2}, which are normally highest in summer when the sky is clear, while smallest in winter and cloudy situations. The GSE data acquired in this research for the selected stations are satellite derived, imagery processed data from the Bureau of Meteorology, Australia. The original data were generated from the Geostationary Meteorological Satellites namely GMS-4, GMS-5, and MTSAT-1R, MTSAT-2 of the Japan Meteorological Agency, and the U.S. National Oceanic & Atmospheric Administration (NOAA) GOES-9 satellite.

The Perth and Brisbane selected in this study are two important cities located in the western and eastern ends of Australia, which respectively are the capitals of the States of Queensland and Western Australia. The Brisbane station is located on the Northern Brisbane River in Queensland and the average annual temperature ranges from 21.0 °C in July to 27.0 °C in February. The Perth airport station is located at the airport of Perth in Western Australia and the average annual temperature ranges from 18.9 °C in July to 23.4 °C in February.

All the datasets from 01/01/2000 to 30/06/2022 (22.5 years) for Brisbane and Perth stations were attained from the Bureau of Meteorology Australia. The acquired input data or predictors included maximum air temperature (T_{max}) and minimum air temperature (T_{min}) used to forecast weekly GSE. The descriptive statistics of the datasets and the latitude and longitude data for both stations are listed in Table 1. Fig. 3 illustrates the study area and the locations of the two stations.

At Brisbane station, the distributions of the observed weekly T_{min}, T_{max}, and GSE datasets were close to normal distributions with a Kurtosis range of [-1.062, −0.8635] and had a platykurtic distribution with respect to Kurtosis <0. At Perth airport station, the weekly T_{min}, T_{max}, and GSE datasets with a Kurtosis range of [−1.278, −0.863] were close to near-normal distributions and had a platykurtic distribution (Kurtosis <0). Besides, the skewness ranges (Brisbane: [−0.1651, 0.2257] and

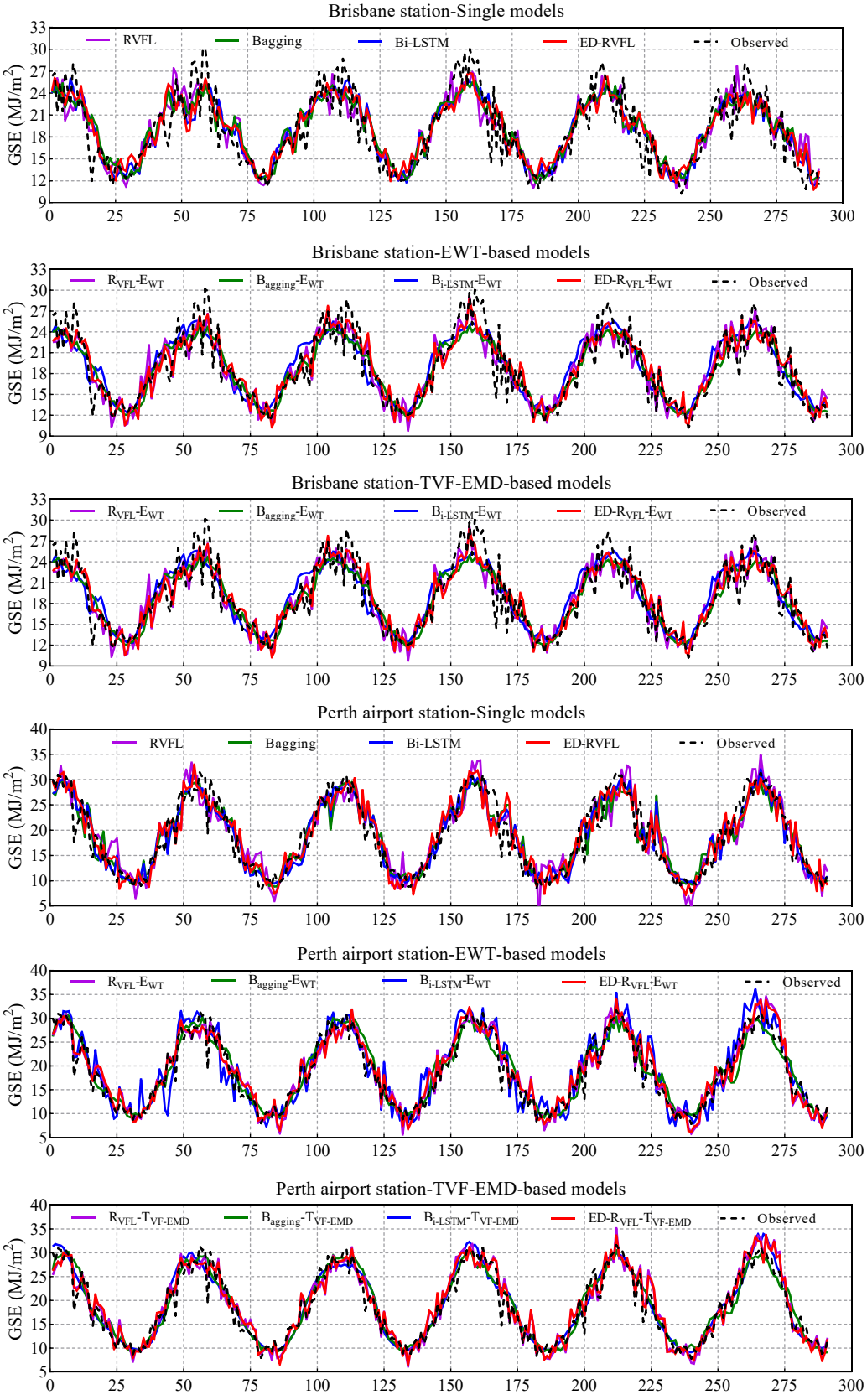
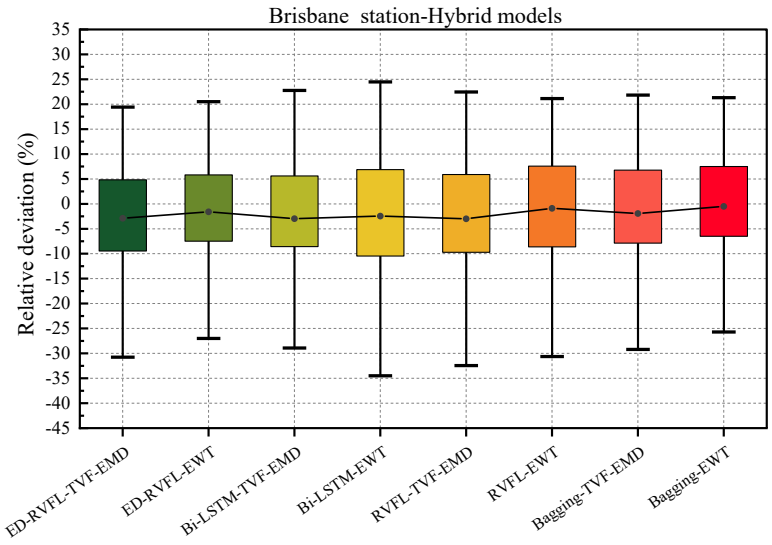
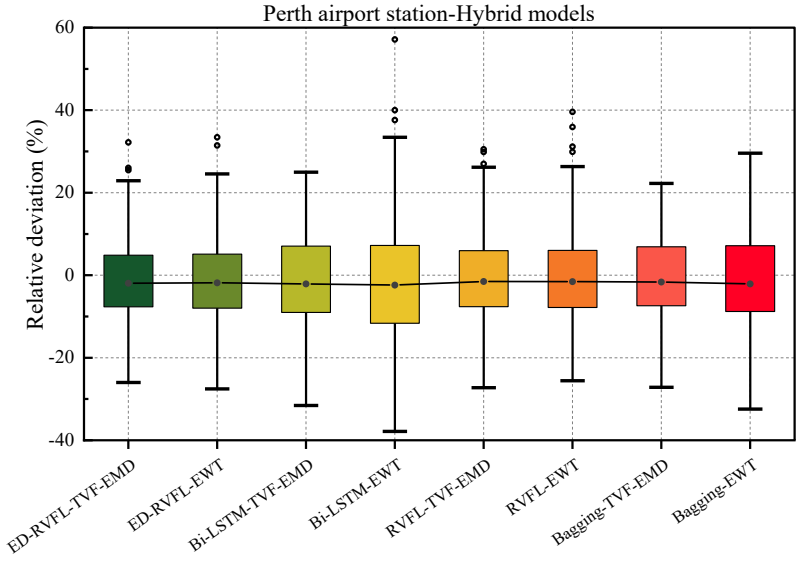


Fig. 11. Trends of the simulated GSE values during the testing period at Brisbane and Perth airport stations.



	ED-R_{VFL}-T_{VF-EMD}	*ED-R_{VFL}-E_{WT}*	*B_{i-LSTM}-R_{VF-EMD}*	*B_{i-LSTM}-E_{WT}*	*R_{VFL}-T_{VF-EMD}*	*R_{VFL}-E_{WT}*	*B_{agging}-T_{VF-EMD}*	*B_{agging}-E_{WT}*
Q25%	*-9.436*	*-7.471*	*-8.555*	*-10.44*	*-9.704*	*-8.618*	*-7.869*	*-6.478*
Median	*-1.329*	*-0.3819*	*-1.88*	*-1.25*	*-2.076*	*0.431*	*0.5351*	*1.281*
Q75%	*4.843*	*5.828*	*5.621*	*6.894*	*5.885*	*7.587*	*6.779*	*7.511*
Mean	*-2.883*	*-1.595*	*-2.962*	*-2.453*	*-2.989*	*-0.894*	*-1.941*	*-0.4952*
IQR	*14.279*	*13.299*	*14.176*	*17.334*	*15.589*	*16.205*	*14.648*	*13.989*



	ED-R_{VFL}-T_{VF-EMD}	ED-R_{VFL}-E_{WT}	B_{i-LSTM}-T_{VF-EMD}	B_{i-LSTM}-E_{WT}	R_{VFL}-T_{VF-EMD}	R_{VFL}-E_{WT}	B_{agging}-T_{VF-EMD}	B_{agging}-E_{WT}
Q25%	*-7.644*	*-7.952*	*-8.983*	*-11.61*	*-7.593*	*-7.81*	*-7.369*	*-8.777*
Median	*-1.921*	*-0.6553*	*0.03347*	*-1.28*	*-0.6392*	*-1.514*	*0.9014*	*0.1506*
Q75%	*4.879*	*5.134*	*7.075*	*7.226*	*5.975*	*6.021*	*6.897*	*7.154*
Mean	*-1.922*	*-1.851*	*-2.094*	*-2.392*	*-1.506*	*-1.533*	*-1.633*	*-2.099*
IQR	*12.523*	*13.086*	*16.058*	*18.836*	*13.568*	*13.831*	*14.266*	*15.931*

Fig. 12. Box plots of relative deviation distribution simulated by the hybrid models for Brisbane and Perth airport stations during the testing period.

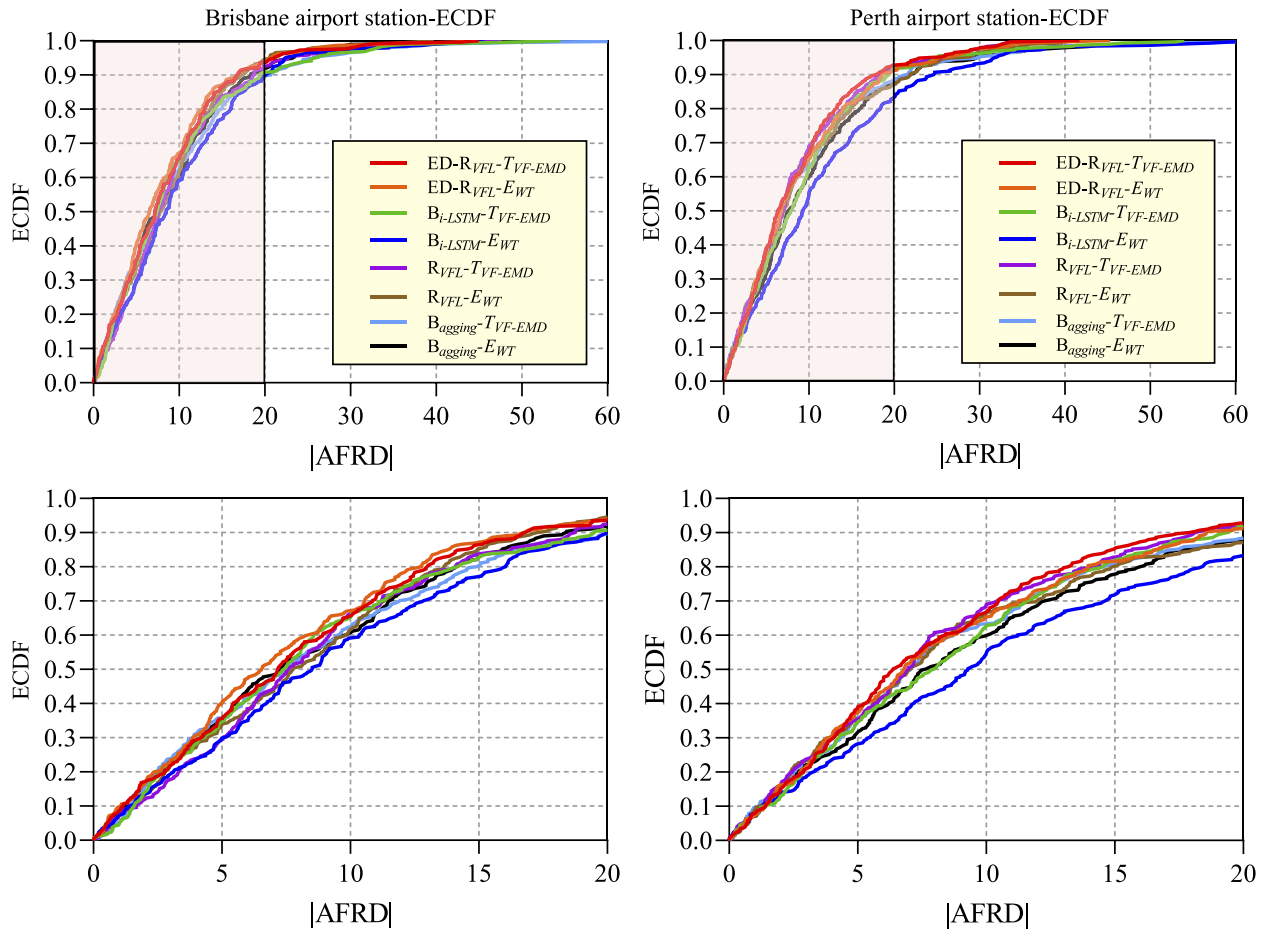


Fig. 13. Empirical cumulative distribution function (ECDF) of absolute forecasting relative deviation, |FARD| (MJ/m^2) for all the hybrid models at Brisbane and Perth airport stations during the testing period.

Perth airport: [0.2033, 0.2787]) for both stations were in an allowable range (i.e., [−2.2, 2.2]) (Geroge & Mallery, 2003; Nie, Bent, & Hull, 1975) for an ML-based forecasting model.

2.10. Intelligent expert framework establishment

In this research, a powerful hybrid model is developed for solar radiation forecasting using air temperature data (T_{max} and T_{min}) for two locations in Australia. This model consists of two pre-processing stages: (1) it uses a powerful feature selection called "extra tree" to detect the effective time lags of each of the predictor signals (Stage 1); and (2) each of the predictors is decomposed using two new and effective decomposition techniques, namely EWT and TVF-EMD, to provide suitable data for feeding the ML models (Stage 2). In this research, a new DL model called Ali was used, for the first time, to solve energy-related problems. Also, RVFL, Bi-LSTM, and Bagging were used for comparisons either in a combined mode with pre-processing steps or in a single mode to compare with the main model. The ED-RVFL, RVFL, and Bi-LSTM models, along with extra tree, were developed in the Python environment with the help of Scikit-learn and Tensorflow libraries, and other techniques (e.g., Bagging) were implemented in the MATLAB environment. The flowchart of the proposed expert system to forecast GSE is shown in Fig. 4. The details on the pre-processing stages of the modeling are described as follows:

Stage 1: data processing and preparing.

Generally, the ground-based meteorological time series datasets have some missing data and outliers, which require some treatments before constructing the predictive ML-based models. In addition, the original daily datasets need to be converted into a weekly scale to forecast the weekly GSE. In this research, the limited missing data and outliers were processed by using the iterative k-nearest neighbor method (Keerin & Boongoen, 2021) and the Grubbs test (Grubbs, 1950; Mishra et al., 2022), respectively. Afterward, the most significant antecedent information or the lagged-time components of each predictor should be accurately determined. Basically, the correlation assesses the relationship between two or more variables. But, the disadvantages of correlational research include no cause and effect, no-inference results, and potentially confounding factors. All these issues are tackled by the newly proposed feature selection technique (i.e., extra tree) for solving the complicated hydrological and environmental problems. Nowadays, similar techniques have been widely used in various fields to solve the optimal feature selection problems (Ahmed et al., 2021; Masrur Ahmed et al., 2021; Prasad et al., 2019). In the current study, the significant lagged-time components were extracted by extra tree feature selection, instead of using the cross-correlation technique. Here, the extra tree by a criterion, namely importance factor specifies the most influential lagged-time sequences to feed the models. The outcomes from the first preprocessing procedure over the predictors (T_{max} and T_{min}) for the two sites are shown in the sun plots in Fig. 5. The values of important factor, as the benchmark criterion are also listed in Fig. 5. The optimal lags obtained in the current stage, listed in the last row of Table 2, are required for the next preprocessing stage.

Stage 2: Predictor decomposing.

Since short-term meteorological or hydrological time series generally are nonstationary and have a lot of noise, using them without any preprocessing in the ML models often does not lead to acceptable results. For this reason, it is necessary to decompose the nonlinear input signals into their intrinsic modes of oscillation before feeding the ML models. In

this research, the two predictors are decomposed into a number of intrinsic mode functions (IMFs) and the corresponding residuals using two new techniques, namely TVF-EMD and EWT. To the best of our knowledge, the TVF-EMD technique is utilized for the first time to forecast GSE in a hybrid model with feature selection and derives from deep learning. This preprocessing scheme can overcome the defects of EMD, including the end effect and mode mixing (K. Wang et al., 2020). Furthermore, the EWT returns the multiresolution analysis (MRA) components as a new advanced extension of the wavelet transform families for signal analysis (Jamei et al., 2022c). TVF-EMD has some tuning parameters including Max_IMF, spline order, bandwidth threshold, and end_flag (Jamei et al., 2022b). The aforementioned optimal parameters are attained by using the trial and error procedure. Also, since the maximum number of decomposition components (Max_IMF) for both stations is 9, the total number of sub-sequences to be fed into the ML models based on the antecedent information chosen for the input variables is 90. The effective EWT tuning parameters include signal type, global trend removal, and polynomial interpolation degree (Jamei et al., 2022b), which are shown in Table 2, along with the optimal parameters of the TVF-EMD. The total numbers of sub-sequences in the EWT-based models for Brisbane (IMFs (T_{max} & T_{min}) = 7) and for Perth airport (IMFs (T_{max}) = 10 and IMFs (T_{min}) = 13) are 70 and 115, respectively. It should be noted that the number of IMFs in the EWT is determined adaptively. As a schematic sample, Fig. 6 shows the decomposed sub-sequences of predictors using the TVF-EMD and EWT schemes for the Perth airport station.

Stage 3: Prerequisites of feeding models.

The data splitting strategy is very important and challenging in constructing the predictive models. So, adopting a suitable strategy according to the length of the time series in forecasting problems can prevent or mitigate the occurrence of overfitting. K-fold cross-validation is one of the most commonly used cross-validation methods to mitigate overfitting. However, this scheme is not recommended for time series due to the gaps it creates in the time series sequence (Cochrane, 2018). Regarding the sufficient length of data (22.5 years) and grid search for tuning ML models, it is possible to avoid the occurrence of overfitting and use the hold-out cross-validation. Thus, 75% of the entire time series is allocated for the training of the models, and the remaining 25% of the dataset is used for model testing. In addition, all the inputs and targets are normalized between 0 and 1 to enhance the efficiency of the models and convergence (Ali et al., 2021). At this stage, the forecasting process using the ML models is carried out for each IMF related to the predictors (T_{min} and T_{max}) for both stations.

After the data required to feed the models are prepared from the previous steps, a strategy is adopted to optimize the results and control overfitting. Pursuing optimal outcomes and tuning the hyperparameters of the ML models are mutually dependent. One of the straightforward techniques that have lately gained a lot of popularity is the random search strategy. The main hyperparameters of ED-RVFL and RVFL are the node numbers, the number of hidden layers, and some regularization coefficients (Shi et al., 2021), while the main hyperparameters of Bi-LSTM (Jamei et al., 2022b) are the number of neurons, number of layers, and learning rate. The tuning of a bagging model is simple and it is developed by defining the fitrensemble function and the "bag" mode in MATLAB. Table 3 lists the optimal hyperparameters of all the hybrid and standalone predictive models gained by the random search strategy. The last stage of modeling is summing up all the forecasts associated with the IMFs to get the weekly GSE.

3. Results assessment and discussion

The performances of the ED-T_{VFL}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, R_{VFL}-$T_{VF\text{-}EMD}$, B_{agging}-$T_{VF\text{-}EMD}$, ED-T_{VFL}-E_{WT}, $B_{i\text{-}LSTM}$-E_{WT}, R_{VFL}-E_{WT}, B_{agging}-E_{WT}, ED-TVFL, Bi-LSTM, RVFL, and bagging models used to forecast weekly GSE for the stations of Perth Airport and Brisbane in Australia were evaluated by using the eight statistical metrics. As shown in Table 4, the ED-T_{VFL}-$T_{VF\text{-}EMD}$ model attained maximum precision for Perth Airport station (R = 0.9665, RMSE = 1.9193, MAE = 1.4898, MAPE = 8.5879, NSE = 0.9288, KGE = 0.9565, I_A = 0.9823, $U_{95\%}$ = 5.2803) in the testing period, following by ED-T_{VFL}-E_{WT} [R = 0.9634, RMSE = 1.9832, MAE = 1.5345, MAPE = 9.0832, NSE = 0.9240, KGE = 0.9583, I_A = 0.9809, $U_{95\%}$ = 5.4685], and R_{VFL}-$T_{VF\text{-}EMD}$ [R = 0.9608, RMSE = 2.0394, MAE = 1.5360, MAPE = 8.7936, NSE = 0.9204, KGE = 0.9484, I_A = 0.9804, $U_{95\%}$ = 5.5988]. The $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, B_{agging}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-E_{WT}, R_{VFL}-E_{WT}, B_{agging}-E_{WT}, ED-TVFL, Bi-LSTM, RVFL, and bagging models (Table 4) were reasonably good to forecast weekly GSE. It was found that the hybrid $T_{VF\text{-}EMD}$ based models had the capabilities to achieve the highest accuracy, followed by the EWT-based models for the Perth Airport station. Further analysis indicated that the standalone models (i.e., ED-TVFL, Bi-LSTM, RVFL, and bagging) had relatively lower accuracy, compared to the hybrid versions. Overall, the ED-T_{VFL}-$T_{VF\text{-}EMD}$ model was the most accurate model for forecasting weekly GSE for the Perth Airport station.

For Brisbane station, the ED-T_{VFL}-E_{WT} model showed better accuracy (R = 0.9218, RMSE = 1.9708, MAE = 1.5230, MAPE = 8.1847, NSE = 0.8476, KGE = 0.8552, I_A = 0.9555, $U_{95\%}$ = 5.4655) (Table 5) in the testing period to forecast weekly GSE than all other models including ED-T_{VFL}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, R_{VFL}-$T_{VF\text{-}EMD}$, B_{agging}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-E_{WT}, R_{VFL}-E_{WT}, B_{agging}-E_{WT}, ED-TVFL, Bi-LSTM, RVFL, and bagging. The results again confirmed that the hybrid E_{WT} and $T_{VF\text{-}EMD}$ models were more accurate for Brisbane station, while the standalone models did not achieve the optimum precision (Table 5).

Fig. 7 shows the comparisons of the TVF-EMD based, EWT based, and standalone/single baseline models for forecasting weekly GSE in forms of binomial distribution plots (left) and scatter plots (right) for both Brisbane and Perth Airport stations. For Brisbane station, ED-T_{VFL}-E_{WT} achieved higher accuracy with a binomial distribution against the observed distribution of the weekly GSE and attained a higher value of R (0.9218) than all other models. The ED-T_{VFL}-$T_{VF\text{-}EMD}$ model showed the best accuracy for Perth Airport station based on a binomial distribution and the scatter plot with R = 0.9665 In summary, the ED-T_{VFL}-E_{WT} mode (for Brisbane station) and the ED-T_{VFL}-$T_{VF\text{-}EMD}$ model (for Perth Airport station) provided better simulations of weekly GSE.

Fig. 8 shows the Taylor diagrams of the reference and forecasted weekly GSE generated by the ED-T_{VFL}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, R_{VFL}-$T_{VF\text{-}EMD}$, B_{agging}-$T_{VF\text{-}EMD}$, ED-T_{VFL}-E_{WT}, $B_{i\text{-}LSTM}$-E_{WT}, R_{VFL}-E_{WT}, B_{agging}-E_{WT}, ED-TVFL, Bi-LSTM, RVFL, and bagging models for Brisbane and Perth airport stations. The Taylor diagrams (Taylor, 2001) provide a thorough and comprehensive assessment by comparing the model simulations with the reference values based on standard deviation and correlation coefficient. As shown in Fig. 8, The GSE from T_{VFL}-$T_{VF\text{-}EMD}$ (green dot) is close to the reference GSE within the arc of 0.95–0.99 for both Brisbane and Perth Airport stations, followed by R_{VFL}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$. Thus, T_{VFL}-$T_{VF\text{-}EMD}$ is the most precise model for weekly GSE forecasting for both stations.

Fig. 9 exhibits the spider plots based on RMSE and MAE (right), and R, NSE, KGE, and IA (left) for the weekly GSE generated by the ED-T_{VFL}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, R_{VFL}-$T_{VF\text{-}EMD}$, B_{agging}-$T_{VF\text{-}EMD}$, ED-T_{VFL}-E_{WT}, $B_{i\text{-}LSTM}$-E_{WT}, R_{VFL}-E_{WT}, B_{agging}-E_{WT}, ED-TVFL, Bi-LSTM, RVFL, and bagging models for Brisbane and Perth Airport stations. The spider plots revealed that the ED-T_{VFL}-E_{WT} model acquired the largest spider web radius of R, NSE, KGE, and IA, and the lowest values of RMSE and MAE for Brisbane station, compared to other models. The ED-T_{VFL}-$T_{VF\text{-}EMD}$ model generated the largest spider web radius based on R, NSE, KGE, and IA, and the lowest values of RMSE and MAE for Perth Airport station. The results demonstrated that the ED-TVFL model, hybridized with the EWT and TVF-EMD decommission methods, improved the modeling accuracy for both stations in the forecast of weekly GSE.

Fig. 10 shows the comparisons of the models in terms of MAPE and $U_{5\%}$ to forecast weekly GSE for both Brisbane and Perth Airport stations. ED-T_{VFL}-E_{WT} and ED-T_{VFL}-$T_{VF\text{-}EMD}$ showed similar performances for Brisbane station, although ED-T_{VFL}-E_{WT} acquired slightly lower MAPE

and $U_{95\%}$ values. The other models attained higher MAPE and $U_{95\%}$ values (or lower accuracy) for Brisbane station. Similar results were obtained for Perth Airport station. ED-T_{VFL}-$T_{VF\text{-}EMD}$ was slightly better than ED-T_{VFL}-E_{WT}, while the other benchmarking models performed poorly in this regard. As shown in Fig. 10, both ED-T_{VFL}-$T_{VF\text{-}EMD}$ and ED-T_{VFL}-E_{WT} provided better forecasting of weekly GSE with reasonably good accuracy for both stations.

Fig. 11 displays the temporal trends of GSE values simulated by different models over the testing period for Brisbane and Perth Airport stations. The ED-T_{VFL}-E_{WT} and ED-T_{VFL}-$T_{VF\text{-}EMD}$ models appeared to be the most accurate in forecasting weekly GSE and showed a similar trend against the observed GSE data for Brisbane station. Similarly, the forecasting trend for Perth Airport station from the ED-T_{VFL}-E_{WT} and ED-T_{VFL}-$T_{VF\text{-}EMD}$ models were also very close to the observed GSE data. Thus, the ED-T_{VFL}-E_{WT} and ED-T_{VFL}-$T_{VF\text{-}EMD}$ models had better accuracy in the time-series plots of weekly GSE for both stations.

A boxplot is an alternative way to show a straightforward comparison and assessment of the models. Fig. 12 displays the boxplots of the forecasted GSE along with Q25%, Median, Q75%, Mean, and IQR for the hybrid models. For the Brisbane station, the ED-T_{VFL}-E_{WT} model exhibited the lowest boxplot distribution with Q25% = −7.471, Median = -0.3819, Q75% = 5.828, Mean = −1.595, and IQR = 13.299 for the forecast of weekly GSE, followed by $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, ED-T_{VFL}-T_{VF}.EMD, B_{agging}-E_{WT}, and B_{agging}-$T_{VF\text{-}EMD}$. The ED-T_{VFL}-$T_{VF\text{-}EMD}$ model acquired better precision [Q25% = −7.644, Median = −1.921, Q75% = 4.879, Mean = −1.922, IQR = 12.523] than other models in the forecast of weekly GSE.

Fig. 13 shows the absolute forecasting relative deviation (|AFRD|) in terms of the empirical cumulative distribution function (ECDF) of the weekly GSE generated by the hybrid version of TVF-EMD, and EWT models (i.e., ED-T_{VFL}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, R_{VFL}-$T_{VF\text{-}EMD}$, B_{agging}-$T_{VF\text{-}EMD}$, ED-T_{VFL}-E_{WT}, $B_{i\text{-}LSTM}$-E_{WT}, R_{VFL}-E_{WT}, and B_{agging}-E_{WT}) for Brisbane and Perth Airport stations. By comparing, the ECDF of the |AFRD| (MJ/m2) with range = 20 indicates that ED-T_{VFL}-$T_{VF\text{-}EMD}$ and ED-T_{VFL}-E_{WT} achieved a very close profile with ECDF convergence above 0.90% for both Brisbane and Perth Airport stations in contrast to the benchmark $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, R_{VFL}-$T_{VF\text{-}EMD}$, B_{agging}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-E_{WT}, R_{VFL}-E_{WT}, and B_{agging}-E_{WT} models, which further demonstrated the better performances of the ED-T_{VFL}-$T_{VF\text{-}EMD}$ and ED-T_{VFL}-E_{WT} models to forecast weekly GSE.

4. Conclusion

In this research, a novel modelling approach was developed based on time-varying filter-based empirical mode decomposition (TVF-EMD) unified with ensemble deep-RVFL (i.e., ED-R_{VFL}-$T_{VF\text{-}EMD}$) to forecast weekly GSE for Brisbane and Perth Airport stations. The extra tree feature selection algorithm was used to determine the most significant antecedent information, or the lagged-time components of each predictor. The TVF-EMD technique was applied to decompose the significant lagged-time components into the IMFs and residual components. Another significant aspect of this study lies in the development of the TVF-EMD with empirical wavelet transformation (EWT) and the ED-T_{VFL}-E_{WT} model. The pre-processing decomposition methods TVF-EMD and EWT were implemented to solve the non-stationarity and non-linearity issues associated with the data. Furthermore, along with the main ML model (i.e., ED-RVFL), the RVFL, Bi-LSTM, and bagging models were hybridized with the TVF-EMD and EWT models, forming the $B_{i\text{-}LSTM}$-$T_{VF\text{-}EMD}$, R_{VFL}-$T_{VF\text{-}EMD}$, B_{agging}-$T_{VF\text{-}EMD}$, $B_{i\text{-}LSTM}$-E_{WT}, R_{VFL}-E_{WT}, and B_{agging}-E_{WT} models for forecasting weekly GSE. The analyses demonstrated the advantages of the extra tree feature selection, ED-TVF and EWT pre-processing models, which significantly improved the forecasting accuracy of the ED-RVFL model. The results confirmed that the ED-T_{VFL}-$T_{VF\text{-}EMD}$ and ED-T_{VFL}-E_{WT} were significantly better than the benchmarking models in the forecast of weekly GSE for both stations. It was noted that the ED-T_{VFL}-$T_{VF\text{-}EMD}$ (R = 0.9665, RMSE = 1.9193 MJ/m^2, and KGE = 0.9565) and ED-T_{VFL}-E_{WT} (R = 0.9218, RMSE = 1.9708 MJ/m^2, and KGE = 0.8552) models achieved 6% − 7% and 7% − 8% improvement in accuracy for Perth airport and Brisbane stations in comparison with the benchmarking models.

In the future studies, more advanced techniques such as quantum algorithms can be used for optimization. Explainable AI models can be integrated to overcome the limitations of the black-box nature to understand and validate the complex relationships among the predictor data during the learning stage. The scope of this research can be further broadened to solve various environmental, hydrological, and agricultural problems, and assist decision makers in accurate forecasts to better cope with the issues related to climate change, agriculture, and energy crisis.

CRediT authorship contribution statement

Mehdi Jamei: Conceptualization, Formal analysis, Visualization, Modeling, Software, Writing up the manuscript, Project administration; **Masoud Karbasi:** Methodology, Writing up the manuscript and manuscript editing; **Mumtaz Ali:** Investigation, Validation, Writing up the manuscript, Supervision; **Anurag Malik:** Methodology, Writing up the manuscript; **Xuefeng Chu:** Supervision, Editing and reviewing; **Zaher Mundher Yaseen:** Investigation, and Writing up the manuscript.

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

The authors do not have permission to share data.

Acknowledgements

The authors are duly acknowledging the Bureau of Meteorology, Australia (BMA) to provide the daily global solar exposure (GSE) data and other relevant input data for this research.

References

Abubakr, M., Akoush, B., Khalil, A., & Hassan, M. A. (2022). Unleashing deep neural network full potential for solar radiation forecasting in a new geographic location with historical data scarcity: A transfer learning approach. *The European Physical Journal Plus, 137*(4), 474.

Acikgoz, H. (2022). A novel approach based on integration of convolutional neural networks and deep feature selection for short-term solar radiation forecasting. *Applied Energy*, *305*(October 2021), 117912. 10.1016/j.apenergy.2021.117912.

Adnan, R. M., Mostafa, R. R., Islam, A. R. M. T., Gorgij, A. D., Kuriqi, A., & Kisi, O. (2021). Improving Drought Modeling Using Hybrid Random Vector Functional Link Methods. *Water, 13*(23), 3379. https://doi.org/10.3390/w13233379

Ahmed, A. A. M., Deo, R. C., Ghahramani, A., Raj, N., Feng, Q., Yin, Z., & Yang, L. (2021). LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios. *Stochastic Environmental Research and Risk Assessment.*. https://doi.org/10.1007/s00477-021-01969-3

Ali, A. I. M., & Mohamed, H. R. A. (2022). Improved P&O MPPT algorithm with efficient open-circuit voltage estimation for two-stage grid-integrated PV system under realistic solar radiation. *International Journal of Electrical Power & Energy Systems, 137*, Article 107805.

Apaydin, H., Feizi, H., Sattari, M. T., Colak, M. S., Shamshirband, S., & Chau, K.-W. (2020). Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting. *Water, 12*(5), 1500. https://doi.org/10.3390/w12051500

Avand, M., Janizadeh, S., Tien Bui, D., Pham, V. H., Ngo, P. T. T., & Nhu, V.-H. (2020). A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *International Journal of Digital Earth, 13*(12), 1408–1429. https://doi.org/10.1080/17538947.2020.1718785

Bailek, N., Bouchouicha, K., Al-Mostafa, Z., El-Shimy, M., Aoun, N., Slimani, A., & Al-Shehri, S. (2018). A new empirical model for forecasting the diffuse solar radiation over Sahara in the Algerian Big South. *Renewable Energy, 117*, 530–537.

Bamisile, O., Oluwasanmi, A., Ejiyi, C., Yimen, N., Obiora, S., & Huang, Q. (2022). Comparison of machine learning and deep learning algorithms for hourly global/

diffuse solar radiation predictions. *International Journal of Energy Research, 46*(8), 10052–10073.
Basilio, S. A., & Goliatt, L. (2022). Gradient Boosting Hybridized with Exponential Natural Evolution Strategies for Estimating the Strength of Geopolymer Self-Compacting Concrete. *Knowledge-Based Engineering and Sciences, 3*(1), 1–16.
Bokde, N. D., Yaseen, Z. M., & Andersen, G. B. (2020). ForecastTB—An R Package as a Test-Bench for Time Series Forecasting—Application of Wind Speed and Solar Radiation Modeling. *Energies, 13*(10), 2578.
Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.
Cheng, W. X., Suganthan, P. N., & Katuwal, R. (2021). Time series classification using diversified ensemble deep random vector functional link and resnet features. *Applied Soft Computing, 112*, Article 107826.
Cochrane, C. (2018). Time Series Nested Cross-Validation. Towards Data Science. acessado 24/03/2019: https://towardsdatascience. com/time-series-nestedcross.
Cogliani, E., Ricchiazzi, P., & Maccari, A. (2007). Physical model SOLARMET for determinating total and direct solar radiation by meteosat satellite images. *Solar Energy, 81*(6), 791–798.
Daubechies, I. (1992). Ten lectures on wavelets, CBMS-NSF Reg. In *Conf. Ser. Appl. Math* (Vol. 61, p. 357).
Dey, S., Dey, A. K., & Mall, R. K. (2021). Modeling Long-term Groundwater Levels By Exploring Deep Bidirectional Long Short-Term Memory using Hydro-climatic Data. *Water Resources Management, 35*(10), 3395–3410. https://doi.org/10.1007/s11269-021-02899-z
El Mghouchi, Y., Chham, E., Zemmouri, E. M., & El Bouardi, A. (2019). Assessment of different combinations of meteorological parameters for predicting daily global solar radiation using artificial neural networks. *Building and Environment, 149*, 607–622.
Eseye, A. T., Zhang, J., & Zheng, D. (2018). Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. *Renewable Energy, 118*, 357–367. https://doi.org/10.1016/j.renene.2017.11.011
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. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2019.01.040
Fan, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X., & Zou, H. (2019b). Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2018.10.018
Gautier, C., Diak, G., & Masse, S. (1980). A simple physical model to estimate incident solar radiation at the surface from GOES satellite data. *Journal of Applied Meteorology and Climatology, 19*(8), 1005–1012.
Geroge, D., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference*. Boston: Allyn & Bacon.
Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3–42.
Ghimire, S., Deo, R. C., Downs, N. J., & Raj, N. (2019). Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia. *Journal of Cleaner Production, 216*, 288–310.
Gilles, J. (2013). Empirical wavelet transform. *IEEE Transactions on Signal Processing, 61*(16), 3999–4010.
Goliatt, L., & Yaseen, Z. M. (2022). Development of a hybrid computational intelligent model for daily global solar radiation prediction. *Expert Systems with Applications, 118295*.
Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*(5–6), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042
Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *The Annals of Mathematical Statistics*, 27–58.
Gu, Q., Chang, Y., Xiong, N., & Chen, L. (2021). Forecasting Nickel futures price based on the empirical wavelet transform and gradient boosting decision trees. *Applied Soft Computing, 109*, Article 107472. https://doi.org/10.1016/j.asoc.2021.107472
Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology, 377*(1–2), 80–91.
Hai, T., Sharafati, A., Mohammed, A., Salih, S. Q., Deo, R. C., Al-Ansari, N., & Yaseen, Z. M. (2020). Global Solar Radiation Estimation and Climatic Variability Analysis Using Extreme Learning Machine Based Predictive Model. *IEEE Access, 8*, 12026–12042. https://doi.org/10.1109/ACCESS.2020.2965303
Huang, X., Hayashi, K., Matsumoto, T., Tao, L., Huang, Y., & Tomino, Y. (2022). Estimation of Rooftop Solar Power Potential by Comparing Solar Radiation Data and Remote Sensing Data—A Case Study in Aichi. *Japan. Remote Sensing, 14*(7), 1742.
Ihianle, I. K., Nwajana, A. O., Ebenuwa, S. H., Otuka, R. I., Owa, K., & Orisatoki, M. O. (2020). A Deep Learning Approach for Human Activities Recognition From Multimodal Sensing Devices. *IEEE Access, 8*, 179028–179038. https://doi.org/10.1109/ACCESS.2020.3027979
Jahani, B., Dinpashoh, Y., & Nafchi, A. R. (2017). Evaluation and development of empirical models for estimating daily solar radiation. *Renewable and Sustainable Energy Reviews, 73*, 878–891.
Jamei, M., Ahmadianfar, I., Chu, X., & Yaseen, Z. M. (2020). Estimation of triangular side orifice discharge coefficient under a free flow condition using data-driven models. *Flow Measurement and Instrumentation*, 101878.
Jamei, M., Ali, M., Malik, A., Prasad, R., Abdulla, S., & Yaseen, Z. M. (2022b). Forecasting Daily Flood Water Level Using Hybrid Advanced Machine Learning Based Time-Varying Filtered Empirical Mode Decomposition Approach. *Water Resources Management*. https://doi.org/10.1007/s11269-022-03270-6
Jamei, M., Karbasi, M., Malik, A., Jamei, M., Kisi, O., & Yaseen, Z. M. (2022c). Long-term multi-step ahead forecasting of root zone soil moisture in different climates: Novel ensemble-based complementary data-intelligent paradigms. *Agricultural Water Management, 269*, Article 107679.
Kaaya, I., Ascencio-Vásquez, J., Weiss, K.-A., & Topič, M. (2021). Assessment of uncertainties and variations in PV modules degradation rates and lifetime predictions using physical models. *Solar Energy, 218*, 354–367.
Kang, H., Yang, S., Huang, J., & Oh, J. (2020). Time Series Prediction of Wastewater Flow Rate by Bidirectional LSTM Deep Learning. *International Journal of Control, Automation and Systems, 18*(12), 3023–3030. https://doi.org/10.1007/s12555-019-0984-6
Kannan, N., & Vakeesan, D. (2016). Solar energy for future world:-A review. *Renewable and Sustainable Energy Reviews, 62*, 1092–1105.
Keerin, P., & Boongoen, T. (2021). Improved knn imputation for missing values in gene expression data. *Computers, Materials & Continua*.
Khosravi, A., Koury, R. N. N., Machado, L., & Pabon, J. J. G. (2018). Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. *Journal of Cleaner Production, 176*, 63–75. https://doi.org/10.1016/j.jclepro.2017.12.065
Kisi, O., Heddam, S., & Yaseen, Z. M. (2019). The implementation of univariable scheme-based air temperature for solar radiation prediction: New development of dynamic evolving neural-fuzzy inference system model. *Applied Energy, 241*, 184–195. https://doi.org/10.1016/j.apenergy.2019.03.089
Kumari, P., & Toshniwal, D. (2021a). Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production, 318*, Article 128566.
Kumari, P., & Toshniwal, D. (2021b). Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *Journal of Cleaner Production, 279*, Article 123285.
Le, X.-H., Nguyen, D.-H., Jung, S., Yeon, M., & Lee, G. (2021). Comparison of Deep Learning Techniques for River Streamflow Forecasting. *IEEE Access, 9*, 71805–71820. https://doi.org/10.1109/ACCESS.2021.3077703
Li, F., Ma, G., Chen, S., & Huang, W. (2021). An Ensemble Modeling Approach to Forecast Daily Reservoir Inflow Using Bidirectional Long- and Short-Term Memory (Bi-LSTM), Variational Mode Decomposition (VMD), and Energy Entropy Method. *Water Resources Management, 35*(9), 2941–2963. https://doi.org/10.1007/s11269-021-02879-3
Li, H., Li, Z., & Mo, W. (2017). A time varying filter approach for empirical mode decomposition. *Signal Processing, 138*, 146–158. https://doi.org/10.1016/j.sigpro.2017.03.019
Li, Y.-H., Harfiya, L. N., Purwandari, K., & Lin, Y.-D. (2020). Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model. *Sensors, 20*(19), 5606. https://doi.org/10.3390/s20195606
Liang, S., Wang, D., Wu, J., Wang, R., & Wang, R. (2021). Method of Bidirectional LSTM Modelling for the Atmospheric Temperature. *Intelligent Automation & Soft Computing, 29*(3), 701–714. 10.32604/iasc.2021.020010.
Maddu, R., Vanga, A. R., Sajja, J. K., Basha, G., & Shaik, R. (2021). Prediction of land surface temperature of major coastal cities of India using bidirectional LSTM neural networks. *Journal of Water and Climate Change, 12*(8), 3801–3819. https://doi.org/10.2166/wcc.2021.460
Manieniyan, V., Thambidurai, M., & Selvakumar, R. (2009). Study on energy crisis and the future of fossil fuels. *Proceedings of SHEE, 10*, 2234–3689.
Masrur Ahmed, A. A., Deo, R. C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., & Yang, L. (2021). Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. *Journal of Hydrology, 599*(August 2020), 126350. 10.1016/j.jhydrol.2021.126350.
Minh-Tuan, N., & Kim, Y.-H. (2019). Bidirectional Long Short-Term Memory Neural Networks for Linear Sum Assignment Problems. *Applied Sciences, 9*(17), 3470. https://doi.org/10.3390/app9173470
Mishra, G., Sulieman, M. M., Kaya, F., Francaviglia, R., Keshavarzi, A., Bakhshandeh, E., … Elmobarak, A. (2022). Machine learning for cation exchange capacity prediction in different land uses. *CATENA, 216*, Article 106404.
Monjoly, S., Andrï, M., Calif, R., & Soubdhan, T. (2017). Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach. *Energy, 119*, 288–298. https://doi.org/10.1016/j.energy.2016.11.061
Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Transactions of the ASABE, 58*(6), 1763–1785. 10.13031/trans.58.10715.
Mostafavi, E. S., Ramiyani, S. S., Sarvar, R., Moud, H. I., & Mousavi, S. M. (2013). A hybrid computational approach to estimate solar global radiation: An empirical evidence from Iran. *Energy, 49*(1), 204–210. https://doi.org/10.1016/j.energy.2012.11.023
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology, 10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6
Nie, N. H., Bent, D. H., & Hull, C. H. (1975). *SPSS: Statistical package for the social sciences, 227*. McGraw-Hill New York.
Olpenda, A. S., Stereńczak, K., & Będkowski, K. (2018). Modeling solar radiation in the forest using remote sensing data: A review of approaches and opportunities. *Remote Sensing, 10*(5), 694.
Pao, Y.-H., Park, G.-H., & Sobajic, D. J. (1994). Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing, 6*(2), 163–180.
Park, K., Jung, Y., Kim, K., & Park, S. K. (2020). Determination of Deep Learning Model and Optimum Length of Training Data in the River with Large Fluctuations in Flow Rates. *Water, 12*(12), 3537. https://doi.org/10.3390/w12123537

Patino, C. M., & Ferreira, J. C. (2015). Confidence intervals: A useful statistical tool to estimate effect sizes in the real world. *Jornal Brasileiro de Pneumologia*. https://doi.org/10.1590/s1806-37562015000000314

Prasad, R., Deo, R. C., Li, Y., & Maraseni, T. (2019). Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *Catena, 177*(August 2018), 149–166. 10.1016/j.catena.2019.02.012.

Qazi, A., Fayaz, H., Wadi, A., Raj, R. G., Rahim, N. A., & Khan, W. A. (2015). The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review. *Journal of Cleaner Production, 104*, 1–12.

Rehamnia, I., Benlaoukli, B., Jamei, M., Karbasi, M., & Malik, A. (2021). Simulation of seepage flow through embankment dam by using a novel extended Kalman filter based neural network paradigm: Case study of Fontaine Gazelles Dam. *Algeria. Measurement, 176*, Article 109219.

Sada, A. Y., Faskari, S. A., Ilyasu, F. B., & Abba, S. I. (2022). Application of Different Membership Function for Short-term Load Demand Estimation: A Neuro-Fuzzy Approach. *Knowledge-Based Engineering and Sciences, 3*(3), 93–100.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing, 45*(11), 2673–2681. https://doi.org/10.1109/78.650093

Şenkal, O. (2010). Modeling of solar radiation using remote sensing and artificial neural network in Turkey. *Energy, 35*(12), 4795–4801. https://doi.org/10.1016/j.energy.2010.09.009

Seyyedattar, M., Ghiasi, M. M., Zendehboudi, S., & Butt, S. (2020). Determination of bubble point pressure and oil formation volume factor: Extra trees compared with LSSVM-CSA hybrid and ANFIS models. *Fuel, 269*, Article 116834. https://doi.org/10.1016/j.fuel.2019.116834

Sharafati, A., Khosravi, K., Khosravinia, P., Ahmed, K., Salman, S. A., Yaseen, Z. M., & Shahid, S. (2019). The potential of novel data mining models for global solar radiation prediction. *International Journal of Environmental Science and Technology, 16*(11), 7147–7164. https://doi.org/10.1007/s13762-019-02344-0

Sharma, V., Yang, D., Walsh, W., & Reindl, T. (2016). Short term solar irradiance forecasting using a mixed wavelet neural network. *Renewable Energy, 90*, 481–492. https://doi.org/10.1016/j.renene.2016.01.020

Shi, Q., Katuwal, R., Suganthan, P. N., & Tanveer, M. (2021). Random vector functional link neural network based ensemble deep learning. *Pattern Recognition, 117*, Article 107978.

Singh, U. K., Jamei, M., Karbasi, M., Malik, A., & Pandey, M. (2022). Application of a modern multi-level ensemble approach for the estimation of critical shear stress in cohesive sediment mixture. *Journal of Hydrology, 607*, Article 127549.

Song, C., Chen, X., Wu, P., & Jin, H. (2021). Combining time varying filtering based empirical mode decomposition and machine learning to predict precipitation from nonlinear series. *Journal of Hydrology, 603*, Article 126914. https://doi.org/10.1016/j.jhydrol.2021.126914

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of Statistics, 24*, 303–329.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research Atmospheres, 106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Tracker, C. A. (2022). Global reaction to energy crisis risks zero carbon transition.

Voyant, C., Notton, G., Duchaud, J. L., Almorox, J., & Yaseen, Z. M. (2020). Solar irradiation prediction intervals based on Box-Cox transformation and univariate representation of periodic autoregressive model. *Renewable Energy Focus*. https://doi.org/10.1016/j.ref.2020.04.001

Wang, K., Fu, W., Chen, T., Zhang, B., Xiong, D., & Fang, P. (2020). A compound framework for wind speed forecasting based on comprehensive feature selection, quantile regression incorporated into convolutional simplified long short-term memory network and residual error correction. *Energy Conversion and Management, 222*(May), Article 113234. https://doi.org/10.1016/j.enconman.2020.113234

Wang, M., & Sheng, X. (2022). Combining empirical wavelet transform and transfer matrix or modal superposition to reconstruct responses of structures subject to typical excitations. *Mechanical Systems and Signal Processing, 163*, Article 108162.

Wehenkel, L., Ernst, D., & Geurts, P. (2006). Ensembles of extremely randomized trees and some generic applications. *Proceedings of Robust Methods for Power System State Estimation and Load Forecasting*.

Willmott, C. J. (1981). On the validation of models. *Physical Geography, 2*(2), 184–194. https://doi.org/10.1080/02723646.1981.10642213

Wu, R., Qin, Z., & Liu, B.-Y. (2022). A systematic analysis of dynamic frequency spillovers among carbon emissions trading (CET), fossil energy and sectoral stock markets: Evidence from China. *Energy, 124176*.

Yin, J., Deng, Z., Ines, A. V. M., Wu, J., & Rasu, E. (2020). Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). *Agricultural Water Management*. https://doi.org/10.1016/j.agwat.2020.106386

Zayed, M. E., Zhao, J., Li, W., Elsheikh, A. H., Abd Elaziz, M., Yousri, D., … Mingxi, Z. (2021). Predicting the performance of solar dish Stirling power plant using a hybrid random vector functional link/chimp optimization model. *Solar Energy, 222*, 1–17.

Zhang, C., & Ma, Y. (2012). *Ensemble Machine Learning*. (C. Zhang & Y. Ma, Eds.). Boston, MA: Springer US. 10.1007/978-1-4419-9326-7.

Zhang, Y., Chiew, F. H. S., Li, M., & Post, D. (2018). Predicting Runoff Signatures Using Regression and Hydrological Modeling Approaches. *Water Resources Research, 54*(10), 7859–7878. https://doi.org/10.1029/2018WR023325

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology, 598*. https://doi.org/10.1016/j.jhydrol.2021.126266