Sains Malaysiana 55(10)(2026): 1690-1702
http://doi.org/10.17576/jsm-2026-5510-08

Predictive Reliability of Denoised QM-Corrected Satellite-Derived Photosynthetically Active Radiation (PAR) under Humid Tropical Conditions

(Kebolehpercayaan Ramalan bagi Sinaran Aktif Fotosintesis (PAR) Terbitan Satelit yang Telah Dibuang Hingar dan Dibetulkan secara QM di bawah Keadaan Tropika Lembap)

ROHAIZA DAUD[1,2,*], FARRAH MELISSA MUHARAM[1], RADHWANE DERRAZ[3], NGAI PAING TAN[4] & NOR AZURA HUSIN[5]

[1]*Department of Agriculture Technology, Faculty of Agriculture, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[2]*Centre for Pre-University Studies, Universiti Malaysia Sarawak (UNIMAS), 94300 Kota Samarahan, Sarawak, Malaysia*
[3]*School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*
[4]*Department of Land Management, Faculty of Agriculture, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[5]*Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

Received: 12 June 2026/Accepted: 20 September 2026

*Corresponding author; email: drohaiza@unimas.my

ABSTRACT

Satellite-derived photosynthetically active radiation (PAR) products provide valuable information for environmental monitoring and agricultural applications, but residual fluctuations may remain even after bias correction, potentially affecting predictive reliability. This study evaluated the influence of multiple denoising approaches, including decomposition-, filtering-, smoothing-, and learning-based methods, on the predictive performance of quantile mapping (QM)-corrected satellite-derived PAR under humid tropical conditions in Peninsular Malaysia. MODIS MCD18A2 Version 061 PAR data were obtained for 14 meteorological stations between 2015 and 2023 and corrected using QM prior to denoising. Predictive performance was assessed using five regression models. Most denoising approaches improved predictive performance relative to the baseline QM-corrected dataset, although the magnitude of improvement varied considerably among methods. Median filtering paired with Random Forest produced the highest mean R^2 (0.748), with a mean MSE of 0.293. Although LSTM yielded a lower numerical MSE (0.004), its lower mean R^2 (0.542) and nearly flat reconstruction indicated variance compression rather than faithful preservation of PAR dynamics. Median filtering was therefore selected as the most reliable overall method. In contrast, Empirical Mode Decomposition (EMD) exhibited inconsistent performance across stations and substantially larger prediction errors. The findings demonstrate that denoising can substantially enhance the predictive reliability of satellite-derived PAR. More importantly, the results suggest that predictive reliability is influenced more by effective signal preservation than by algorithmic complexity alone. Under humid tropical conditions, relatively simple denoising approaches may provide more robust improvements than computationally intensive alternatives for PAR-based environmental modelling applications.

Keywords: Denoising; machine learning; MODIS MCD18A2; photosynthetically active radiation; quantile mapping

ABSTRAK

Produk sinaran aktif fotosintesis (PAR) berasaskan satelit menyediakan maklumat penting untuk pemantauan alam sekitar dan aplikasi pertanian, namun lebihan turun naik masih boleh wujud walaupun selepas pembetulan bias, sekali gus berpotensi menjejaskan kebolehpercayaan ramalan. Kajian ini menilai pengaruh pelbagai pendekatan penyahhingaran, termasuk kaedah berasaskan penguraian, penapisan, pelicinan dan pembelajaran, terhadap prestasi ramalan PAR berasaskan satelit yang telah diperbetulkan menggunakan pemetaan kuantil (QM) di bawah keadaan tropika lembap di Semenanjung Malaysia. Data PAR MODIS MCD18A2 Versi 061 diperoleh bagi 14 stesen meteorologi untuk tempoh 2015 hingga 2023 dan diperbetulkan menggunakan QM sebelum proses penyahhingaran dijalankan. Prestasi ramalan dinilai menggunakan lima model regresi. Kebanyakan kaedah penyahhingaran meningkatkan prestasi ramalan berbanding set data asas yang telah diperbetulkan menggunakan QM, walaupun tahap peningkatan berbeza antara kaedah. Penapisan median yang digandingkan dengan hutan rawak menghasilkan purata R^2 tertinggi (0.748) dan purata MSE

sebanyak 0.293. Walaupun LSTM menghasilkan MSE yang lebih rendah (0.004), purata R^2 yang lebih rendah (0.542) dan pembinaan semula hampir mendatar menunjukkan pemampatan varians, bukannya pengekalan dinamik PAR. Oleh itu, penapisan median dipilih sebagai kaedah yang paling boleh dipercayai secara keseluruhan. Sebaliknya, Penguraian Mod Empirik (EMD) menunjukkan prestasi yang tidak tekal antara stesen serta menghasilkan ralat ramalan yang lebih tinggi. Keputusan kajian menunjukkan bahawa penyahhingaran dapat meningkatkan kebolehpercayaan ramalan PAR berasaskan satelit. Yang lebih penting, kebolehpercayaan ramalan didapati lebih dipengaruhi oleh pemeliharaan maklumat temporal yang bermakna berbanding kerumitan algoritma semata-mata. Di bawah keadaan tropika lembap, pendekatan penyahhingaran yang lebih ringkas berpotensi memberikan peningkatan yang lebih mantap bagi aplikasi pemodelan alam sekitar berasaskan PAR.

Kata kunci: MODIS MCD18A2; pembelajaran mesin; pemetaan kuantil; penyahhingaran; sinaran aktif fotosintesis

INTRODUCTION

Photosynthetically active radiation (PAR) represents the portion of solar radiation utilised by plants during photosynthesis and plays an important role in regulating plant growth, biomass accumulation, and crop productivity (McCree 1972). Accurate PAR information is therefore essential for agricultural monitoring, ecological assessment, and environmental modelling applications (Liu et al. 2025; Zhao et al. 2023). In recent years, satellite-derived PAR products have become increasingly valuable because they provide spatially continuous observations and long-term temporal records over large geographical areas, thereby overcoming the limitations of sparse ground-based measurement networks (Liu et al. 2025; Singh et al. 2022). Among the available products, the Moderate Resolution Imaging Spectroradiometer (MODIS)-derived PAR dataset has been widely applied in environmental and agricultural studies due to its long-term consistency, accessibility, and ability to support large-scale spatiotemporal analysis (Wang et al. 2020).

However, despite these advantages, satellite-derived PAR products remain susceptible to various sources of uncertainty, particularly under humid tropical conditions. Persistent cloud cover, high atmospheric moisture (Gould et al. 2019), aerosol variability, and rapidly changing weather conditions may introduce inconsistencies and fluctuations within satellite-derived PAR signals (Wang et al. 2020). These uncertainties may subsequently affect the reliability of downstream analyses, including environmental assessment and prediction applications (Finazzi, Rodeschini & Tedesco 2025). Previous studies have shown that bias correction techniques such as quantile mapping (QM) can improve the agreement between satellite-derived and ground-observed environmental datasets by reducing systematic errors and distributional inconsistencies (Bailey et al. 2025; Katiraie-Boroujerdy et al. 2020).

In this study, quantile mapping (QM)-corrected satellite-derived PAR data were used as the baseline dataset prior to the application of denoising methods. Although QM correction reduces distributional inconsistencies and systematic bias, residual temporal fluctuations may still remain within the corrected PAR signals (Bailey et al. 2025). Such residual noise may influence the consistency of predictive modelling performance, particularly when machine learning approaches are used to analyse environmental datasets (Pelletier et al. 2017).

To minimise these fluctuations, denoising approaches have increasingly been explored in environmental signal processing to suppress noise fluctuations while preserving meaningful temporal information (Mariethoz et al. 2015). Various denoising techniques, including decomposition-based, filtering-based, smoothing-based, and learning-based approaches, have demonstrated potential in improving signal stability and reducing noise-related variability in environmental and satellite-derived time-series datasets (Hird & McDermid 2009; Priyadarshi et al. 2018; Samal, Babu & Das 2021). Nevertheless, the influence of denoising methods on the predictive reliability of corrected satellite-derived PAR datasets under humid tropical conditions remains insufficiently explored.

To address this gap, the study brought station-specific QM correction, ten denoising methods and five regression models into a single station-based analysis. The models comprised Linear Regression, Random Forest, Gradient Boosting, Neural Network and Bagging Regressor (Finazzi, Rodeschini & Tedesco 2025; Kumari & Singh 2023). Rather than judging denoising from signal smoothness alone, the analysis examined whether changes in signal characteristics improved the ability of corrected satellite-derived PAR to represent ground-derived reference PAR without erasing meaningful temporal variation. Applying the same workflow across fourteen stations allowed simple and learning-based methods to be compared under a common humid tropical setting.

MATERIALS AND METHODS

The study workflow is shown in Figure 1, covering data preprocessing, QM correction, denoising, regression modelling, and performance evaluation.

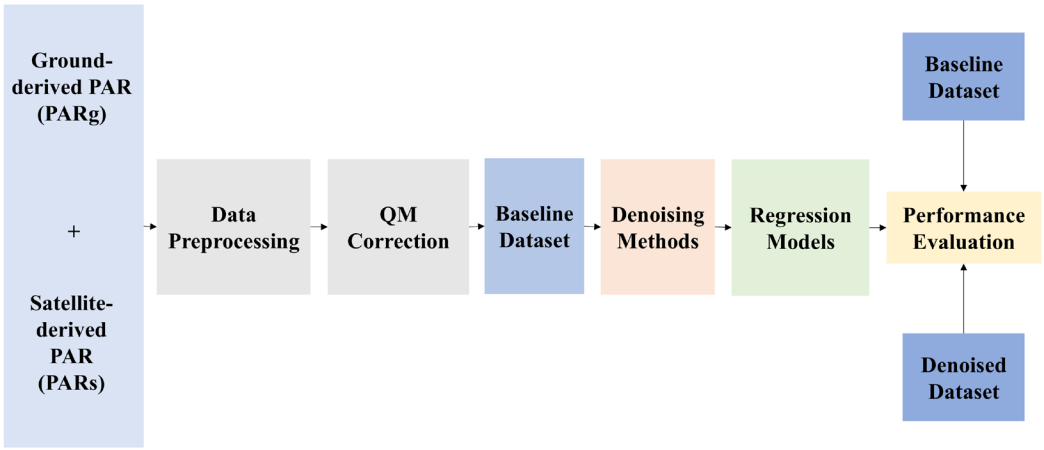


FIGURE 1. Flowchart of the methodological framework applied in this study

STUDY AREA

Peninsular Malaysia (PM), also known as West Malaysia, is located at the southern margin of the Asian continental plate (He et al. 2021) (Figure 2). It lies between 1°N–7°N and 99°E–105°E within the humid tropical zone, where incoming solar radiation remains relatively high throughout the year (Tan et al. 2021). Due to its equatorial position, day length varies only slightly, averaging about 12 h per day (Shavalipour et al. 2013). The regional climate is mainly influenced by the Northeast Monsoon (NEM) and Southwest Monsoon (SWM). The NEM, from November to March, is generally associated with higher rainfall, frequent cloud cover, and occasional flooding, while the SWM, from May to August, is comparatively drier (Tan et al. 2021). The inter-monsoon periods further contribute to short-term atmospheric variability. These seasonal changes affect cloud formation and incoming solar radiation, making Peninsular Malaysia a suitable region for evaluating denoising approaches for satellite-derived PAR under humid tropical conditions.

DATA ACQUISITION AND PREPROCESSING

GROUND-DERIVED REFERENCE PAR (PARg)

Ground-derived reference PAR (PARg) was estimated from daily global solar radiation (SR) records obtained from fourteen meteorological stations in Peninsular Malaysia during 2015-2023. As direct PAR measurements were unavailable, PARg was expressed on an energy basis and estimated using a PAR-to-global-solar-radiation ratio of 0.42 reported in a previous tropical study conducted in Singapore (Tan & Ismail 2016). The coefficient was adopted as a regional approximation because Singapore and Peninsular Malaysia share an equatorial humid tropical setting. The calculation was performed using Equation (1):

$$PAR_g = \mathrm{SR} \times 0.42$$

where PARg and SR are expressed in MJ M^{-2} day^{-1}. PARg therefore represents a ground-derived estimate based on measured global solar radiation rather than a direct PAR measurement. The station characteristics are summarized in Table 1.

SATELLITE-DERIVED PAR (PARs)

Satellite-derived PAR (PARs) data were obtained from the MODIS MCD18A2 Version 061 product for the period 2015-2023. The product provides spatially continuous PAR estimates at 1 km spatial resolution and 3-hourly temporal resolution (Wang et al. 2020). The data were accessed from the NASA Land Processes Distributed Active Archive Center (LP DAAC) and extracted based on the locations of the meteorological stations used in this study. The extracted PARs values, originally expressed in W m^{-2}, were converted to MJ m^{-2} day^{-1} to match the unit of the ground-derived PAR dataset.

Several preprocessing steps were performed before the correction and denoising analyses. This included spatial alignment to WGS 1984/UTM Zones 47N and 48N, temporal consistency checking, station-level pairing between PARg and PARs, and removal of incomplete observations. Outliers were subsequently identified and handled using the interquartile range (IQR) method. The processed dataset was subsequently subjected to quantile mapping (QM), denoising and predictive modelling. Data extraction and preprocessing were conducted using Google Earth Engine, ArcMap 10.8.2, and Python. For temporal visualization, monthly mean PAR values were calculated from the daily observations across the study period.

Before denoising, PARs were corrected independently for each station using empirical Quantile Mapping Mean Bias Correction (QM_MBC). For each station, empirical cumulative distribution functions were constructed from the paired PARs and PARg records. Each daily PARs value was assigned its cumulative

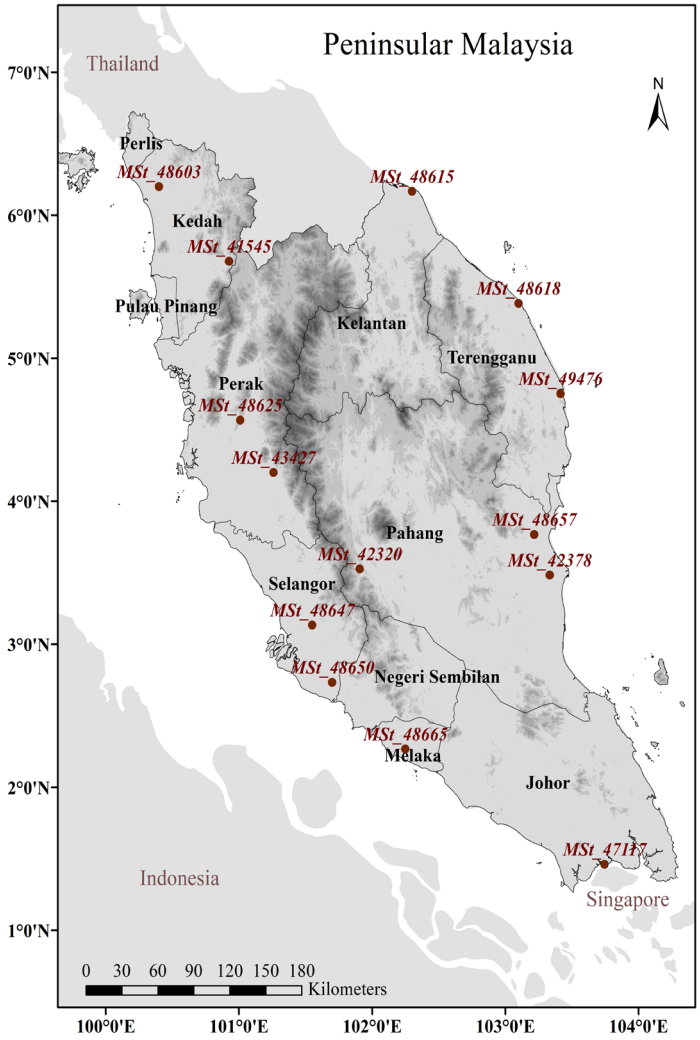


FIGURE 2. Location of Peninsular Malaysia and distribution of meteorological stations

TABLE 1. Characteristics of meteorological stations used in this study

Station Code	Location	Coordinates (Lat N°, Lon E°)	Elevation (m)
MSt_47117	Johor Bahru, Johor	1°27’N, 103°44’E	15.30
MSt_49476	Dungun, Terengganu	4°45’N, 103°24’E	3.00
MSt_42378	Pekan, Pahang	3°29’N, 103°20’E	4.00
MSt_48657	Kuantan, Pahang	3°46’N, 103°13’E	15.20
MSt_48618	Kuala Terengganu, Terengganu	5°23’N, 103°06’E	5.20
MSt_48615	Kota Bharu, Kelantan	6°10’N, 102°18’E	4.40
MSt_48665	Melaka	2°16’N, 102°15’E	8.50
MSt_42320	Bentong, Pahang	3°31’N, 101°54’E	97.00
MSt_48650	Sepang, Selangor	2°44’N, 101°42’E	16.10
MSt_48647	Subang, Selangor	3°08’N, 101°33’E	16.60
MSt_43427	Tapah, Perak	4°12’N, 101°15’E	35.10
MSt_48625	Ipoh, Perak	4°34’N, 101°06’E	40.10
MSt_41545	Baling, Kedah	5°40’N, 100°55’E	52.00
MSt_48603	Alor Setar, Kedah	6°12’N, 100°24’E	3.90

probability in the PARs distribution and mapped to the corresponding quantile of the inverse PARg distribution:

$$PAR_{s,c,j,d} = F^{-1}_{PAR_{g,j}}\left[F_{PAR_{s,j}}\left(PAR_{s,j,d}\right)\right]$$

where $PAR_{s,c,j,d}$ is the QM-corrected satellite-derived PAR for station j on day d, $PAR_{s,j,d}$ is the corresponding uncorrected satellite-derived PAR, $F_{PAR_{s,j}}$ is the station-specific empirical cumulative distribution function of PARs, and $F_{PAR_{s,j}}$ is the inverse empirical cumulative distribution function of PARg. No separate calibration and validation subsets were defined at the QM stage; all available paired daily records were used to establish the station-specific mapping.

DENOISING METHODS

To improve the quality of satellite-derived PAR (PARs) and minimise unwanted fluctuations, a range of denoising techniques representing decomposition-, filtering-, smoothing-, and learning-based approaches were evaluated (Table 2). The methods were selected to examine different strategies for reducing noise while retaining the temporal characteristics required for reliable prediction. This consideration is particularly relevant under humid tropical conditions, where persistent cloud cover and atmospheric variability may introduce substantial uncertainty into satellite-derived PAR observations (Gould et al. 2019).

Empirical Mode Decomposition (EMD) and Ensemble Empirical Mode Decomposition (EEMD) were employed as decomposition-based approaches. These methods decompose the original signal into a series of intrinsic mode functions, allowing noise-related fluctuations to be separated from meaningful temporal variability (Huang et al. 1998; Wu & Huang 2009). Owing to their adaptive nature, EMD and EEMD are widely applied to nonlinear and non-stationary environmental time-series data, including atmospheric and remote sensing observations (Zhang et al. 2022).

Several filtering-based techniques were also evaluated, including Fast Fourier Transform (FFT), Kalman filtering, wavelet transform, and Wiener filtering. Mean and median filters were additionally applied as smoothing approaches to reduce short-term irregular fluctuations while maintaining the overall temporal behaviour of the PAR series. FFT has been widely used to suppress high-frequency noise in remotely sensed time-series datasets, whereas wavelet-based approaches enable localised noise removal across multiple temporal scales (Cooley & Tukey 1965; Mallat 1989). Kalman and

TABLE 2. Summary of denoising methods evaluated in this study

Group	Denoising Method	Main Function	References
Smoothing-based	Mean Filtering	Reduces short-term fluctuations through averaging	Hird & McDermid (2009)
	Median Filtering	Reduces irregular fluctuations and minimize anomalous observations	Hird & McDermid (2009)
Filtering-based	Fast Fourier Transform (FFT)	Suppresses high-frequency signal noise in the frequency domain	Cooley & Tukey (1965)
	Kalman Filtering	Performs adaptative recursive noise filtering	Kalman (1960)
	Wavelet Transform	Reduces noise while preserving temporal signal characterictics	Mallat (1989)
	Wiener Filtering	Minimizes noise based on statistical signal estimation	Wiener (1949)
Decomposition-based	Emperical Mode Decomposition (EMD)	Decomposes nonlinear and nonstationary signals into intrinsic components	Huang et al. (1998)
	Ensemble Emperical Mode Decomposition (EEMD)	Improves decomposition stability by reducing mode mixing	Wu & Huang (2009)
Learning-based	Autoencoder (AE)	Reconstructus signal structure while reducing noise contamination	Vincent et al. (2010)
	Long Short-Term Memory (LSTM)	Captures temporal dependencies and reconstructs nonlinear signal patterns	Hochreiter & Schmidhuber (1997)

Wiener filters were included because of their effectiveness in estimating underlying signals from noisy observations through recursive and statistical filtering procedures (Kalman 1960; Wiener 1949).

To represent learning-based denoising strategies, Autoencoder (AE) and Long Short-Term Memory (LSTM) models were implemented. These approaches are capable of learning complex nonlinear patterns and temporal dependencies within environmental signals, enabling reconstruction of the underlying signal while suppressing noise contamination (Hochreiter & Schmidhuber 1997; Vincent et al. 2010). Similar learning-based denoising approaches have been successfully applied to environmental time-series prediction problems, including air pollution forecasting under conditions of missing and noisy observations (Samal, Babu & Das 2021).

The principal parameter settings used for each denoising method are summarised in Table 2. Mean, median, FFT, Kalman, wavelet, Wiener, EMD, and EEMD were applied separately to each station series, whereas AE and LSTM processed all fourteen stations series jointly. The resulting datasets were evaluated using the same signal-quality and regression framework.

REGRESSION MODELS

To examine whether denoising improved predictive performance, regression models were applied to both the baseline and denoised PAR datasets. The baseline analysis used QM-corrected satellite-derived PAR before denoising, while the post-denoising analysis used the corresponding datasets after noise reduction. This comparison allowed the influence of denoising on PAR prediction reliability to be assessed more clearly.

Five regression models were tested: Linear Regression (LR), Random Forest (RF), Gradient Boosting (GB), Neural Network (NN), and Bagging Regressor (BR). These models were selected to represent different levels of model complexity, from a simple linear approach to ensemble-based and neural network methods commonly used in environmental prediction studies (Breiman 1996, 2001; Friedman 2001; Goodfellow, Bengio & Courville 2016).

For each station, the paired PARs and PARg observations were randomly divided into training (80%) and testing (20%) subsets using a fixed random state of 42. The same splitting procedure was applied to the baseline and all denoised datasets to ensure comparability. Predictor standardisation was fitted using the training PARs data and then applied to the test data, while PARg remained on its original processed scale. Model performance was evaluated exclusively on the test subset using MSE and R^2. A random split was used because the analysis assessed correspondence across the observed range of tropical conditions rather than future-time forecasting.

PERFORMANCE EVALUATION

The denoising methods were evaluated using two aspects: signal quality and predictive performance. Signal quality was assessed using signal-to-noise ratio (SNR), standard deviation (STD), and mean absolute deviation (MAD). These metrics were used to examine how well each method reduced residual noise while retaining the main temporal pattern of the PAR signal. Predictive performance was evaluated using the coefficient of determination (R^2) and mean squared error (MSE). The metrics were calculated for both the baseline and denoised

TABLE 3. Evaluation metrics used for signal quality and predictive performance assessment

Metric Category	Metric	Formula	Description
Signal Quality	Signal-to-Noise Ratio (SNR)	$SNR = 10\log_{10}\left(\frac{\sum_{i=1}^{n} y_i^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\right)$	Measures the strength of the signal relative to residual noise
Signal Quality	Standard Deviation (STD)	$STD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$	Measures the dispersion of signal values
Signal Quality	Mean Absolute Deviation (MAD)	$MAD = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i - \bar{x}\rvert$	Measures the average absolute deviation of signal values from the mean
Predictive Performance	Mean Squared Error (MSE)	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Measures the average squared prediction error
Predictive Performance	Coefficient of Determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	Measures the proportion of variance explained by the model

datasets to compare model performance before and after denoising. Higher R^2 and SNR values and lower MSE were interpreted as improved performance, whereas changes in STD and MAD were considered together with temporal-pattern retention to avoid treating variance compression as improved signal quality (Chicco, Warrens & Jurman 2021).

where y_i represents the observed value; $\widehat{y_i}$ represents the predicted value; $\bar{y}$ represents the mean observed value; x_i represents the signal value; $\overline{x}$ represents the mean signal value; and n represents the total number of observations.

RESULTS AND DISCUSSION

CHARACTERISTICS OF QM-CORRECTED PAR SIGNALS

Figure 3 presents the mean QM-corrected PAR values across the 14 meteorological stations in Peninsular Malaysia. The corrected PAR values ranged from 5.09 MJ m^{-2} day^{-1} at MSt_47117 to 8.63 MJ m^{-2} day^{-1} at MSt_43427, with an overall mean of 7.31 MJ m^{-2} day^{-1}. Most stations exhibited values between 7.12 and 7.88 MJ m^{-2} day^{-1}, indicating relatively consistent PAR availability across the study area after QM correction.

Despite the relatively narrow range of corrected PAR values, several stations exhibited noticeably higher or lower PAR availability than the overall average. Stations such as MSt_43427 and MSt_42378 recorded comparatively higher mean values of 8.63 and 8.04 MJ m^{-2} day^{-1}, respectively, whereas MSt_47117, MSt_42320 and MSt_41545 showed lower PAR levels of 5.09, 6.66, and 6.96 MJ m^{-2} day^{-1}, respectively. These differences may reflect local variations in cloud persistence, rainfall frequency, and atmospheric conditions that influence the amount of solar radiation reaching the surface (Gould et al. 2019; Wang et al. 2020). Similar spatial and temporal variability in PAR availability has been reported in previous studies, highlighting the influence of environmental and atmospheric conditions on PAR distribution (Akitsu et al. 2022; Zhao et al. 2023). The retention of these station-level differences suggests that QM correction reduced systematic bias without excessively altering the underlying spatial characteristics of the PAR signal.

The corrected dataset served as the baseline input for subsequent denoising analyses. Maintaining the spatial variability observed among stations is important because excessive smoothing or signal modification may reduce the representativeness of local PAR conditions and subsequently influence predictive performance. Therefore, the denoising methods evaluated in this study were expected to reduce residual noise while preserving the essential temporal and spatial characteristics of the corrected PAR signal.

NOISE CHARACTERISTICS OF PARg AND PARs DATASETS

Table 4 summarises the noise characteristics of the ground-derived reference PAR (PARg) and QM-corrected satellite-derived PAR (PARs) datasets across the 14 meteorological stations. Differences

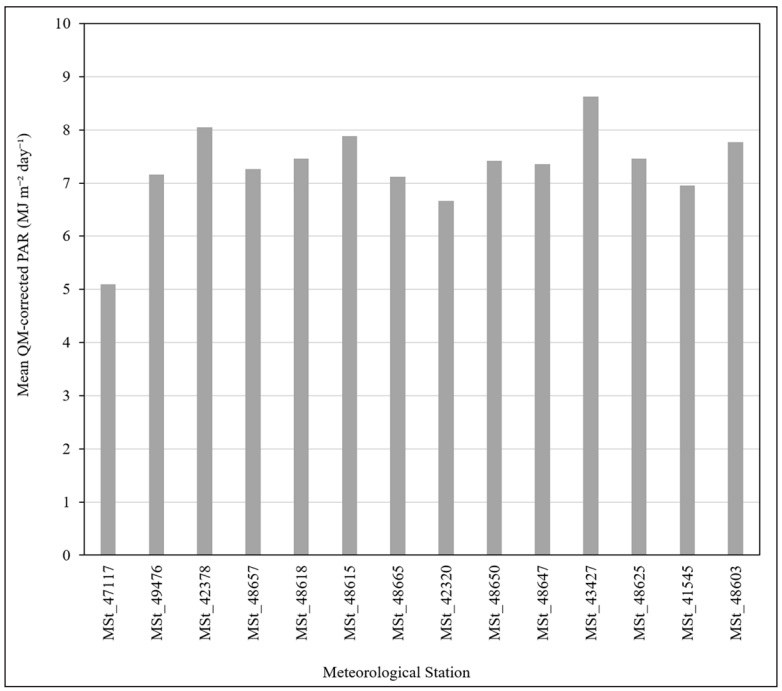


FIGURE 3. Spatial variability of mean QM-corrected PAR across the selected stations in Peninsular Malaysia

between the two datasets were evident for all three noise indicators, namely standard deviation (STD), signal-to-noise ratio (SNR), and mean absolute deviation (MAD).

In general, PARs exhibited higher STD and MAD values than PARg at most stations. The STD values of PARs ranged from approximately 1.233 to 2.664 MJ m^{-2} day^{-1}, whereas PARg ranged from 0.472 to 2.326 MJ m^{-2} day^{-1}. Similarly, MAD values were generally higher in PARs than in PARg, indicating greater short-term fluctuations and larger deviations from the central tendency of the signal. These findings suggest that residual variability remained present within the corrected satellite-derived dataset despite the application of QM correction. This is consistent with previous work showing that bias correction can reduce systematic and distributional errors, although short-term noise or local signal instability may remain in environmental time-series data (Bailey et al. 2025).

The SNR values showed a similar station-to-station variation. The highest SNR for PARg was observed at MSt_42320 (14.259 dB), whereas the lowest was recorded at MSt_48618 (3.098 dB). For PARs, SNR values were generally lower, ranging from 2.729 dB to 5.468 dB, indicating weaker signal quality compared with the reference dataset. The lower SNR in PARs implies that part of the satellite-derived signal remained affected by noise-related variability, which may reduce the stability of subsequent prediction models. Similar challenges associated with noise contamination in environmental and satellite-derived datasets have been reported in previous studies, highlighting the importance of signal enhancement prior to subsequent analysis and prediction (Barth et al. 2024; Finazzi, Rodeschini & Tedesco 2025).

The persistence of higher variability and lower signal quality in PARs highlights the potential benefit of additional signal processing prior to predictive modelling. In this context, denoising is not intended to remove all variability from the PAR signal, but to reduce irregular fluctuations while preserving meaningful temporal behaviour. This balance is particularly important for environmental time-series data because excessive smoothing may obscure meaningful signal characteristics and reduce the information available for subsequent interpretation and prediction (Mariethoz et al. 2015). Consequently, several denoising approaches were evaluated in the subsequent analysis to determine whether residual fluctuations could be reduced while preserving the essential characteristics of the PAR signal.

INFLUENCE OF DENOISING ON SIGNAL QUALITY

Figure 4 compares the monthly mean patterns of the original QM-corrected PAR series with the median- and LSTM-denoised series. Median filtering retained the broader seasonal progression of the original signal while reducing local fluctuations. In contrast, the LSTM reconstruction remained almost constant throughout the year and did not reproduce the seasonal amplitude evident in the original series.

TABLE 4. Noise evaluation metrics for PARg and PARs datasets across the 14 stations

Station	PARg			PARs		
	STD (MJ m^{-2} day^{-1})	MAD (MJ m^{-2} day^{-1})	SNR (dB)	STD (MJ m^{-2} day^{-1})	MAD (MJ m^{-2} day^{-1})	SNR (dB)
MSt_47117	1.237	1.016	3.459	1.567	1.252	2.729
MSt_49476	1.567	0.992	4.807	2.392	1.903	3.149
MSt_42378	1.701	1.078	4.662	2.664	2.111	2.977
MSt_48657	1.779	1.439	4.080	2.325	1.829	3.121
MSt_48618	2.326	2.002	3.098	2.382	1.916	3.025
MSt_48615	2.084	1.683	3.752	2.400	1.902	3.259
MSt_48665	1.775	1.446	4.188	1.909	1.513	3.894
MSt_42320	0.472	0.128	14.259	1.476	1.172	4.564
MSt_48650	1.918	1.567	3.865	1.936	1.543	3.829
MSt_48647	1.833	1.465	3.950	1.618	1.286	4.477
MSt_43427	1.772	1.392	4.724	1.619	1.282	5.171
MSt_48625	1.514	1.229	4.899	1.573	1.252	4.716
MSt_41545	1.019	0.508	6.617	1.233	0.970	5.468
MSt_48603	1.855	1.525	4.176	1.823	1.455	4.250

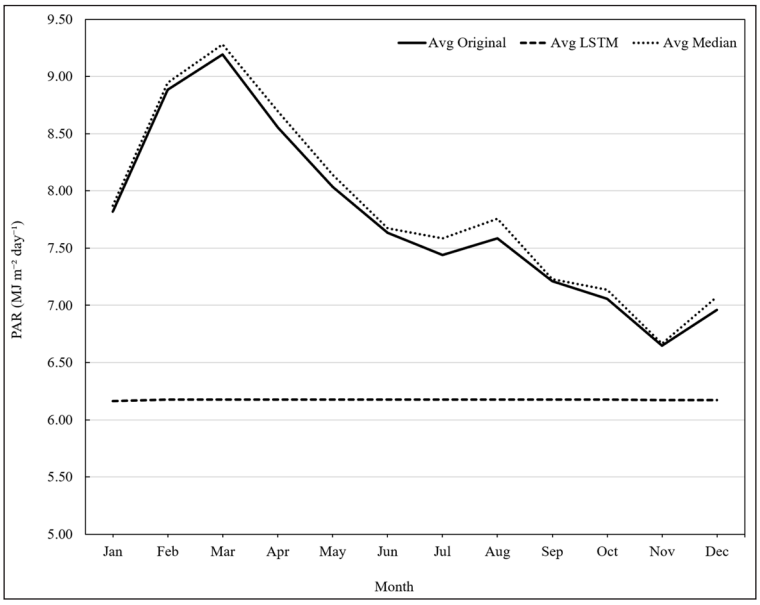


FIGURE 4. Monthly mean patterns of the original QM-corrected satellite-derived PAR and the corresponding median- and LSTM-denoised series averaged acros fourten stations during 2015 to 2023

The nearly flat LSTM profile indicates substantial compression of temporal variance rather than improved signal preservation. This explains why it is very low MSE should not be interpreted independently of R^2 and the reconstructed temporal pattern. Median filtering provided a more balanced result because it reduced short-term irregularities without removing the seasonal variation relevant to PAR dynamics (Akitsu et al. 2022).

COMPARATIVE PERFORMANCE OF DENOISING APPROACHES

Table 5 summarises the predictive performance of the baseline and denoised datasets, while Figures 5 and 6 compare the corresponding mean R^2 and MSE values, respectively. For the baseline dataset, Linear Regression was the best-performing model, with a mean R^2 of 0.207 and a mean MSE of 2.208. Most denoising methods improved upon this baseline, although the magnitude of improvement varied among methods. Mean R^2 ranged from 0.131 to 0.748, while mean MSE ranged from 0.004 to 1.616 when the exceptionally large EMD error was excluded. EMD was excluded from Figure 6 because its exceptionally large mean MSE (9299.660) would compress the graphical scale and obscure differences among the remaining methods; the result is reported in Table 5. These results indicate that effective denoising depended not only on noise suppression but also on the retention of temporal characteristics relevant to PAR prediction (Mariethoz et al. 2015).

Among the evaluated approaches, median filtering produced the strongest overall performance. The median-denoised dataset achieved the highest mean R^2 value (0.748) and one of the lowest mean MSE values (0.293), with Random Forest identified as the best-performing regression model (Table 5). Compared with the baseline dataset, the application of median filtering substantially improved predictive reliability while reducing prediction error. The superior performance of the median filter may be attributed to its ability to suppress local irregular fluctuations without altering the dominant temporal structure of the PAR signal, which is consistent with previous remote sensing time-series studies showing that denoising can reduce noise while preserving meaningful temporal characteristics (Hird & McDermid 2009; Priyadarshi et al. 2018). Such behaviour is particularly advantageous in humid tropical environments, where short-term atmospheric variability associated with cloud cover and weather fluctuations can introduce noise into satellite-derived observations (Gould et al. 2019).

Learning-based denoising also improved predictive performance, with mean R^2 values of 0.542 for LSTM and 0.538 for AE. LSTM yielded the lowest numerical mean MSE (0.004), whereas AE produced a mean MSE of 0.458. However, the reconstructed LSTM series showed pronounced variance compression and an almost flat temporal profile (Figure 4). Its small prediction error therefore partly reflected the restricted amplitude of the reconstructed series rather than faithful preservation of PAR variability. This interpretation is consistent with its lower mean R^2 relative to median filtering (0.542 versus 0.748). Accordingly, LSTM was not considered the most reliable overall method despite its lower numerical MSE.

Intermediate performance was observed for mean filtering, FFT, Kalman filtering, Wiener filtering, wavelet transformation, and EEMD. These methods generally

produced mean R^2 values ranging from approximately 0.23 to 0.37, while reducing prediction errors relative to the baseline dataset. Similar improvements in satellite-derived time-series quality following denoising have been reported for MODIS-based remote sensing products (Lu, Liu & Liang 2007; Priyadarshi et al. 2018). The results indicate that successful denoising requires a balance between noise suppression and retention of environmentally meaningful temporal information (Mariethoz et al. 2015). Excessive smoothing or inappropriate signal decomposition may reduce fluctuations but can also weaken signal characteristics that contribute to predictive performance.

In contrast, EMD produced the weakest overall performance, with a mean R^2 of 0.131 and a mean MSE of 9299.660. Inspection of the denoised outputs traced the exceptionally large error to reconstruction instability at several stations, particularly MSt_42320, MSt_42378, MSt_41545 and MSt_47117. At these stations, removal of the first intrinsic mode function generated unrealistic negative or excessively large reconstructed PARg values, which substantially inflated the prediction error. The poor EMD performance therefore reflected station-specific decomposition and reconstruction failure rather than a consistent scaling difference across the dataset. The contrasting results are relevant to humid tropical environments, where frequent cloud cover and rapid atmospheric fluctuations can introduce short-lived irregularities into satellite-derived PAR observations (Gould et al. 2019). Median filtering was well suited to this signal structure because it reduced isolated fluctuations while retaining the broader temporal pattern.

TABLE 5. Comparative predictive performance of denoising methods and their best-performing regression models

Method	Best model	Mean MSE	Mean R^2
Baseline	Linear regression	2.208	0.207
Mean	Neural network	0.597	0.372
Median	Random forest	0.293	0.748
Wiener	Neural network	1.616	0.231
Wavelet	Neural network	0.978	0.226
FFT	Linear regression	1.051	0.303
Kalman	Neural network	1.134	0.262
EMD	Neural network	9299.660	0.131
EEMD	Linear regression	1.586	0.227
AE	Linear regression	0.458	0.538
LSTM	Linear regression	0.004	0.542

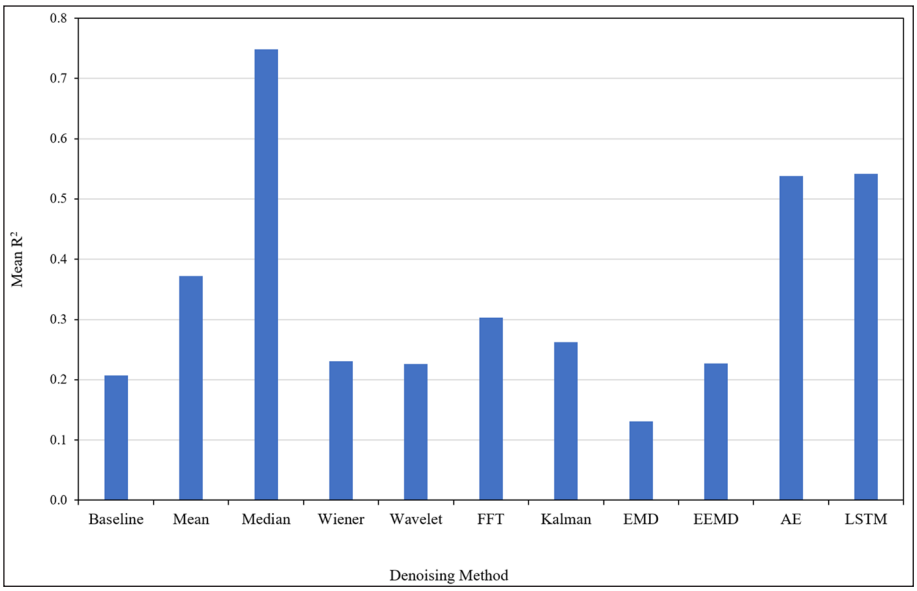


FIGURE 5. Comparison of mean R^2 values across different denoising methods

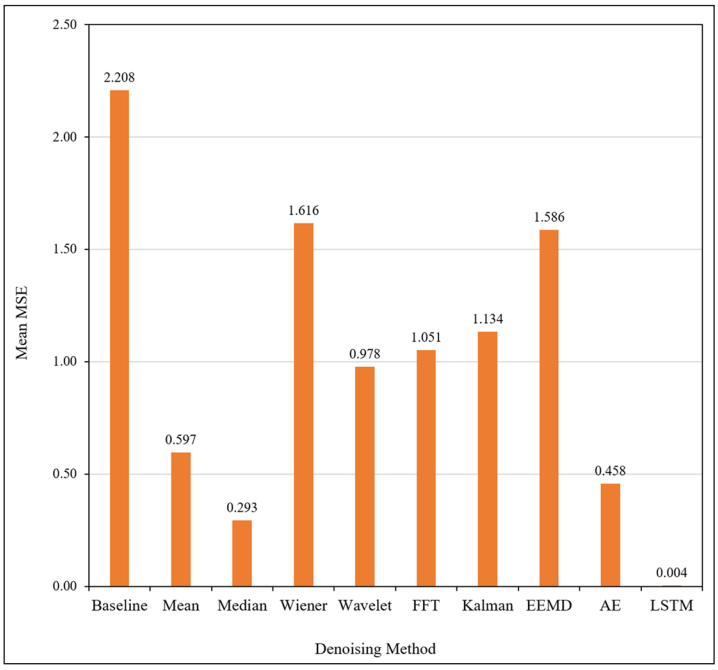


FIGURE 6. Comparison of mean MSE values across denoising methods, excluding EMD due to substantially larger error values

In comparison, decomposition-based EMD became unstable at several stations, whereas LSTM excessively compressed the temporal variation. Thus, increasing algorithmic complexity did not necessarily improve reliability under the conditions represented by this dataset. As the atmospheric drivers of individual fluctuations were not tested directly, these links should be interpreted as plausible signal-based explanations rather than causal relationships.

CONCLUSION

This study compared ten denoising methods for improving the predictive reliability of QM-corrected satellite-derived PAR across fourteen stations in Peninsular Malaysia. Most methods improved upon the baseline QM-corrected dataset, but their effectiveness depended on preserving meaningful temporal variation. Median filtering combined with Random Forest produced the highest mean R^2 (0.748) with a mean MSE of 0.293 and retained the broader temporal pattern of the original signal. Although LSTM yielded a lower numerical MSE (0.004), its nearly flat reconstruction indicated substantial variance compression. EMD performed least reliably because reconstruction instability at several stations produced exceptionally large errors.

These findings show that a relatively simple denoising method can provide more reliable PAR prediction than computationally intensive alternatives when it balances noise reduction with signal preservation. Nevertheless, PARg was estimated from measured global solar radiation using a regional conversion coefficient rather than obtained through direct PAR measurements, and the analysis was limited to fourteen stations in Peninsular Malaysia. Future studies should validate these findings using direct PAR observations and a wider range of tropical environments.

ACKNOWLEDGEMENTS

We gratefully acknowledge the Malaysian Meteorological Department (MetMalaysia) for providing ground-based solar radiation datasets that were essential for this study.

REFERENCES

Akitsu, T.K., Nasahara, K.N., Ijima, O., Hirose, Y., Ide, R., Takagi, K. & Kume, A. 2022. The variability and seasonality in the ratio of photosynthetically active radiation to solar radiation: A simple empirical model of the ratio. *International Journal of Applied Earth Observation and Geoinformation* 108: 102724. https://doi.org/10.1016/j.jag.2022.102724

Bailey, M.D., Nychka, D.W., Sengupta, M., Yang, J., Xie, Y., Habte, A. & Bandyopadhyay, S. 2025. Adapting quantile mapping to bias correct solar radiation data. *Solar Energy* 288: 113220. https://doi.org/10.1016/j.solener.2024.113220

Barth, A., Brajard, J., Alvera-Azcárate, A., Mohamed, B., Troupin, C. & Beckers, J-M. 2024. Ensemble reconstruction of missing satellite data using a denoising diffusion model: Application to chlorophyll a concentration in the Black Sea. *Ocean Science* 20: 1567-1584. https://doi.org/10.5194/os-20-1567-2024

Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5-32. https://doi.org/10.1023/A:1010933404324

Breiman, L. 1996. Bagging predictors. *Machine Learning* 24(2): 123-140. https://doi.org/10.1007/BF00058655

Chicco, D., Warrens, M.J. & Jurman, G. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science* 7: e623. https://doi.org/10.7717/peerj-cs.623

Cooley, J.W. & Tukey, J.W. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation* 19(90): 297-301. https://doi.org/10.2307/2003354

Finazzi, F., Rodeschini, J. & Tedesco, L. 2025. Discussion on assessing predictability of environmental time series with statistical and machine learning models. *Environmetrics* 36: e2900. https://doi.org/10.1002/env.2900

Friedman, J.H. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29: 1189-1232. https://doi.org/10.1214/aos/1013203451

Goodfellow, I., Bengio, Y. & Courville, A. 2016. *Deep Learning*. Cambridge: MIT Press.

Gould, R.W., Ko, D.S., Ladner, S.D., Lawson, T.A. & Macdonald, C.P. 2019. Comparison of satellite, model, and *in situ* values of photosynthetically available radiation (PAR). *Journal of Atmospheric and Oceanic Technology* 36(4): 535-555. https://doi.org/10.1175/JTECH-D-18-0096.1

He, Q., Chun, K.P., Tan, M.L., Dieppois, B., Liew, J., Klaus, J., Fournier, M., Massei, N. & Yetemen, O. 2021. Tropical drought patterns and their linkages to large-scale climate variability over Peninsular Malaysia. *Hydrological Processes* 35(9): e14356. https://doi.org/10.1002/hyp.14356

Hird, J.N. & McDermid, G.J. 2009. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment* 113(1): 248-258. https://doi.org/10.1016/j.rse.2008.09.003

Hochreiter, S. & Schmidhuber, J. 1997. Long short-term memory. *Neural Computation* 9(8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, N.E., Shen, Z., Long, S., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. & Liu, H.H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A* 454(1971): 903-995. https://doi.org/10.1098/rspa.1998.0193

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82(1): 35-45. https://doi.org/10.1115/1.3662552

Katiraie-Boroujerdy, P.S., Rahnamay Naeini, M., Akbari Asanjan, A., Chavoshian, A., Hsu, K.L. & Sorooshian, S. 2020. Bias correction of satellite-based precipitation estimations using quantile mapping approach in different climate regions of Iran. *Remote Sensing* 12(12): 2102. https://doi.org/10.3390/rs12132102

Kumari, S. & Singh, S.K. 2023. Machine learning-based time series models for effective CO_2 emission prediction in India. *Environmental Science and Pollution Research* 30: 116601-116616. https://doi.org/10.1007/s11356-022-21723-8

Liu, J., Cai, Y., Pei, X. & Yu, X. 2025. Advances in research and application of techniques for measuring photosynthetically active radiation. *Remote Sensing* 17(10): 1765. https://doi.org/10.3390/rs17101765

Lu, X., Liu, R. & Liang, S. 2007. Removal of noise by wavelet method to generate high quality temporal data of terrestrial MODIS products. *Photogrammetric Engineering & Remote Sensing* 73(10): 1129-1139.

Mallat, S.G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7): 674-693. https://doi.org/10.1109/34.192463

Mariethoz, G., Linde, N., Jougnot, D. & Rezaee, H. 2015. Feature-preserving interpolation and filtering of environmental time series. *Environmental Modelling & Software* 72: 71-76. https://doi.org/10.1016/j.envsoft.2015.07.001

McCree, K.J. 1972. Action spectrum, absorptance and quantum yield of photosynthesis in crop plants. *Agricultural Meteorology* 9: 191-216. https://doi.org/10.1016/0002-1571(71)90022-7

Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sicre, C. & Dedieu, G. 2017. Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sensing* 9(2): 173. https://doi.org/10.3390/rs9020173

Priyadarshi, N., Chowdary, V.M., Srivastava, Y.K., Das, I.C. & Jha, C.S. 2018. Reconstruction of time series MODIS EVI data using de-noising algorithms. *Geocarto International* 33(10): 1095-1113. https://doi.org/10.1080/10106049.2017.1333535

Samal, K.K.R., Babu, K.S. & Das, S.K. 2021. Temporal convolutional denoising autoencoder network for air pollution prediction with missing values. *Urban Climate* 38: 100872. https://doi.org/10.1016/j.uclim.2021.100872

Shavalipour, A., Hakemzadeh, M.H., Sopian, K., Mohamed Haris, S. & Zaidi, S.H. 2013. New formulation for the estimation of monthly average daily solar irradiation for the tropics: A case study of Peninsular Malaysia. *International Journal of Photoenergy* 2013: 174671. https://doi.org/10.1155/2013/174671

Singh, R.K., Vader, A., Mundy, C.J., Søreide, J.E., Iken, K., Dunton, K.H., Castro de la Guardia, L., Sejr, M.K. & Bélanger, S. 2022. Satellite-derived photosynthetically available radiation at the coastal Arctic seafloor. *Remote Sensing* 14(20): 5180. https://doi.org/10.3390/rs14205180

Tan, M.L., Juneng, L., Tangang, F.T., Chung, J.X. & Radin Firdaus, R.B. 2021. Changes in temperature extremes and their relationship with ENSO in Malaysia from 1985 to 2018. *International Journal of Climatology* 41(S1): E2564-E2580. https://doi.org/10.1002/joc.6864

Tan, P.Y. & Ismail, M.R. 2016. Photosynthetically active radiation and comparison of methods for its estimation in equatorial Singapore. *Theoretical and Applied Climatology* 123(3-4): 873-883. https://doi.org/10.1007/s00704-015-1399-9

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. & Manzagol, P.A. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research* 11: 3371-3408.

Wang, D., Liang, S., Zhang, Y., Gao, X., Brown, M.G.L. & Jia, A. 2020. A new set of MODIS land products (MCD18): Downward shortwave radiation and photosynthetically active radiation. *Remote Sensing* 12(1): 168. https://doi.org/10.3390/RS12010168

Wiener, N. 1949. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: John Wiley & Sons.

Wu, Z. & Huang, N.E. 2009. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis* 1(1): 1-41. https://doi.org/10.1142/S1793536909000047

Zhang, X., Yang, Y., Jin, M., Lujie, Z. & Zijian, D. 2022. Rayleigh lidar signal denoising method combined with WT, EEMD and LOWESS to improve retrieval accuracy. *Remote Sensing* 14(23): 6175. https://doi.org/10.3390/rs14143270

Zhao, P., Bai, Y., Zhang, Z., Wang, L., Guo, J. & Wang, J. 2023. Differences in diffuse photosynthetically active radiation effects on cropland light use efficiency calculated via contemporary remote sensing and crop production models. *Ecological Informatics* 73: 101948. https://doi.org/10.1016/j.ecoinf.2022.101948