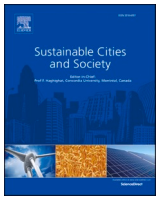
Sustainable Cities and Society 99 (2023) 104964

Contents lists available at ScienceDirect

Sustainable Cities and Society

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



Unravelling the effects of dynamic urban thermal environment on utility-scale floating photovoltaic electricity generation

Ziyi Huang [a], Rui Zhu [b,*], Linlin You [c]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, 181 Chatham Road South, Kowloon, Hong Kong, China
[b] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632, Republic of Singapore
[c] School of Intelligent Systems Engineering, Sun Yat-Sen University, Shenzhen 518107, China

ARTICLE INFO

Keywords:
Photovoltaic conversion efficiency
Decarbonization
Urban thermal environment
Machine learning
GIScience
Remote sensing

ABSTRACT

The utility-scale photovoltaic systems have been widely integrated for sustainable urban development. However, the varying photovoltaic conversion efficiency (PVCE) affected by dynamic urban thermal environment causes uncontrollable uncertainty in electricity generation, which challenges installed-capacity planning and load-balancing operations. To tackle this problem, this study develops a PVCE estimation model containing four hierarchical modules. First, photovoltaic surface temperatures (PVSTs) are retrieved from satellite imagery, and meteorological features are collected to represent the dynamic thermal environment. Second, the contribution of each feature to the PVST estimation is evaluated, referring to the Mean Decrease in Impurity, Permutation Importance, and SHapley Addictive exPlanations. Third, machine learning models using Support Vector Machine, Random Forest, and XGBoost are developed to establish robust regressions between PVSTs and the selected features, enabling an accurate estimation of PVST spatio-temporal heterogeneity. Finally, the spatio-temporally corresponding PVCEs are calculated, resulting in a refined estimation of annual electricity generation. The investigation of four floating PV systems in Singapore found that their PVCEs vary insignificantly throughout the year, which is probably because of its stable climate and the cooling effect of water. The proposed model is simple and effective, demonstrating its impact on PV potential estimation when the urban thermal effect becomes significant.

1. Introduction

1.1. Background and motivation

The International Energy Agency has suggested that solar photovoltaic (PV) will become the largest energy resource by 2027, with an increase of about 1000 Gigawatts between 2023 and 2027 (Bahar & Analyst, 2023). Accurately estimating the utility-scale or distributed PV potential is crucial for effective PV deployment planning and evaluating the electricity generation of installed PV systems (Zhang et al., 2019; Zhu et al., 2023a). Previous studies focused on extracting large-scale PV areas from satellite imagery (Li et al., 2021; Wang et al., 2018; Zhu et al., 2023b), constructing three-dimensional building models with solar distribution information (Yan et al., 2023; Zhu et al., 2019), and exploring the technical routes for PV deployment in the transition to solar cities (Wong et al., 2016; Zhong et al., 2021), which has built a scientific foundation for PV potential estimation. It has also been suggested that the thermal environment affects PV conversion efficiency (PVCE) significantly (Dubey et al., 2013). However, obtaining PVCE requires the measurement of the electrical characteristics and solar radiation, which is impractical and costly when applied to large-scale PV systems. Due to this reason, dynamic PVCE has rarely been incorporated into the estimation models. Since the urban heat island effect creates an altered thermal environment characterized by higher land surface temperature and increased thermal mass (Ji et al., 2014; McCarthy et al., 2010), it is imperative to estimate the effects of the dynamic thermal environment on the PVCE.

To accurately quantify the PV electricity generation, a feasible approach is building regressions between thermal-related features (e.g., PV surface temperatures (PVSTs) ambient temperature, wind speed, and relative humidity) and the PVCE (Kim et al., 2019). However, it cannot be applied to the newly constructed PV farms due to the lack of historical

* Corresponding author.
E-mail address: zhur@ihpc.a-star.edu.sg (R. Zhu).

https://doi.org/10.1016/j.scs.2023.104964
Received 27 June 2023; Received in revised form 29 August 2023; Accepted 25 September 2023
Available online 26 September 2023
2210-6707/© 2023 Elsevier Ltd. All rights reserved.

PVCE data. Previous studies assumed a constant PVCE to estimate PV potential (Morales Pedraza, 2022), which ignored the fact that the dynamic PVCE is influenced by the thermal environment (Skoplaki & Palyvos, 2009). Therefore, without considering the dynamic effect of the thermal environment on PVCE may cause a large deviation in PV potential estimation compared to reality. Particularly, the estimated PV electricity production may be significantly smaller than expected based on a designated installed capacity, especially in cities where the temperatures are constantly hot (e.g., Singapore) or where there are big changes in the thermal environment over time and space (e.g., Dubai). Consequently, inaccurate electricity production estimation leads to unsatisfied operating income and incorrect environmental benefits prediction such as greenhouse gas emission reduction for installed utility-scale PV systems. Although the dynamic thermal environment can be monitored by high-quality instruments and sensors in the floating PV testbed, few studies estimated the effects of the thermal environment on PVCE for installed utility-scale PV systems, which usually cover large areas. To tackle this challenge, this study aims to develop a simple and effective method to accurately estimate PV electricity generation under a dynamic thermal environment.

1.2. PVST estimation models

Over the past few decades, studies have proposed many methods to estimate the relationship between the thermal environment and PVST. The most widely used approach is to determine empirical correlations by remaining other thermal environment features as a constant under the design-specific and simplified conditions (Skoplaki & Palyvos, 2009). However, this method cannot be applied to the installed large-scale floating PV systems due to the complex interaction of the thermal environment. Some other studies focus on the electrical equivalent circuits. Since the PVST dependence of the electrical characteristics on short-circuit current and open-circuit voltage of specific PV modules can be obtained after experiments, the PVST can be derived by the corresponding current and voltage under different thermal environments (Motahhir et al., 2018; Rashidi et al., 2021). However, this method is time and cost intensive, e.g., deploying sensors to obtain the current and voltage for PVST derivation. Another approach assumes that the PV module is a simplified multilayer model and estimates the PVST based on the energy balance model (Akhsassi et al., 2018). For instance, the finite element and finite difference methods simplify the heat transfer process between the PV modules and the ambient environment and between different layers of the PV modules (Aly et al., 2018a; 2018b), while the computational fluid dynamics method accounts for both the heat transfer and fluid flow to estimate the PVST (Kim & Nam, 2019). However, the simplified multilayer model is only suitable for the specific PV modules and this method cannot be generalized to other types of PV modules or different regions.

Recently, some studies estimated PVSTs affected by the thermal environment by developing appropriate approximation models based on artificial intelligence techniques. A study utilized the neural network to estimate the PVST from the measured PVST and solar radiation with an accuracy of 96 % (Hegazy et al., 2019); however, the accuracy highly depends on the measured PVST, which takes a long time to acquire for installed large-scale floating PV systems. Another study used ambient temperature, solar radiation, relative humidity, wind speed, and electricity generation to build the PVST estimation models using the Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Regression Tree Ensemble. While, the result showed that SVM has the smallest root mean square error of 3.2 °C in the testing dataset (May May Tzuc et al., 2018), which highly influences the accuracy of the PV electricity generation estimation. Artificial neural network (ANN) has also been developed to estimate the PVST by using the meteorological records of wind speed, humidity, cloud cover, and the measured results of outdoor air temperature, solar radiation, and PVST as input parameters with the R^2 of 0.97 (Jung et al., 2020). However, this method is not able to estimate continuously spatio-temporal distribution of PVSTs. Missing such capability makes it difficult to accurately identify heterogenous PVCE over time and space and hinders the planning of PV installation. To tackle the above problems, this study aims to determine significant influential factors of the utility-scale floating PVSTs and accurately estimate PVSTs by reliably input parameters.

1.3. Estimation of PV electricity generation

Previous studies proposed that (i) the PVCE of the crystalline silicon PV cells will reduce by 0.3 %−0.45 % with a 1 °C increase in the PVST under a standard testing condition (25 °C) (Kaldellis et al., 2014); and (ii) unlike ground-mounted and rooftop PV systems, floating PV systems present competitiveness in improving PVCE, which benefits from the cooling effects of the ambient thermal environment (Dörenkämper et al., 2021). Therefore, the estimation of PV electricity generation requires consideration of the impact of the dynamic thermal environment. Although some studies have attempted to develop the PV electricity generation estimation model utilizing continuous thermal-related features based on long short-term memory-based machine learning (ML) methods, these models can only ensure the accuracy of very short-term estimation (e.g., ranging from the next 5 min to next 24 h) because they are highly dependant on the most recent meteorological data (Jung et al., 2020; Rana et al., 2016). However, the continuously frequent measurement is impractical to be applied the utility-installed PV farms. Currently, some studies estimated the PVCE from the thermal environment based on ML. A study proposed ANNs to estimate the PVCE by using ambient temperature, solar radiation, wind speed, and relative humidity as input parameters with the lowest root-mean-square error (RMSE) of 0.017 (Arslan, 2023). However, it is a data-driven method requiring sufficient data to guarantee the model's accuracy. To accurately estimate the PV electricity generation, we can estimate the dynamic PVCE due to the high correlation between the PVST and PVCE (Skoplaki & Palyvos, 2009).

1.4. Land surface temperature retrieval method

Unlike deploying many sensors, the land surface temperature (LST) retrieval method is an economical and effective alternative to observing the PVSTs. Easily accessible remote sensing images cover large areas, which can provide the observation of the entire utility-scale PV areas. We retrieved the LST on the PV surfaces as the training data in the estimation model. Previous studies developed many retrieval methods through one or more thermal bands of remote sensing images (Qin et al., 2001; Yu et al., 2014). Some studies used only one thermal infrared band to retrieve the LST, such as single-channel and mono-window algorithms; however, they rely on accurate upwelling atmospheric radiances (Sekertekin & Bonafoni, 2020) and near-surface air temperature (Qin et al., 2001), respectively. In contrast, some other studies utilize more than one TIR band, such as split window and multi-channel algorithms, to minimize the atmospheric effects (Becker & Li, 1990); however, they are limited by the quantity and quality of TIR bands (Majumder et al., 2021; Yu et al., 2014). To accurately retrieve the LST, we utilized the radiative transfer equation (RTE) algorithm to eliminate the atmospheric effects (Gillespie et al., 1998) in the Landsat series images, which currently have the highest spatial resolution available for free.

1.5. Contribution

This study is innovative in (i) developing machine learning-based methods to estimate the utility-scale floating PVST, (ii) unravelling the effects of dynamic thermal environment on PVST estimation, and (iii) improving the accuracy of the PV electricity generation estimation. Meanwhile, the raw PVST dataset was obtained from open-access satellite images, suggesting a favourable generalization capability to any other regions.

The following sections of this paper are organized as follows. Section 2 develops a machine learning-based PVST estimation method and evaluates the effects of a dynamic thermal environment on the PVST. Section 3 introduces the PV electricity generation estimation. Section 4 exhibits the empirical investigation based on four floating PV farms in Singapore. Section 5 compares the PV electricity generation by using assumed and estimated PVCE. Finally, Section 6 presents the discussion and draws a conclusion.

2. Machine learning-based estimation of PVSTs

This study utilizes the satellite images to retrieve the LSTs on floating PV surfaces as the ground truth as an economical and effective alternative to observing the PVSTs of the installed large-scale floating PV farms. Due to the constraints on temporal resolution and cloud cover of the satellite images, the remote sensing data alone cannot provide continuous LST observations. As a result, this study develops a machine learning-based estimation approach to model the relationship between observed LSTs and thermal-related features and then generates dense estimated PVSTs for further analysis.

This study proposes a research framework to estimate PV electricity generation on large-scale floating PV systems under a dynamic thermal environment in the following steps (Fig. 1). First, the observed LSTs are used as the training data to estimate the PVST. To achieve this, LSTs on floating PV surfaces are retrieved from the thermal bands of remote sensing images. Second, the PVST estimation models are built based on the ML models to accurately estimate PVSTs under a dynamic thermal environment. The estimated PVSTs are validated by using a set of performance evaluation metrics to identify the best estimation model. Third, the effects of the urban thermal environment on the PVST are evaluated based on the best estimation model, and the floating PV electricity generation is better estimated by adapting to the dynamic PVCE.

2.1. Calculating the LSTs on floating PV surfaces

To build the PVST estimation model, we first extract the floating PV areas from the remote sensing images using the Support Vector Machine (SVM) classification. Next, the LSTs on floating PV surfaces are computed by using the RTE (Sobrino et al., 2004). The normalized difference vegetable index (*NDVI*) is calculated in Eq. (1) as follows (Townshend & Justice, 1986):

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

where *NIR* and *R* represent the digital number of near-infrared and red bands, respectively. The proportion of vegetation coverage (P_v) can be described in Eq. (2) as follows,

$$P_v = \frac{NDVI - (NDVI)_s}{(NDVI)_v - (NDVI)_s} \tag{2}$$

where $(NDVI)_s$ and $(NDVI)_v$ determine whether a pixel represents vegetation or non-vegetation by setting a threshold, which are set as 0.05 and 0.7, respectively. This study classified the land surfaces into three types including built-up land, natural surfaces, and waterbody to calculate the emissivity of each pixel Peng et al., 2020). The built-up land includes the infrastructure elements such as roads and buildings, while the natural surfaces encompass natural land surfaces, farmland, and woodland. The pixels with an *NDVI* value less than and equal to 0 are designated as waterbody, while the ones with an *NDVI* value exceeding 0.7 represent natural surfaces. The in-between pixels, i.e., those with an *NDVI* value ranging between 0 and 0.7 are regarded as the built-up land. The emissivity of built-up land (ε_b), natural surfaces (ε_s) and waterbody (ε_w) can be calculated (Eqs. (3)–((5)) as follows,

$$\varepsilon_b = 0.9589 + 0.086P_v - 0.0671P_v^2 \tag{3}$$

$$\varepsilon_s = 0.9625 + 0.0614P_v - 0.0461P_v^2 \tag{4}$$

$$\varepsilon_w = 0.995 \tag{5}$$

$$B(LST) = \frac{DN_{The} - Lu - t(1-\varepsilon)L_d}{t\varepsilon} \tag{6}$$

Eq. (6) describes the blackbody thermal radiance calculation, where *Lu* and L_d correspond to the effective bandpass upwelling and downwelling radiance (W/(m^2 sr μm)), respectively; *t* represents the band average atmospheric transmission; DN_{The} is the digital number in the atmospherically corrected thermal band. After that, *LST* can be calculated in Eq. (7) as follow,

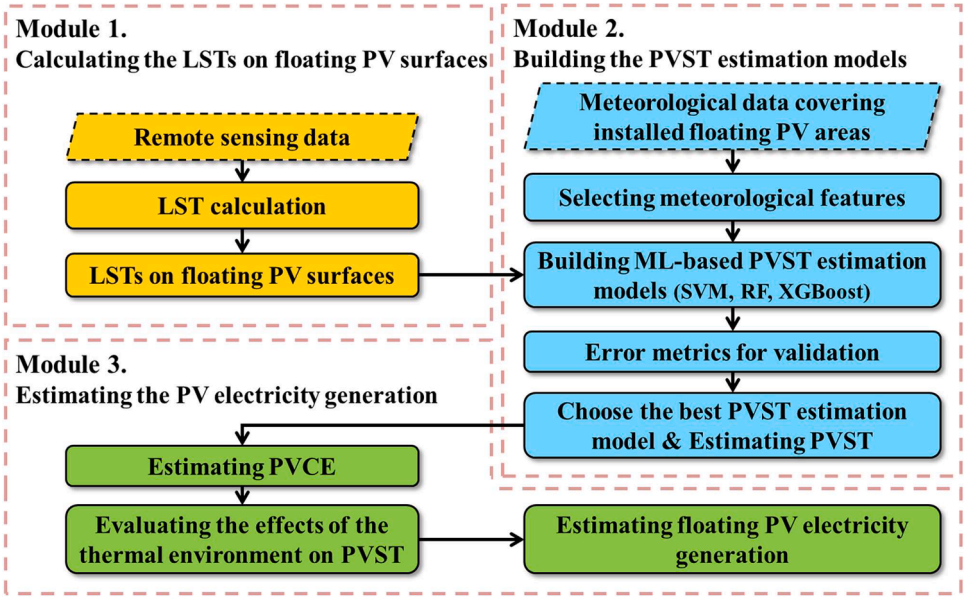


Fig. 1. Research framework built by three interconnected modules.

$$LST = \frac{K_2}{ln\left[\frac{K_1}{B(LST)} + 1\right]} - 273.15 \tag{7}$$

where K_1 (W/(m^2 sr μm)) and K_2 (K) are the constant of the thermal band corresponding to the specific sensors, which can be found in the metadata file of satellite images. Finally, the observed LSTs on floating PV surfaces are extracted by the PV areas to build the PVST estimation models.

2.2. Building the PVST estimation models

2.2.1. Selecting appropriate urban thermal-related features

The thermal environment is defined as the ambient thermal environment on the floating PV areas consisting of several thermal-related features, as summarized in Table 1. Notably, some thermal-related features are obtained based on other features such as heat index temperature and dew point temperature (Camuffo, 2019; National Oceanic & Atmospheric Administration, 2005), which might introduce noise resulting in accuracy reduction. To enhance the model's interpretability, highly correlated features are first removed based on the correlation coefficient. Next, the remaining features are selected by using the predictive power score (PPS) and Boruta to drop unimportant features. PPS evaluates the estimated capacity of a single feature to the PVST estimation using a decision tree-based model, ranging from 0 (no estimated capacity) to 1 (perfect estimation), which can capture the non-linear relationships compared with the correlation coefficient (Ferber, 1956). Boruta creates shadow features by randomly shuffling the real ones to introduce randomness and then uses a random forest classifier to remove the features less important than the best shadow feature after a series of trials, which follows a binominal distribution (Masrur Masrur Ahmed et al., 2021).

2.2.2. Machine learning-based estimation models

To estimate the continuous PVSTs, the supervised ML is utilized to build the PVST estimation models because of the constraints on the temporal resolution of Landsat images resulting in limited observed LSTs. SVM is widely used to estimate the PVSTs (May May Tzuc et al., 2018), which aims to create a function to approximate the relationship between features and the target variable by reflecting the original features in high-dimensional space (Fan et al., 2020) (Eq. (8)) and minimize the distance between the hyperplane and the nearest point from it (Eq. (9)).

$$f(x) = w^T x_i + b \tag{8}$$

$$\begin{aligned} &\boldsymbol{min}\frac{1}{2}ws^2 + C\sum_{i=1}^{n}\beta_i \\ &s.t.\ y_i\left(w^T x_i + b\right) \geq 1 - \beta_i,\ \beta_i \geq 0,\ i = 1,\ 2,\ 3,\ \ldots,\ n \end{aligned} \tag{9}$$

where w and b represent the weight and bias, respectively; C refers to the penalty parameter controlling the degree of punishing the samples whose errors go beyond the given value; n and β_i are the numbers of training samples and the slack variable controlling the insensitive zone used to fit the training dataset, respectively. The number of support vectors and computation complexity reduces with the increasing slack variable value. To map the points to higher dimensional space, we used the kernel of Gaussian radial basis function (RBF), because linear combinations of RBF can approximate almost any function (Anyanwu et al., 2023).

Table 1
Abbreviations of the thermal-related features.

No.	Abbreviation	Features
1	AT	Air temperature in degrees Celsius
2	H	Relative humidity in percentage
3	WS	Wind speed in kilometres per hour
4	V	Visibility in kilometres
5	P	Atmospheric pressure in millibars
6	CC	Cloud cover amount in percentage
7	UVI	Ultraviolet Index
8	R	Precipitation in millimetres
9	HIT	Heat index temperature in degrees Celsius
10	DPT	Dew point temperature in degrees Celsius

This study uses two widely used tree-based methods including Random Forest (RF) and XGBoost to estimate the PVSTs. Each decision tree in RF searches for the best feature out of a random subset of features to reduce the correlation between decision trees (Liaw & Wiener, 2002). The final RF estimated results are the average of the estimation from decision trees. The impurity score of mean square error (MSE) is calculated (Eq. (10)) to evaluate the decision trees as follows (Strobl et al., 2007):

$$MSE = \frac{(y_i - \widehat{y}_i)^2}{n} \tag{10}$$

where y_i and $\widehat{y}_i$ represent the true and estimated PVSTs, respectively, and n is the number of the training dataset. In addition, the bootstrap aggregating method is utilized to replace the random samples. The observations of a sample can be used as a validation dataset for the corresponding decision tree. On the other hand, XGBoost is a scalable end-to-end system for tree gradient boosting by simplifying the regularized learning objective and algorithm of the approximate framework for parallelization Chen & Guestrin, 2016). XGBoost can be described (Eqs. (11), ((12)) as follows:

$$\widehat{y}_i = \Phi(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{11}$$

$$F = \left\{f(x) = w_{q(x)}\right\},\ \left(q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T\right) \tag{12}$$

where F, K and T represents the space of regression tree, the additive functions used for estimation, and the number of leaves in each regression tree, respectively; f_k corresponds to a separate tree structure q and leaf weight w, while w_i represents a continuous score on i-th leaf in each regression tree; m is the number of thermal-related features. The final estimation is the sum of the score w. The $\widehat{y}_i^{(t)}$ is the estimated PVST of the i-th leaf at the t-th iteration, and the f_t is an additive term to minimize the loss function, which can be approximated (Eq. (13)) as follows:

$$\mathscr{L}^{(t)} \approx \sum_{i=1}^{n}\left[l\left(y_i, \widehat{y}^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{13}$$

where l evaluates the variation between the target y_i and the sum of the estimation at the $(i-1)$-th iteration $\widehat{y}^{(t-1)}$ and the approximate value of f_t in differentiable convex. The symbols g_i and h_i represent the first and second order gradient statistics on the loss function, respectively. The simplified objective can be obtained after eliminating the constant term in Eq. (14), while the penalty of model complexity Ω aims to prevent overfitting by smoothening the weight w, which can be obtained (Eq. (15)) as follows:

$$\widetilde{\mathscr{L}}^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{14}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda w^2 \tag{15}$$

where γ and λ are the L1 and L2 regularization terms, respectively, which are used to avoid overfitting.

To improve the estimation accuracy, grid search combined with cross-validation searches through the specific parameter grid for each

model in the independent validation dataset to achieve the best hyperparameter combination. Apart from the testing dataset taking account of 20 %, the remaining data is the training and validation dataset, which are utilized in the 5-fold cross-validation for hyperparameter tuning and model selection.

2.2.3. Evaluation metrics for validation

The best PVST estimation model is selected by the comparison of the evaluation metrics, including the coefficient of determination (R^2), mean absolute error (MAE), and RMSE between the observed LSTs and estimated PVSTs, which are calculated Eqs. (16)–((18)) to measure the accuracy of the estimation as follows (Akhsassi et al., 2018):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(LST - PVST_{est})^2}{\sum_{i=1}^{n}(\overline{LST} - PVST_{est})^2} \tag{16}$$

$$MAE = \frac{\sum_{i=1}^{n}|PVST_{est} - LST|}{n} \tag{17}$$

$$RMSE = \sqrt{\frac{(PVST_{est} - LST)^2}{n}} \tag{18}$$

Where LST, $PVST_{est}$, and $\overline{LST}$ represent the observed LST, estimated PVST, and average observed LST (°C), respectively; n is the observation number in the testing dataset.

2.3. The effects of the thermal environment on PVST

In this study, the contributions of different features to the PVST estimation are evaluated by feature importance (FI) scores, including three indices. First, the Mean Decrease in Impurity (MDI) observes the MSE reduction to quantify the contribution of each feature, which can be obtained inherently during the model training without additional computation. It is noticeable that MDI might overestimate the features with higher cardinality and is only effective within ensemble tree-based algorithms (Strobl et al., 2007). Second, the Permutation Importance (PI) determines the FI by evaluating the estimation variation by randomly shuffling a single feature at a time. Significant variation in the PVST estimation indicates high FI; however, PI might overstate the importance of correlated features (Gregorutti et al., 2017). Third, the SHapley Addictive exPlanations (SHAP) method averages the marginal contribution of each feature to the estimation over all possible feature combinations, which, however, is computationally intensive for high-dimensional datasets (Lundberg & Lee, 2017). Therefore, comprehensively using the three indices can unravel the effects of the dynamic thermal environment on PVST.

3. Estimation of the PV electricity generation

To accurately quantify the dynamic PVCE associated with specific time and location, PVST needs to be taken into account, as it determined the PV performance variation compared to the temperature in standard test conditions (T_{stc}, 25 °C). Thus, the dynamic estimated PVCE ($\eta(PVST)$) can be obtained based on the $PVST$ generated through the machine learning-based estimation model under dynamic thermal environments in Eq. (19) as follows (Dubey et al., 2013):

$$\eta(PVST) = \eta(T_{stc})[1 + \beta_{stc}(PVST - T_{stc})] \tag{19}$$

where the T_{stc}, $\eta(T_{stc})$ and β_{stc} represent the nominal temperature, PVCE and temperature coefficient at maximum power point at standard test conditions, respectively. They can be obtained from the nominal data from the manufacturer. Meanwhile, Eq. (19) is assumed effective when $|PVST - T_{stc}| \leq 20$ °C.

In this study, the dynamically estimated PVCE can be obtained by the estimated PVSTs, while the static PVCE is assumed to be 20 % for comparison. Therefore, the monthly average estimated PV electricity generation, E_{pv} (kWh) can be estimated in Eq. (20) as follows (Skoplaki & Palyvos, 2009):

$$E_{pv} = A \times SR \times \eta \times PR \tag{20}$$

where the A, SR, and η represent the floating PV areas (m^2), monthly solar radiation (kWh/m^2/month), and monthly average PVCE (%), respectively. PR is the performance ratio of a solar PV system with a typical value of 80 %, which considers the system losses such as inverter losses, cable losses, and dust effects (Berwal et al., 2017).

4. Empirical investigation

4.1. Study area

Singapore with only 5.6 million population (Koh, 2022) has consumed more than 53.5 terawatt-hours (TWh) electricity in 2021 (Energy Market Authority, 2022). Since Singapore has high annual average solar irradiance of 1663 kWh/m^2 (Doshi et al., 2013), the Singaporean government has established an initiative to increase solar PV deployment to 2 gigawatt-peak (GWp) by 2030 to facilitate the sustainable goals of the Green Plan 2030 and Net Zero Emissions by 2050 (Dörenkämper et al., 2021). Due to the limited land at 734.3 km^2 (Singapore Land Authority, 2023), the utility-scale PV systems are mainly deployed on the reservoirs and offshores. In this study, we investigated four floating PV farms including three ones on the reservoir and one offshore with a total installed capacity of 68 MWp, which have been operated since 2021 (Bi & Law, 2023; Liang et al., 2022). Fig. 2 shows the details of them including distribution, installed capacity, and covering area.

4.2. Data collection

For the urban thermal datasets of Singapore, ten thermal-related features from 2009 to 2022 are collected from Meteorological Service Singapore (National Environment Agency, 2023) and World Weather Online (2023), which are presented in Table 2. The distribution of the meteorological stations observing AT, H, and WS is shown in the Appendix A. The Landsat images with no cloud cover or as little cloud as possible were collected from earth explorer (United States Geological Survey, 2023), including Landsat 7, 8, and 9, because of their highest spatial resolution of 30 metres available for free at present. Each satellite passes over Singapore at around 11:16 AM local time every 16 days; however, the imaging time of Landsat 7 has been advanced by around 1.5 h from April 2022 due to the orbit variation. The scale of four floating PV farms is varying as shown in Fig. 2. Thus, the number of sampling points of floating PV farms at the Tengeh Reservoir, Strait of Johor, Bedok Reservoir, and Lower Seletar Reservoir is 446, 51, 22, and 16, respectively. Table 3 demonstrates the details of the collected Landsat images of each floating PV farm. Meanwhile, the atmospheric correction parameters were collected to compute the LST from the National Aeronautics and Space Administration (Kafer et al., 2020). The solar radiation in Singapore is obtained from the Annual Climatological Report in Singapore (2014–2018) at one-day intervals (Meteorological Sevice Singapore, 2022) and National Solar Radiation Database (2016–2020) at one-hour intervals (National Solar Radiation Database, 2020) for the PV electricity generation estimation.

4.3. Data preprocessing

Since the atmospheric effects lead to huge bias to the LST retrieval, original Landsat images require radiometric calibration and atmospheric correction, which aim to convert the raw digital numbers in the images to top-of-atmosphere (TOA) reflectance values (Chander et al., 2009) and remove or minimize the influence of atmospheric effects by

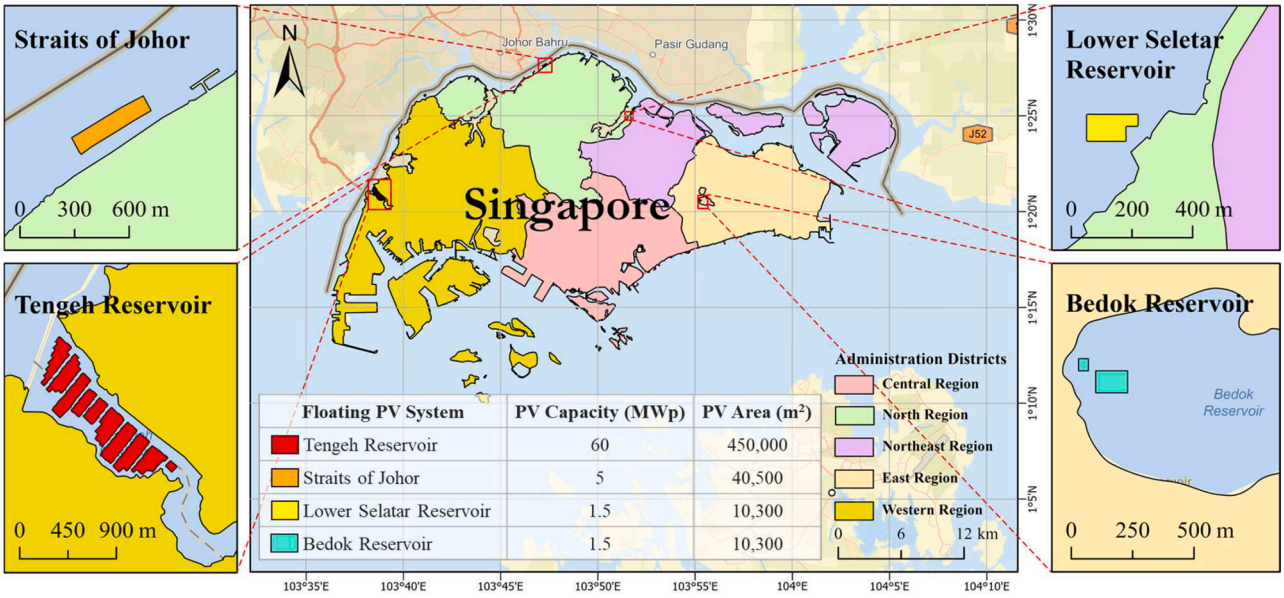


Fig. 2. The investigated four floating PV farms in Singapore.

Table 2
Collected urban thermal-related features in Singapore (2009–2022) (National Environment Agency, 2023; World Weather Online, 2023).

Source	Spatial resolution	Urban thermal-related feature	Temporal resolution
Meteorological Service Singapore	At weather-station level	Air temperature Relative humidity Wind speed Precipitation	One-minute interval
		Ultraviolet index (daytime)	One-hour interval
World Weather Online	At Singapore's level	Cloud cover amount Atmospheric pressure Visibility Heat index temperature Dew point temperature	One-hour interval

Table 3
The captured time and the number of collected satellite images from each Landsat mission (United States Geological Survey, 2023).

Location of floating PV farms	Number of collected images	Captured time (dd/mm/yyyy)
Tengeh Reservoir	7	12/02/2022, 20/02/2022, 16/03/2022, 24/03/2022, 01/04/2022, 16/09/2022, 02/08/2023
Strait of Johor	5	03/07/2021, 23/10/2021, 20/02/2022, 08/03/2022, 15/06/2023
Bedok Reservoir	13	17/06/2021, 03/07/2021, 15/10/2021, 08/03/2022, 24/03/2022, 17/04/2022, 04/06/2022, 30/07/2022, 01/09/2022, 30/04/2023, 14/05/2023, 22/05/2023, 30/05/2023
Lower Seletar Reservoir	13	25/06/2021, 03/07/2021, 15/10/2021, 23/10/2021, 20/02/2022, 01/04/2022, 17/04/2022, 04/06/2022, 15/08/2022, 22/10/2022, 12/04/2023, 30/04/2023, 14/05/2023

Fast Line-of-Sight Atmospheric Analysis of Hypercubes (Katkovsky et al., 2018), respectively. To reduce the computational time, two operations were made sequentially. First, the subsets of the radiometrically calibrated satellite images were implemented in the ENVI software before executing the atmospheric correction, which cover the PV and surrounding ground to compute the LST at and near the PV panels. Second, only the urban thermal-related data during the day was kept, while the remaining data were verified to drop the rows with missing or null values and wrong time sequences. After that, the thermal-related data were linearly interpolated using data corresponding to the interval before and after the Landsat imaging time. Meanwhile, the air temperature, relative humidity, and wind speed corresponding to the floating PV locations are computed by using the simple Kriging interpolation as an example shown in Fig. 3 for the date of 20 Feb. 2022 at the Tengeh Reservoir, 23 Oct. 2021 at the Straits of Johor, 15 Oct. 2021 at the Bedok Reservoir, and 15 Aug. 2022 at the Lower Seletar Reservoir, respectively. An accuracy assessment based on RMSE was performed to compare the performance of different variogram models including the Gaussian, Linear, Spherical, Power, Exponential, and Hole-effect, which aims to assign weights to the known points in the vicinity of unsampled points, while the remaining thermal-related features including V, P, CC, UVI, R, HIT, and DPT, are considered to have the same values in the whole of Singapore at the imaging time. Overall, the interpolation results indicate that the RMSEs of air temperature, relative humidity, and wind speed range from 0.41 °C to 1.43 °C, from 3.31 % to 11.87 %, and from 0.75 m/s to 2.96 m/s, respectively (Appendix B).

4.4. Feature selection for PVST estimation model

As described in Section 2.2.1, Pearson and Spearman's rank correlation coefficient measure the linear and monotonic relationships between pairs of thermal-related features and PVST for feature selection (Fig. 4(a, b)). The coefficients in the first column on the left exhibit the effects of each feature on the PVST estimation, which indicates that HIT, DPT, and CC are the top 3 linear correlated features to PVST. Since almost all the p-values are smaller than 0.05, it indicates that the coefficients between the thermal environment and PVST are statistically significant (Fig. 4(c, d)). However, highly correlated features may result in a low model's interpretability as described in Section 2.3. Therefore, HIT and DPT will not be used for building the PVST estimation models due to their dependence on AT, H, and WS (Camuffo, 2019; National

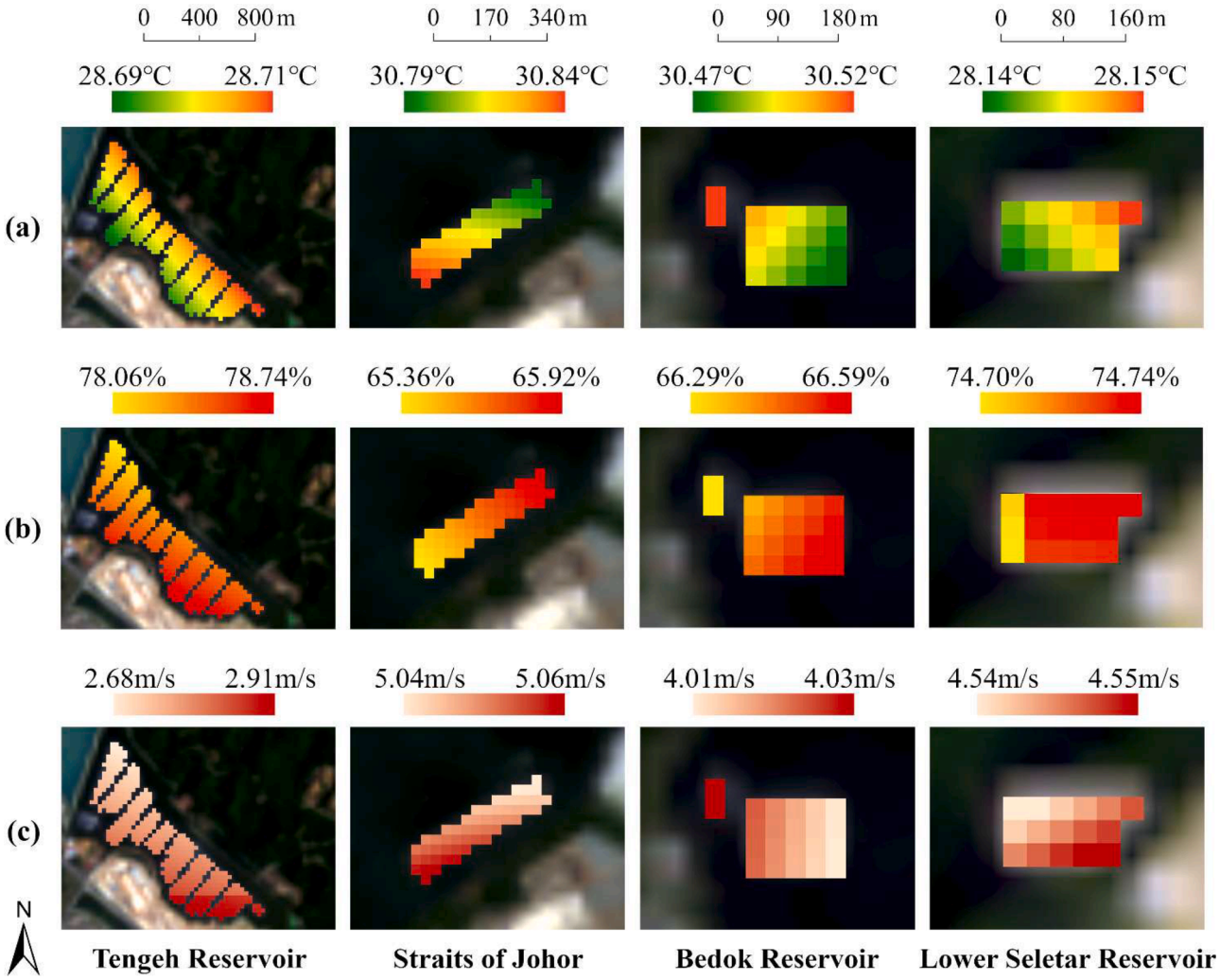


Fig. 3. Simple Kriging interpolation results. (a) AT. (b) H. (c) WS.

Oceanic & Atmospheric Administration, 2005). This means that eight features are maintained based on PPS and Boruta. Although CC is highly correlated to H and PVST and has a relatively high PPS to estimate the PVST, Boruta suggests that CC had a lower feature importance than the best shadow feature in each of the 20 trials as well as V, UVI, and R, which indicates they are unimportant. Meanwhile, the PPS heatmap shows a low estimation capacity of V and UVI (Fig. 4(e)). This suggests that the four features will not be considered either. As a result, AT, H, WS, and P were selected to build the PVST estimation models.

4.5. The estimated PVSTs

The observed LSTs on four floating PVs were merged to train the estimation model because (i) the scale of the floating PVs varies greatly, and (ii) a single floating PV cannot provide sufficient samples. The observed 3871 records were divided into three sub-datasets for training, validation, and testing. The testing dataset accounts for 20 %, while the remaining is utilized 5-fold cross-validation to avoid the estimation models overfitting in the small dataset. Fig. 5(a) shows the observed LSTs on floating PV surfaces retrieved from the Landsat imagery for the dates of 16 Sept. 2022 at the Tengeh Reservoir, 23 Oct. 2021 at the Straits of Johor, 17 June 2021 at the Bedok Reservoir, and 20 Feb. 2022 at the Lower Seletar Reservoir, respectively. Fig. 5(b–d) present the estimated PVSTs based on the RF, SVM, and XGBoost, respectively, which are close to the observed LSTs with the largest difference less than 1 °C. The three PVST estimation methods also capture the characteristic of higher PVSTs in the central PV area from the observation results, while the edge of floating PV surfaces has lower PVSTs benefiting from the cooling effects of ambient thermal environments.

4.6. Evaluation metrics

In this study, the RF estimation model exhibits the best performance during the 5-fold cross-validation with the R^2 of 0.929, the MAE of 0.648, and the RMSE of 0.873. The yellow and red bars in Fig. 6(a–c) show the evaluation metrics of R^2, MAE, and RMSE for the training and testing datasets, respectively. The three estimation models show similar evaluation results with insignificant variation between these two datasets indicating that they are not overfitting. The R^2 of the testing dataset of three models indicates that their estimated PVSTs perfectly explains the observed LSTs on floating PV surfaces. The MAE of the three models has a range from 0.725 to 0.751 in the testing dataset, while the RMSE is from 0.961 to 0.975 (Fig. 6(d–f)). Although the R^2 of the RF is a little lower than that of the XGBoost, the difference between them is too little to affect the PVST estimation. Overall, the PVST estimation model based on the RF has a better performance with a lower MAE and a lower RMSE in the testing dataset. Therefore, RF will be used for evaluating the effects of the dynamic thermal environment on the PVST estimation because it exhibits the best performance in the training, validation, and testing datasets.

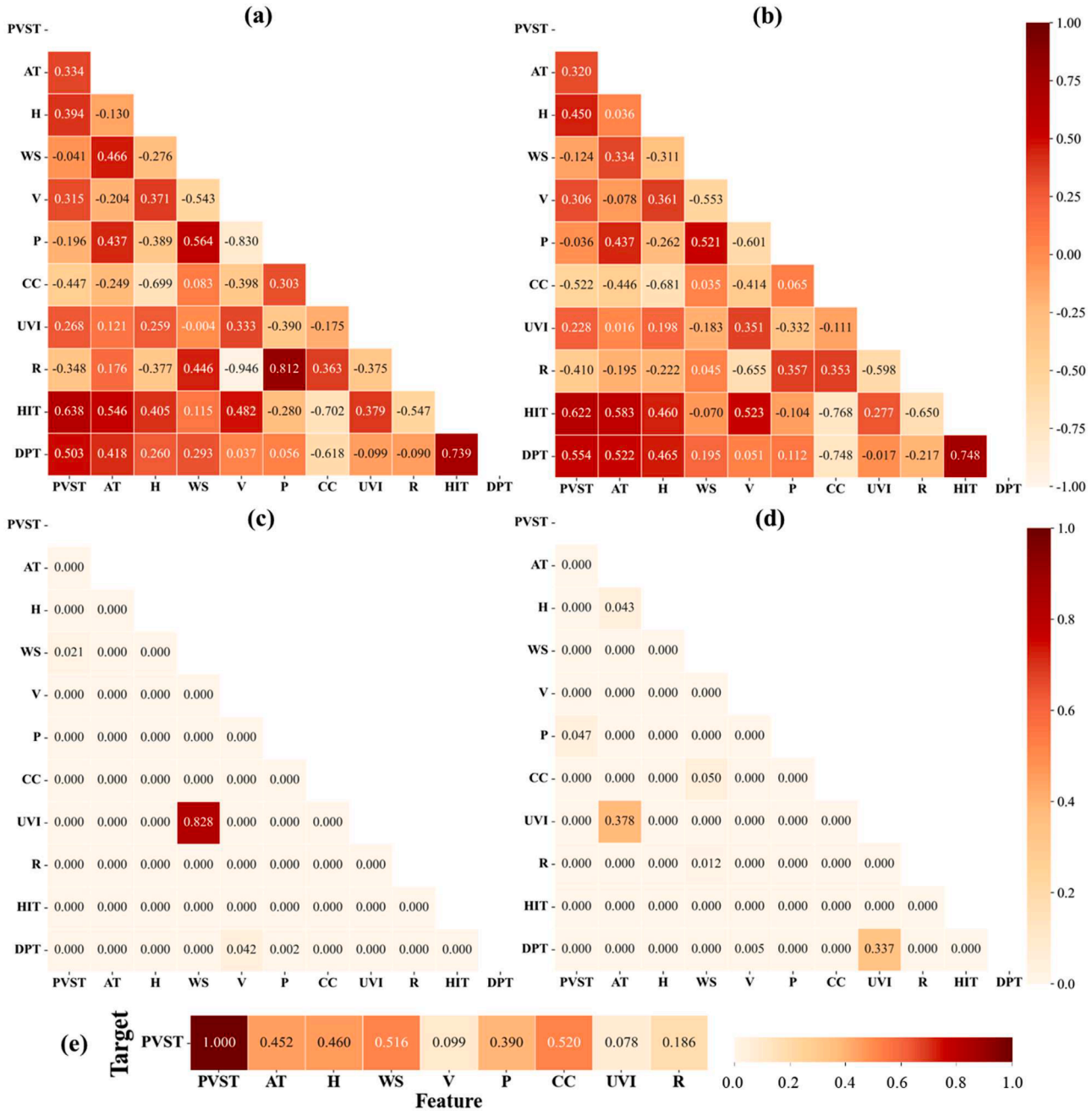


Fig. 4. Feature selection heatmaps. (a) Pearson correlation coefficients. (b) Spearman's rank correlation coefficients. (c) P-value corresponding to Pearson correlation coefficients. (d) P-value corresponding to Spearman's rank correlation coefficients. (e) PPS scores of the remaining eight variables and PVST.

4.7. Effects of urban thermal environment on PVST estimation

Comprehensively considering the pros and cons of the feature-importance methods, this study depicts the feature importance of PVST estimation based on MDI, PI, and SHAP values (Fig. 7). The results suggest that AT and H are the most important two features to estimate the PVST. Specifically, their contribution accounts for 63.6 %, 70.8 %, and 68.7 % regarding MDI, PI, and SHAP, respectively. As described in Section 2.3, MDI may overestimate the feature with higher cardinality (Strobl et al., 2007), while PI may overstate the FI of highly correlated features (Gregorutti et al., 2017). Thus, P has the highest cardinality followed by WS, resulting in higher contributions of them in MDI compared to the other two methods. Pearson correlation coefficient of AT and WS is 0.47, while that of AT and P is 0.44 (Fig. 4(a)). Therefore, the FI of AT in PI is higher than the other two methods.

Fig. 8 displays the distribution of SHAP values of each instance for the four urban thermal-related features. Overall, PVST was observed to be positively correlated with AT, whereas PVST was negatively correlated with P and WS, respectively. The SHAP values distribution of H is complicated, and an extremely high or low value of H generally shows a positive correlation with PVST, while the effect of intermediate feature values is indeterminate. Despite the inherent strengths and weaknesses of each feature importance method, Fig. 7 basically demonstrates a consistency in the contribution of each feature, thereby indicating the reliability of the results. Overall, the contribution to the PVST estimation is ordered by AT, H, P, and WS with approximate values of 35.1 %, 33.6 %, 18.4 %, and 12.9 % based on the SHAP values (Fig. 7(c)), respectively, in this study.

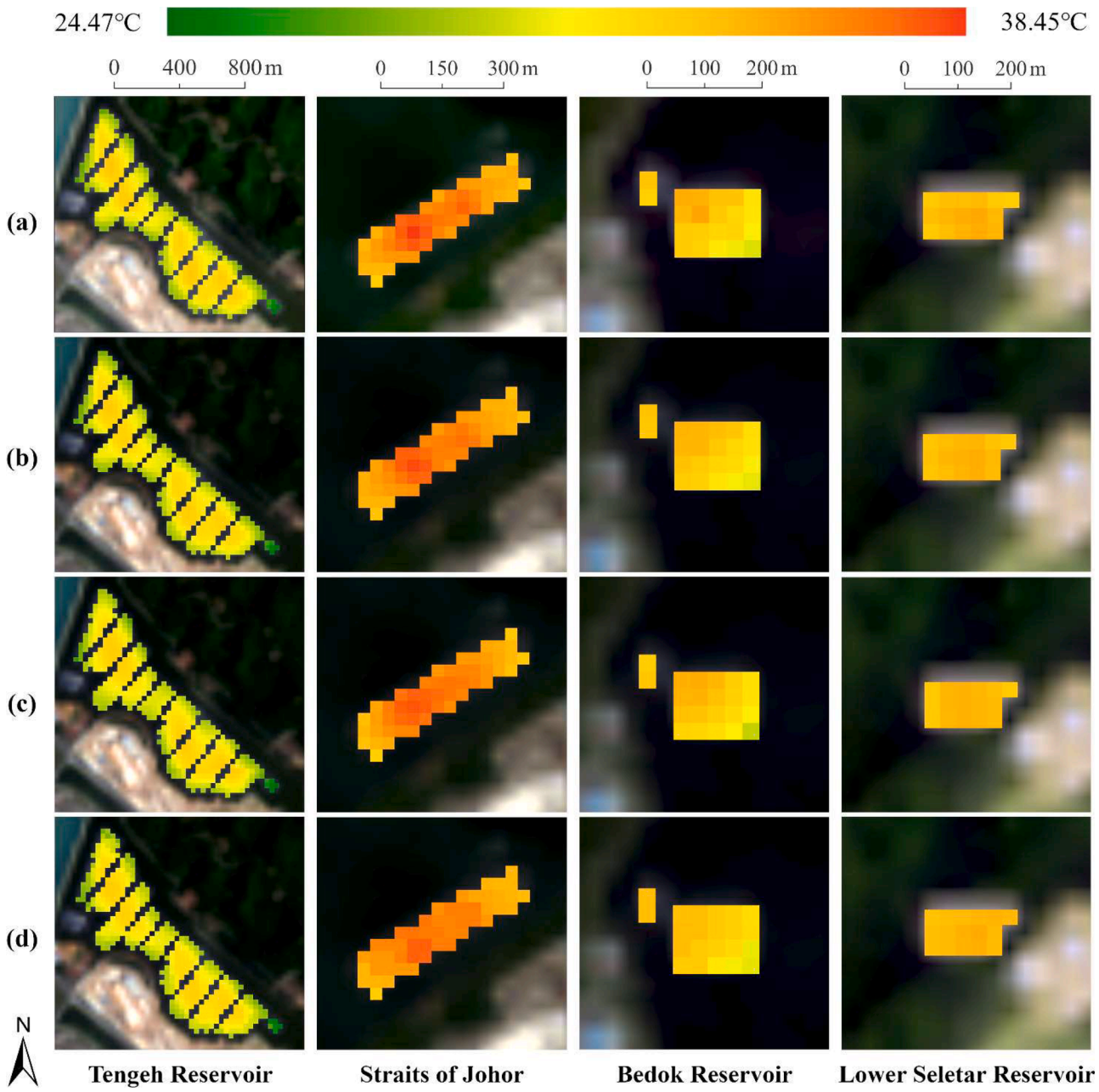


Fig. 5. Observed LSTs versus estimated PVSTs from the Landsat imagery for the dates of 16 Sept. 2022 at the Tengeh Reservoir, 23 Oct. 2021 at the Straits of Johor, 17 June 2021 at the Bedok Reservoir, and 20 Feb. 2022 at the Lower Seletar Reservoir, respectively. (a) Observed LSTs. (b) Estimated PVSTs using RF. (c) Estimated PVSTs using SVM. (d) Estimated PVSTs using XGBoost.

5. Estimation of the floating PV electricity generation

Fig. 9 depicts the estimated monthly PVCE ranging from 20.82 % to 20.94 % in Singapore based on the estimated monthly PVST. The thermal-related features of average monthly AT, H, P, and WS determined by the meteorological statistics from 2009 to 2022 (National Environment Agency, 2023; World Weather Online, 2023) were used as input thermal-related features to obtain the monthly PVSTs from the RF estimation model. Due to the cooling effects of the floating PV, the estimated PVCE was close to the nominal PVCE at the standard test conditions (21.4 %) (Trinasolar, 2022). The PVCE has little change due to the steady variation of the thermal-related conditions in different months of the year in Singapore (National Environment Agency, 2023; World Weather Online, 2023). Also, the dynamically estimated PVCE is consistently larger than the static PVCE of 20 % due to the cooling effects of the floating PV surface (Dörenkämper et al., 2021), which improves the estimation of PV electricity generation. It is found that the estimated monthly PVCEs have relatively high values ($\geq$ 20.9 %) in January, February, November, and December. Meanwhile, the difference between the estimated PVSTs of each month ranges between 34.0 °C and 36.1 °C, while the average monthly AT has a range from 28.2 °C to 30.5 °C. Although the PVCE had little change throughout the year in this study, the estimated PV electricity generation can be a drastic difference if the total PV area is large enough, such as the utility-scale floating PVs, or the floating PVs deployed in high-latitude locations.

Finally, without modelling the effects of the thermal environment by simply assuming that the PVCE is 20 % and the performance ratio is 80 %, the statically estimated monthly electricity generation ranges from 9567 to 13,286 megawatt-hours (MWh), as the yellow bars shown in Fig. 10. In comparison, by incorporating the effects of the thermal

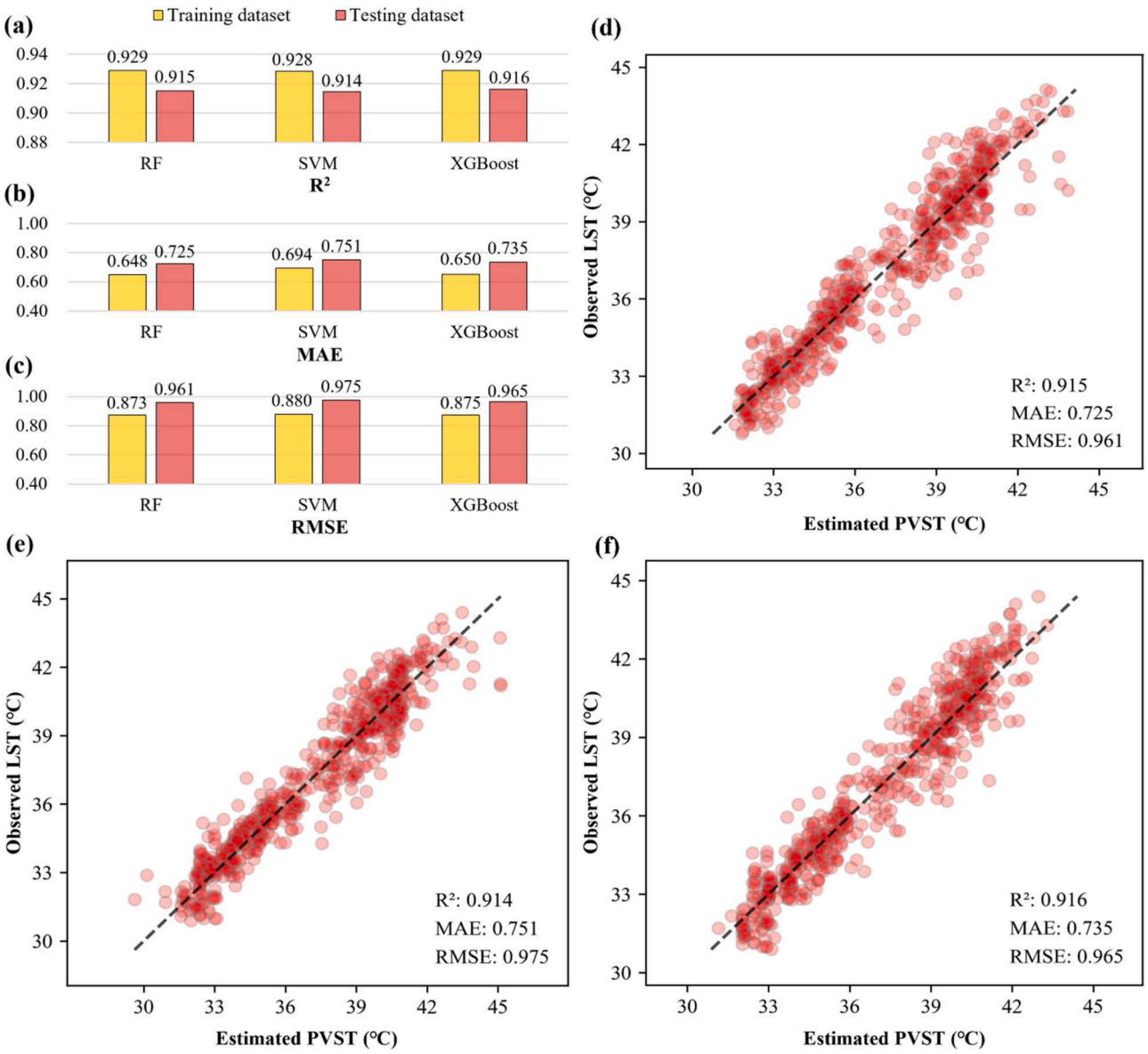


Fig. 6. The evaluation metrics and estimation error for PVST of three estimation models. (a) R^2. (b) MAE. (c) RMSE. (d) The estimation error of RF for the testing dataset. (e) The estimation error of SVM for the testing dataset. (f) The estimation error of XGBoost for the testing dataset.

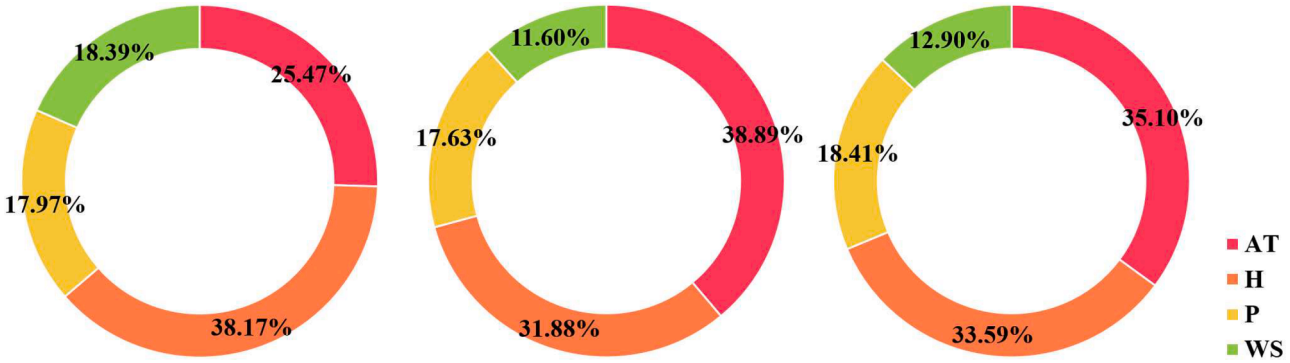


Fig. 7. Feature importance on PVST estimation in RF. (a) MDI. (b) PI. (c) SHAP values.

environment on the PVCE, the dynamically estimated monthly electricity generation ranges from 10,000 to 13,877 MWh as the red bars shown in Fig. 10. Specifically, the dynamically estimated electricity generation is higher than the statically estimated one in each month (Table 4). It is found that the monthly average electricity generation has a peak in March because of relatively high estimated PVCE and solar

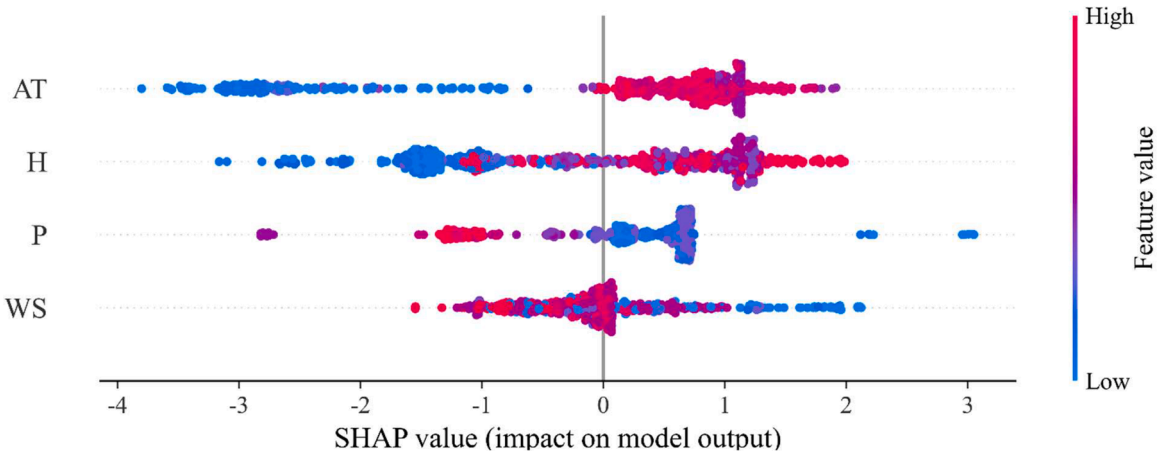


Fig. 8. The distribution of SHAP values for PVST.

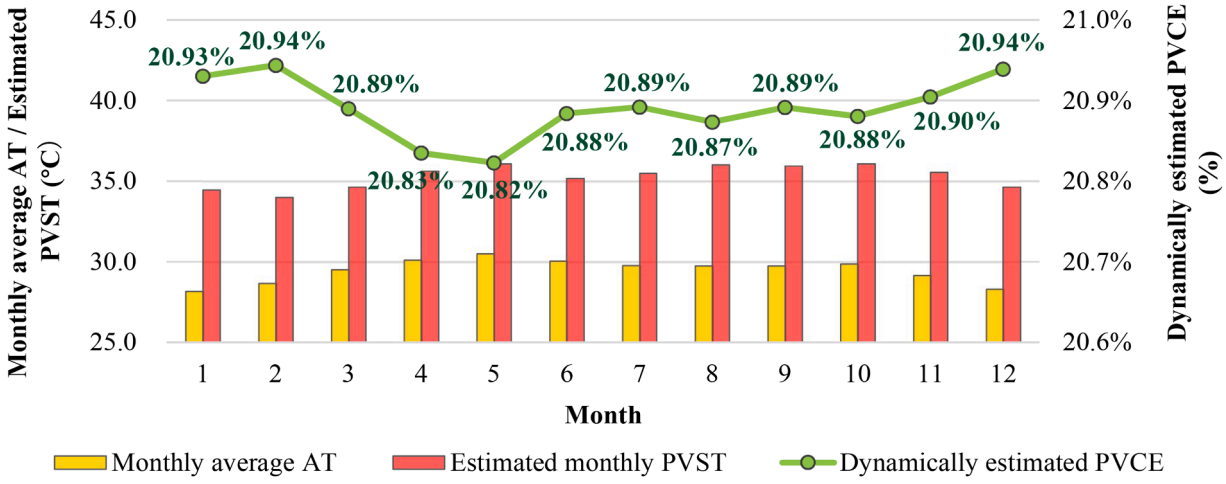


Fig. 9. Estimated monthly PVCE based on the estimated PVST in Singapore.

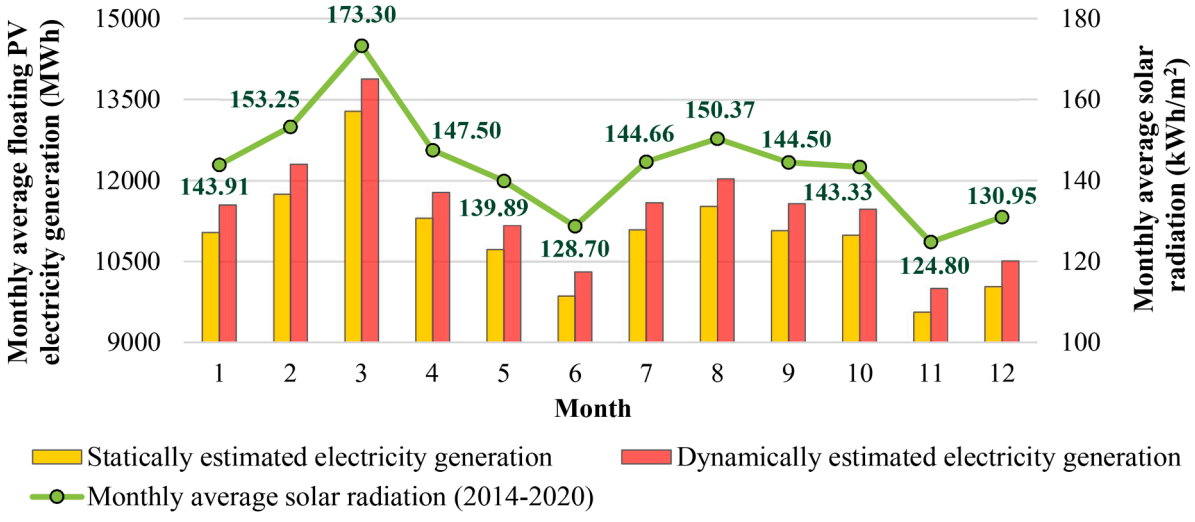


Fig. 10. The electricity generation estimation of four floating PV farms in Singapore (static versus dynamic PVCE).

radiation as the dot on the green line corresponding to March shown in Fig. 9 and Fig. 10, respectively. Overall, the dynamically estimated annual electricity generation is 5888 MWh higher than the statically estimated annual electricity generation.

6. Discussion and conclusion

This study proposed a novel estimation approach for unravelling the relationship between observed LSTs on floating PV surfaces from remote

Table 4
The statically and dynamically estimated electricity generation and their difference.

Month	Statically estimated electricity generation (MWh)	Dynamically estimated electricity generation (MWh)	Difference (dynamically -statically estimated electricity generation) (MWh)
Jan.	11,033	11,546	513
Feb.	11,749	12,303	554
Mar.	13,286	13,877	591
Apr.	11,308	11,780	472
May	10,725	11,166	441
June	9867	10,303	436
July	11,091	11,585	494
Aug.	11,528	12,031	503
Sept.	11,078	11,572	494
Oct.	10,988	11,472	484
Nov.	9567	10,000	433
Dec.	10,039	10,510	471

sensing images and thermal-related features to provide continuous PVSTs for large geographical scale floating PV systems, which does not require consideration of indoor lab conditions. It is useful for the accurate estimation of PV electricity generation. Three estimation models for estimating the PVSTs (with a lower RMSE between 0.96 °C and 0.97 °C in the testing dataset) show similar results, which improves the estimation to achieve a highly accurate PVCE and PV electricity generation. Based on the RF estimation model, the air temperature has the highest estimated capacity of 35.1 % to the PVST estimation in Singapore, while relative humidity, atmospheric pressure, and wind speed account for 33.6 %, 18.4 %, and 12.9 %, respectively. Also, the dynamically estimated PVCE consistently exceeds 20 % due to the cooling effects of floating PV surface, which enables the dynamically estimated annual PV electricity generation to be 5888 MWh higher than the statically estimated one.

This study is vital in three aspects. First, we developed a machine learning-based approach to accurately estimating the PVST under dynamic thermal environments and proposing various evaluation metrics for validation, which successfully generate PVSTs from thermal-related features. Second, this study provides a comprehensive and in-depth understanding of the contribution of each thermal-related feature on the PVST estimation under real and dynamic geographical conditions. The results are particularly useful for the accurate estimation of PV electricity generation due to the correlation between the PVST and PVCE, which proposes an alternative approach to evaluate the effects of the urban thermal environment on PV electricity generation. Third, this study achieves reliable dynamic PVCEs by considering the PVST variation, which is significant to estimate the utility-scale PV electricity generation. In this regard, our approach provides accessible and accurate PV electricity generation of large-scale installed PV farms for different stakeholders such as the owner, operator, and manufacturer. Thus, our approach is important to plan PV installation considering life-cycle economic feasibility (long-term impact) and to operate dynamic grid-load balancing (short-term impact). Meanwhile, it can provide a benchmark to reveal potential PV system failures with the PV electricity generation estimation.

Appendix A. The distribution of weather stations in Singapore

Fig. A.1, Table B.1, Table B.2, Table B.3

The repeat cycle of satellite image acquisition restricts the retrieval of continuous PVSTs. Although Landsat 8 and 9 operating on the same orbit can be regarded as halving the repeat cycle, it is difficult to obtain appropriate satellite images with no cloud or as little cloud as possible. Meanwhile, the similar satellite imaging time leads to the lack of daily variations. Nevertheless, the available images from June 2021 to August 2023 can provide sufficient observed LSTs on floating PV surfaces at the air temperatures between 28.2 °C and 30.5 °C that covered the annual air temperature range in Singapore, which makes the built model capable to estimate PVCE in different thermal-related conditions. In addition, the thermal-related features have little variation in Singapore during the daytime. For example, the range of air temperature corresponding to the imaging time covers both the average air temperature for each two-hour interval from 8:00 to 18:00 every day and the monthly average air temperature from 2009 to 2022. Therefore, it is possible to build a reliable machine learning-based model under the constraints of remote sensing data, which can improve the PVST estimation accuracy.

Notably, the Landsat image is widely available and transferable to other regions with the advantages of free access and easy retrieval, which brings good generalization capacity to the PVST estimation model. For estimations with finer spatial resolution, future work can collect the commercial satellite imagery from the thermal sensors with a higher resolution, such as the Orbita hyperspectral satellites with a spatial resolution of 10 metres and a revisit period of one day, to achieve more observed LSTs on floating PV surfaces for training, testing, and validation to build the PVST estimation model. This study used the empirical formula to estimate the PVCE based on PVST and the nominal parameters at the standard test conditions, which is only valid when the PVST is less than 45 °C. To estimate more accurate PV electricity generation, we plan to develop PVCE estimation models in the future based on the PV configuration information and the real PVCE as well as PV electricity generation data.

In conclusion, this study is innovative in proposing a model that accurately and easily estimates the PVST from the dynamic thermal environments and estimates the PV electricity generation on installed large-scale PV systems. This study should be the first one to quantify the effects of the thermal environment on PVST in a dynamic geographical environment and longtime duration. This study is significant to obtain dynamic and more accurate PVCE. The proposed methodology is independent and can be generalized and transported to other places; this is particularly useful for various applications related to floating PV deployment.

Declaration of Competing Interest

Our paper has not been previously published and is not currently under consideration by another journal. We have no conflicts of interest to disclose. We have obtained all necessary permissions for any data, images, or materials used in our paper.

Data availability

Data will be made available on request.

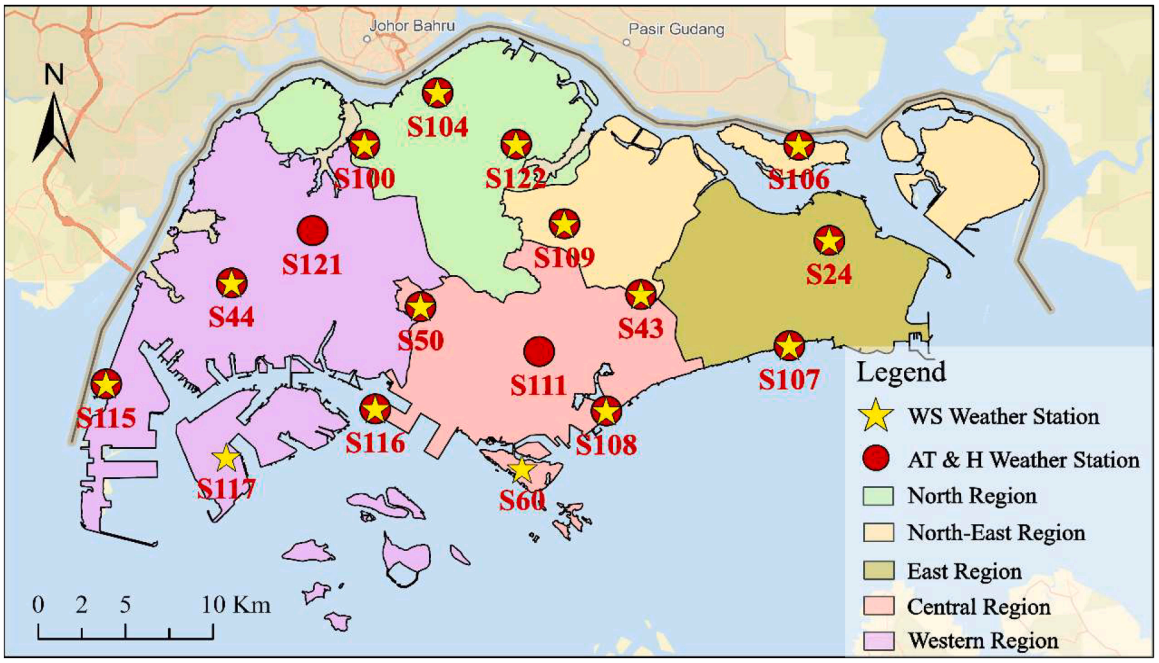


Fig. A.1. The distribution of weather stations in Singapore.

Appendix B. The Simple Kriging interpolation results

Table B.1
The hyperparameters and RMSE evaluation of simple kriging interpolation of air temperature. Overall, the RMSE of the air temperature in the best hyperparameter combinations is lower than 1 °C.

Date	Number of closest points	Number of lags	Variogram model	RMSE (°C)
2021 - 06 - 17	12	3	Gaussian	0.68
2021 - 06 - 25	13	4	Gaussian	0.77
2021 - 07 - 03	3	7	Gaussian	0.94
2021 - 10 - 15	11	4	Gaussian	0.48
2021 - 10 - 23	4	5	Spherical	0.99
2022 - 02 - 12	8	16	Hole-effect	0.41
2022 - 02 - 20	2	7	Gaussian	0.58
2022 - 03 - 08	7	5	Gaussian	0.93
2022 - 03 - 24	11	3	Gaussian	0.67
2022 - 04 - 01	9	12	Hole-effect	0.55
2022 - 04 - 17	3	4	Gaussian	1.43
2022 - 06 - 04	13	10	Gaussian	0.51
2022 - 07 - 30	2	7	Hole-effect	0.66
2022 - 08 - 15	4	4	Gaussian	0.53
2022 - 09 - 01	2	5	Spherical	0.80
2022 - 09 - 16	4	4	Gaussian	0.47
2022 - 10 - 22	6	14	Power	0.58
2023 - 04 - 12	3	14	Gaussian	0.47
2023 - 04 - 30	2	3	Gaussian	0.65
2023 - 05 - 14	3	7	Gaussian	0.81
2023 - 05 - 22	3	4	Gaussian	0.72
2023 - 05 - 30	6	12	Hole-effect	0.75
2023 - 06 - 15	9	6	Spherical	0.75
2023 - 06 - 23	4	13	Hole-effect	0.77
2023 - 08 - 02	3	5	Gaussian	0.59

Table B.2
The hyperparameters and RMSE evaluation of simple kriging interpolation of relative humidity. In general, the RMSE of the relative humidity in the best hyperparameter combinations ranges from 3.4 % to 11.9 %. Most of them are concentrated in the interval from 3 % to 7 %.

Date	Number of closest points	Number of lags	Variogram model	RMSE (%)
2021 - 06 - 17	13	4	Gaussian	3.35
2021 - 06 - 25	7	4	Gaussian	6.66
2021 - 07 - 03	4	4	Gaussian	5.99

(*continued on next page*)

Table B.2 (*continued*)

Date	Number of closest points	Number of lags	Variogram model	RMSE (%)
2021 - 10 - 15	8	10	Spherical	5.73
2021 - 10 - 23	4	5	Gaussian	6.89
2022 - 02 - 12	3	8	Power	3.31
2022 - 02 - 20	8	8	Gaussian	4.15
2022 - 03 - 08	6	7	Gaussian	4.59
2022 - 03 - 24	12	12	Gaussian	10.12
2022 - 04 - 01	12	8	Power	10.14
2022 - 04 - 17	13	6	Gaussian	11.87
2022 - 06 - 04	5	5	Gaussian	5.47
2022 - 07 - 30	5	5	Gaussian	5.45
2022 - 08 - 15	4	6	Spherical	6.99
2022 - 09 - 01	7	4	Gaussian	8.12
2022 - 09 - 16	5	4	Gaussian	5.16
2022 - 10 - 22	5	3	Spherical	5.19
2023 - 04 - 12	5	10	Hole-effect	2.16
2023 - 04 - 30	6	14	Hole-effect	5.89
2023 - 05 - 14	4	6	Gaussian	5.28
2023 - 05 - 22	4	7	Gaussian	3.66
2023 - 05 - 30	4	13	Power	6.40
2023 - 06 - 15	6	14	Gaussian	3.94
2023 - 06 - 23	4	3	Gaussian	4.09
2023 - 08 - 02	4	5	Gaussian	4.37

Table B.3
The hyperparameters and RMSE evaluation of simple kriging interpolation of wind speed. The RMSE of the wind speed ranges from 0.75 m/s to 2.96 m/s, and most of them are concentrated in the interval from 1 to 2 m/s.

Date	Number of closest points	Number of lags	Variogram model	RMSE (m/s)
2021 - 06 - 17	2	11	Hole-effect	1.76
2021 - 06 - 25	6	5	Power	1.18
2021 - 07 - 03	3	12	Gaussian	1.67
2021 - 10 - 15	4	13	Hole-effect	1.70
2021 - 10 - 23	2	5	Spherical	1.90
2022 - 02 - 12	5	3	Gaussian	1.15
2022 - 02 - 20	3	9	Hole-effect	1.61
2022 - 03 - 08	3	6	Spherical	0.75
2022 - 03 - 24	6	10	Hole-effect	1.03
2022 - 04 - 01	3	6	Hole-effect	2.96
2022 - 04 - 17	2	14	Linear	1.96
2022 - 06 - 04	4	3	Gaussian	1.37
2022 - 07 - 30	3	4	Gaussian	1.44
2022 - 08 - 15	9	5	Gaussian	2.44
2022 - 09 - 01	8	9	Hole-effect	1.65
2022 - 09 - 16	6	11	Hole-effect	1.62
2022 - 10 - 22	9	8	Hole-effect	1.67
2023 - 04 - 12	2	12	Gaussian	1.54
2023 - 04 - 30	5	8	Gaussian	2.53
2023 - 05 - 14	7	4	Linear	1.42
2023 - 05 - 22	4	11	Hole-effect	1.75
2023 - 05 - 30	2	3	Spherical	1.22
2023 - 06 - 15	4	3	Gaussian	1.28
2023 - 06 - 23	4	5	Gaussian	1.60
2023 - 08 - 02	4	7	Hole-effect	1.84

References

Akhsassi, M., El Fathi, A., Erraissi, N., Aarich, N., Bennouna, A., Raoufi, M., et al. (2018). Experimental investigation and modeling of the thermal behavior of a solar PV module. *Solar Energy Materials and Solar Cells, 180*, 271–279. https://doi.org/10.1016/j.solmat.2017.06.052

Aly, S. P., Ahzi, S., Barth, N., & Figgis, B. W. (2018a). Two-dimensional finite difference-based model for coupled irradiation and heat transfer in photovoltaic modules. *Solar Energy Materials and Solar Cells, 180*, 289–302. https://doi.org/10.1016/j.solmat.2017.06.055

Aly, S. P., Barth, N., Figgis, B. W., & Ahzi, S. (2018b). A fully transient novel thermal model for in-field photovoltaic modules using developed explicit and implicit finite difference schemes. *Journal of Computational Science, 27*, 357–369. https://doi.org/10.1016/j.jocs.2017.12.013

Anyanwu, G. O., Nwakanma, C. I., Lee, J. M., & Kim, D. S. (2023). RBF-SVM kernel-based model for detecting DDoS attacks in SDN integrated vehicular network. *Ad hoc Networks, 140*, Article 103026. https://doi.org/10.1016/j.adhoc.2022.103026

Arslan, E. (2023). Applying regression techniques to determine mathematical equations of exergy, electricity, and energy values of photovoltaic thermal collector. *Solar Energy, 255*, 369–380. https://doi.org/10.1016/j.solener.2023.02.020

Bahar, H., & Analyst, S. (2023). Renewables 2022 analysis and forecast to 2027. Retrieved 19 June 2023 from https://www.iea.org/.

Becker, F., & Li, Z. L. (1990). Towards a local split window method over land surfaces. *International Journal of Remote Sensing, 11*(3), 369–393. https://doi.org/10.1080/01431169008955028

Berwal, A. K., Kumar, S., Kumari, N., Kumar, V., & Haleem, A. (2017). Design and analysis of rooftop grid tied 50kW capacity Solar Photovoltaic (SPV) power plant. *Renewable and Sustainable Energy Reviews, 77*, 1288–1299. https://doi.org/10.1016/j.rser.2017.03.017

Bi, C., & Law, A. W. K. (2023). Co-locating offshore wind and floating solar farms-Effect of high wind and wave conditions on solar power performance. *Energy, 266*, Article 126437. https://doi.org/10.1016/j.energy.2022.126437

Camuffo, D. (2019). Chapter 19 - measuring time of wetness and moisture in materials. D. Camuffo. *Microclimate for cultural heritage* (pp. 459–482). Elsevier. https://doi.org/10.1016/B978-0-444-64106-9.00019-5.

Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment, 113*(5), 893–903. https://doi.org/10.1016/j.rse.2009.01.007

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Dörenkämper, M., Wahed, A., Kumar, A., de Jong, M., Kroon, J., & Reindl, T. (2021). The cooling effect of floating PV in two different climate zones: A comparison of field test data from the Netherlands and Singapore. *Solar Energy, 219*, 15–23. https://doi.org/10.1016/j.solener.2021.03.051

Doshi, T., Sebastian, N., Souza, D., Nguyen, L., Teo, H., & Guan. (2013). The economics of solar PV in Singapore. *Journal of Engineering Technology, 2*(1). https://doi.org/10.5176/2251-3701_2.1.33

Dubey, S., Sarvaiya, J. N., & Seshadri, B. (2013). Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world – a review. *Energy Procedia, 33*, 311–321. https://doi.org/10.1016/j.egypro.2013.05.072

Energy Market Authority. (2022,. Oct. 21, 2022). Singapore energy statistics 2022. Retrieved 19 June 2023 from https://www.ema.gov.sg/singapore-energy-statistics/Ch03/index3.

Fan, S., Cao, S., & Zhang, Y. (2020). Temperature prediction of photovoltaic panels based on support vector machine with pigeon-inspired optimization. *Complexity, 2020*, 1–12. https://doi.org/10.1155/2020/9278162 (New York, N.Y.).

Ferber, R. (1956). Are correlations any guide to predictive value? *Applied Statistics, 5*(2), 113–121. https://doi.org/10.2307/2985494

Gillespie, A., Rokugawa, S., Matsunaga, T., Cothern, J. S., Hook, S., & Kahle, A. B. (1998). A temperature and emissivity separation algorithm for advanced spaceborne thermal emission and reflection radiometer (ASTER) images. *IEEE Transactions on Geoscience and Remote Sensing, 36*(4), 1113–1126. https://doi.org/10.1109/36.700995

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing, 27*(3), 659–678. https://doi.org/10.1007/s11222-016-9646-1

Hegazy, A., El Shenawy, E. T., & Ibrahim, M. A. (2019). Determination of the PV module surface temperature based on neural network using solar radiation and surface temperature. *ARPN Journal of Engineering and Applied Sciences, 14*(2), 494–503.

Ji, F., Wu, Z., Huang, J., & Chassignet, E. P. (2014). Evolution of land surface air temperature trend. *Nature Climate Change, 4*(6), 462–466. https://doi.org/10.1038/nclimate2223

Jung, D. E., Lee, C., Kim, K. H., & Do, S. L. (2020). Development of a predictive model for a photovoltaic module's surface temperature. *Energies, 13*(15), 18. https://doi.org/10.3390/en13154005

Kaldellis, J. K., Kapsali, M., & Kavadias, K. A. (2014). Temperature and wind speed impact on the efficiency of PV installations. Experience obtained from outdoor measurements in Greece. *Renewable Energy, 66*, 612–624. https://doi.org/10.1016/j.renene.2013.12.041

Katkovsky, L., Martinov, A., Siliuk, V., Ivanov, D., & Kokhanovsky, A. (2018). Fast atmospheric correction method for hyperspectral data. *Remote Sensing, 10*(11), 1698. https://doi.org/10.3390/rs10111698 (Basel, Switzerland).

Kim, G. G., Choi, J. H., Park, S. Y., Bhang, B. G., Nam, W. J., Cha, H. L., et al. (2019). Prediction model for PV performance with correlation analysis of environmental variables. *IEEE Journal of Photovoltaics, 9*(3), 832–841. https://doi.org/10.1109/JPHOTOV.2019.2898521

Kim, J., & Nam, Y. (2019). Study on the cooling effect of attached fins on PV using CFD simulation. *Energies, 12*(4), 758. https://doi.org/10.3390/en12040758

Koh, E.C. (2022). Population trends 2022. Retrieved 19 June 2023 from https://www.singstat.gov.sg/publications/population/population-trends.

Li, P., Zhang, H., Guo, Z., Lyu, S., Chen, J., Li, W., et al. (2021). Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning. *Advances in Applied Energy, 4*, Article 100057. https://doi.org/10.1016/j.adapen.2021.100057

Liang, Y., Saner, C. B., Siew, H. Q. J., Lee, H. H. L., Vanmathi, D. O. P., Tan, X. W., et al. (2022). Optimal configuration and economic-environmental analysis of floating photovoltaic: An empirical case study in Singapore. In *Proceedings of the IEEE sustainable power and energy conference (iSPEC)* (pp. 1–5). https://doi.org/10.1109/iSPEC54162.2022.10033039

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18–22. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768–4777).

Majumder, A., Setia, R., Kingra, P. K., Sembhi, H., Singh, S. P., & Pateriya, B. (2021). Estimation of land surface temperature using different retrieval methods for studying the spatiotemporal variations of surface urban heat and cold islands in Indian Punjab. *Environment, Development and Sustainability, 23*(11), 15921–15942. https://doi.org/10.1007/s10668-021-01321-3

Masrur Ahmed, A. A., Deo, R. C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., et al. (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*, Article 126350. https://doi.org/10.1016/j.jhydrol.2021.126350 (Amsterdam).

May Tzuc, O., Bassam, A., Mendez-Monroy, P. E., & Dominguez, I. S. (2018). Estimation of the operating temperature of photovoltaic modules using artificial intelligence techniques and global sensitivity analysis: A comparative approach. *Journal of Renewable and Sustainable Energy, 10*(3), Article 033503. https://doi.org/10.1063/1.5017520

McCarthy, M. P., Best, M. J., & Betts, R. A. (2010). Climate change in cities due to global warming and urban effects. *Geophysical Research Letters*, (9), 37. https://doi.org/10.1029/2010GL042845

Morales Pedraza, J. (2022). Chapter 3 - solar energy for electricity generation. J. Morales Pedraza. *Non-conventional energy in North America* (pp. 137–174). Elsevier. https://doi.org/10.1016/B978-0-12-823440-2.00006-8.

Motahhir, S., El Hammoumi, A., & El Ghzizal, A. (2018). Photovoltaic system with quantitative comparative between an improved MPPT and existing INC and P&O methods under fast varying of solar irradiation. *Energy Reports, 4*, 341–350. https://doi.org/10.1016/j.egyr.2018.04.003

National Environment Agency. (2023). Realtime weather readings across Singapore. Retrieved 19 June 2023 from https://data.gov.sg/dataset/realtime-weather-readings.

National Oceanic and Atmospheric Administration. (2005). What is the heat index? National oceanic and atmospheric administration. Retrieved 19 June 2023 from https://www.weather.gov/ama/heatindex.

Peng, X., Wu, W., Zheng, Y., Sun, J., Hu, T., & Wang, P. (2020). Correlation analysis of land surface temperature and topographic elements in Hangzhou, China. *Scientific Reports, 10*(1), 10451. https://doi.org/10.1038/s41598-020-67423-6

Qin, Z., Dall'Olmo, G., Karnieli, A., & Berliner, P. (2001). Derivation of split window algorithm and its sensitivity analysis for retrieving land surface temperature from NOAA-advanced very high resolution radiometer data. *Journal of Geophysical Research, 106*(D19), 22655–22670. https://doi.org/10.1029/2000JD900452

Rana, M., Koprinska, I., & Agelidis, V. G. (2016). Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Conversion and Management, 121*, 380–390. https://doi.org/10.1016/j.enconman.2016.05.025

Rashidi, S., Esfahani, J. A., & Hosseinirad, E. (2021). Assessment of solar chimney combined with phase change materials. *Journal of the Taiwan Institute of Chemical Engineers, 124*, 341–350. https://doi.org/10.1016/j.jtice.2021.03.001

Sekertekin, A., & Bonafoni, S. (2020). Land surface temperature retrieval from Landsat 5, 7, and 8 over rural areas: Assessment of different retrieval algorithms and emissivity models and toolbox implementation. *Remote Sensing, 12*(2), 294. https://doi.org/10.3390/rs12020294

Singapore Land Authority. (2023). Total land area of Singapore. Retrieved 19 June 2023 from https://data.gov.sg/dataset/total-land-area-of-singapore.

Skoplaki, E., & Palyvos, J. A. (2009). On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Solar Energy, 83*(5), 614–624. https://doi.org/10.1016/j.solener.2008.10.008

Sobrino, J. A., Jiménez-Muñoz, J. C., & Paolini, L. (2004). Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment, 90*(4), 434–440. https://doi.org/10.1016/j.rse.2004.02.003

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(1), 25. https://doi.org/10.1186/1471-2105-8-25

Townshend, J. R. G., & Justice, C. O. (1986). Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing, 7*(11), 1435–1445. https://doi.org/10.1080/01431168608948946

Trinasolar. (2022). Datasheet of vertex bifacial dual glass monocrystalline module (TSM-DEG21C.20). Retrieved 19 June 2023 from https://static.trinasolar.com/.

United States Geological Survey. (2023). Earth explorer. Retrieved 19 June 2023 from https://earthexplorer.usgs.gov/.

Wang, M., Cui, Q., Sun, Y., & Wang, Q. (2018). Photovoltaic panel extraction from very high-resolution aerial imagery using region–line primitive association analysis and template matching. *ISPRS Journal of Photogrammetry and Remote Sensing, 141*, 100–111. https://doi.org/10.1016/j.isprsjprs.2018.04.010

Wong, M. S., Zhu, R., Liu, Z., Lu, L., Peng, J., Tang, Z., et al. (2016). Estimation of Hong Kong's solar energy potential using GIS and remote sensing technologies. *Renewable Energy, 99*, 325–335. https://doi.org/10.1016/j.renene.2016.07.003

World Weather Online. (2023). Historical marine, sailing and surfing weather api. Retrieved 19 June 2023 from https://www.worldweatheronline.com/.

Yan, L., Zhu, R., Kwan, M. P., Luo, W., Wang, D., Zhang, S., et al. (2023). Estimation of urban-scale photovoltaic potential: A deep learning-based approach for constructing three-dimensional building models from optical remote sensing imagery. *Sustainable Cities and Society, 93*, Article 104515. https://doi.org/10.1016/j.scs.2023.104515

Yu, X., Guo, X., & Wu, Z. (2014). Land surface temperature retrieval from landsat 8 TIRS-comparison between radiative transfer equation-based method, split window algorithm and single channel method. *Remote Sensing, 6*(10), 9829–9852. https://doi.org/10.3390/rs6109829

Zhang, J., Xu, L., Shabunko, V., Tay, S. E. R., Sun, H., Lau, S. S. Y., et al. (2019). Impact of urban block typology on building solar potential and energy use efficiency in tropical high-density city. *Applied Energy, 240*, 513–533. https://doi.org/10.1016/j.apenergy.2019.02.033

Zhong, T., Zhang, Z., Chen, M., Zhang, K., Zhou, Z., Zhu, R., et al. (2021). A city-scale estimation of rooftop solar photovoltaic potential based on deep learning. *Applied Energy, 298*, Article 117132. https://doi.org/10.1016/j.apenergy.2021.117132

Zhu, R., Guo, D., Wong, M. S., Qian, Z., Chen, M., Yang, B., et al. (2023b). Deep solar PV refiner: A detail-oriented deep learning network for refined segmentation of photovoltaic areas from satellite imagery. *International Journal of Applied Earth*

Observation and Geoinformation, 116, Article 103134. https://doi.org/10.1016/j.jag.2022.103134

Zhu, R., Kwan, M. P., Perera, A. T. D., Fan, H., Yang, B., Chen, B., et al. (2023a). GIScience can facilitate the development of solar cities for energy transition. *Advances in Applied Energy, 10*, Article 100129. https://doi.org/10.1016/j.adapen.2023.100129

Zhu, R., You, L., Santi, P., Wong, M. S., & Ratti, C. (2019). Solar accessibility in developing cities: A case study in Kowloon East, Hong Kong. *Sustainable Cities and Society, 51*, Article 101738. https://doi.org/10.1016/j.scs.2019.101738