Contents list available at CBIORE journal website

International Journal of Renewable Energy Development

Journal homepage: https://ijred.cbiore.id

Research Article

Clustering-based assessment of solar irradiation and temperature attributes for PV power generation site selection: A case of Indonesia's Java-Bali region

Yusak Tanoto[a], Gregorius Satia Budhi[b*], Sean Frederick Mingardi[b]

[a]*Electrical Engineering Department, Faculty of Industrial Technology, Petra Christian University, Indonesia*

[b]*Informatics Department, Faculty of Industrial Technology, Petra Christian University, Indonesia*

Abstract. This study presents clustering-based assessments of solar attributes for locating potential solar photovoltaic (PV) power plant sites using *k*-means and density-based spatial clustering of applications with noise (DBSCAN) by examining the yearly average single-attribute and three-attribute clustering on a dataset of long-term hourly-based direct and diffuse irradiation, ambient temperature, and solar PV power output from 2005 to 2022. Three-attribute clustering enables stakeholders to better understand the characteristics of a cluster by collectively identifying three solar attributes and the magnitude of each attribute in an area or cluster. The presence of this information, which constitutes the clusters, suggests that these attributes have different effects on solar PV output power in different clusters. Although *k*-means is an effective method for investigating potential locations for PV power plant placements, DBSCAN offers users an alternative method for accomplishing a similar goal. In the case of three-attribute clustering of direct irradiation with *k*-means and DBSCAN, the 18-year mean value of clusters with the highest yearly average value is achieved at very similar values of 0.305 kW/m^2 and 0.310 kW/m^2, respectively. It turns out that only six years of direct irradiation had an annual mean value of less than 0.305 kW/m^2. This finding implies that in the long run, the solar resources in terms of direct irradiation will typically surpass 0.3 kW/m^2/MW installed capacity over all areas suitable for PV power plants. While focusing on the Java-Bali region, Indonesia, the findings, and methods appear to be of broader interest to policymakers, particularly in developing countries where solar PV is considered an option for sustainable energy generation.

Keywords: Clustering, solar PV, site selection, solar attributes, long-term

Received: 5th Dec 2023; Revised: 18th Feb 2024; Accepted: 29th Feb 2024; Available online: 3rd March 2024

1. Introduction

Solar photovoltaics (PV) are one of the most technologically and commercially promising renewable energy-based technologies driving the transition from heavy reliance on fossil fuels to the widespread use of green energy (Shubbak, 2019). Solar PV has shown the fastest growth in power capacity addition among other renewable energy-based technologies worldwide over the last decade from 2013–2022 (IRENA, 2023). Solar PV capacity is expected to exceed the capacity of coal by 2027, with utility-scale solar PV being the least expensive option among PV system configurations (IEA, 2023). Consequently, higher solar penetration is expected in studies on generation capacity planning scenarios, based on the opinions of electricity industry planners and policymakers.

Africa and Asia have some of the highest solar PV power generation potential (Tlhalerwa, Mulalu, 2019; WEF, 2023). However, despite the excellent solar resources in many developing countries, particularly in Africa and Asia, the need to address numerous challenges and barriers other than technological issues affects investment decision-making (Sreenath *et al.*, 2022; Laldjebaev *et al.*, 2021), and the capacity addition of utility-scale solar PV in many of these countries and territories has been slow. Although databases containing global solar attribute data are now freely accessible online, efforts to incorporate higher solar penetration into the power generation supply mix still necessitate selecting an appropriate dataset, followed by adequate assessment and specific evaluation up to a certain level based on needs.

To date, many studies have been conducted to assess the technical potential of solar PV to make investment decisions. Such high-level studies frequently examine the temporal variability of solar resource datasets from various regions. In this regard, the solar PV output temporal dataset is critical; for example, assessing potential PV contributions in electricity industry generation capacity planning (Tanoto *et al.*, 2020) or analyzing the reliability-cost trade-offs for electricity industry planning with high PV penetrations in emerging economies (Tanoto *et al.*, 2021). Understanding the long-term temporal variability of solar PV modelling is crucial for system planners and operators who analyze system ramp-up and ramp-down capabilities, particularly at high solar penetration levels (Kumar *et al.*, 2022; Tanoto *et al.*, 2017).

Clustering methodologies have been used to identify and classify similar solar attributes within a given temporal and/or spatial coverage (Gutierrez *et al.*, 2022). Scholars have revealed a meaningful and manageable distribution of solar attributes

* Corresponding author
Email: greg@petra.ac.id (G. S.Budhi)

https://doi.org/10.61435/ijred.2024.59998
ISSN: 2252-4940/© 2024.The Author(s). Published by CBIORE

within a region or jurisdiction by clustering the solar attributes of that region, which can make the exploration of solar PV potential more efficient (Zagouras *et al.*, 2014). Clustering methods have been used to evaluate PV system performance (Mahmoud, Gan, 2022; Tsafarakis *et al.*, 2018), assess PV power output patterns (Munshi, Mohamed, 2016), study the energy trilemma of future electricity industry generation with high PV penetration (Tanoto *et al.*, 2020), and forecast PV power output or solar resources (Azimi *et al.*, 2016; Wang *et al.*, 2016; Cheng *et al.*, 2016; Feng *et al.*, 2019; Fu *et al.*, 2019). Although clustering analysis is widely used in the study of solar PV potential, its application in assessing suitable PV power plant locations at specific geographical boundaries is limited.

A combination of affinity propagation and a k-means clustering algorithm was used to select candidate sites for PV power plants, considering solar variability and geographic smoothing effects (Zagouras *et al.*, 2014). The authors created and validated regions of coherent solar quality clustering maps for Lanai Island, Hawaii, based on 15 years of 30-minute temporal global horizontal irradiance (GHI) gridded data from satellite imagery. Although the study provided important insights into selecting locations for PV power plants, it mapped the GHI as the only factor for clustering analysis.

A combination of analytical hierarchical Process (AHP) and density-based clustering (DBC) using a geographic information system (GIS) was used to select suitable sites for solar farms In Ghana (Agyekum *et al.*, 2021). The authors identified three macro-clusters with varying total suitable areas and mean annual GHI values. Although the methodology used in this study appears to be appropriate for developing countries interested in deploying additional solar PV capacity into their national energy mix, the authors only used the yearly average values of the GHI as a representative parameter of solar resources, while considering other GIS-related parameters such as natural and social hazards, land use coverage, major settlement areas, protected areas, economic sustainability, and infrastructure proximity.

A combination of fuzzy c-means and an improved particle swarm optimization algorithm (IDPSO-FCM) was proposed to determine the applicable geographical regions of PV modules based on the regional clustering of environmental factors (Li *et al.*, 2022). The authors evaluated the consistency of the field reliability of the PV modules by considering the several workload and natural wear dimensions, such as solar radiation, precipitation, wind speed, temperature, relative humidity, and dust, along with cycling temperature and UV radiation. The authors used data from 2010 to 2019 to apply the methods to 31 provinces in mainland China. Although this methodology may be useful for assessing suitable PV plant sites based on the field reliability in different climatic regions, the study did not explicitly define the temporal granularity of the required data. Furthermore, numerous factors were involved in the analysis, including those related to the degradation of module performance, which are difficult to determine in developing countries.

Previously, a similar regional clustering analysis study was conducted to predict the field reliability and service lifetime of PV modules in 31 Chinese provincial regions (Liu *et al.*, 2020). The authors predicted the average annual power degradation of PV modules in different cluster regions by focusing on the comprehensive influence of various factors on the field reliability of PV modules. The clustering model described in (Li *et al.*, 2022) is based on various workloads and natural wear factors.

Pakistan's solar energy potential in terms of solar farm size was evaluated in three stages: solar mapping, zone evaluation, and regional clustering. (Amjad, 2020). The authors used QGIS and R to perform solar mapping and zone evaluation using the Energy Sector Management Assistance Program (ESMAP) dataset for solar irradiance, whereas the density-based spatial clustering of applications with noise (DBSCAN) package in R software was used for clustering. The authors have demonstrated that the methodology is adaptable to any spatial setting. However, the study only analyzed yearly direct normal irradiance (DNI) and did not specify the data duration.

Another study used the First-Generation Original Anatolian Honeybees' Investment Decision Support methodology (1GOAHIDSM) to develop a framework for selecting and clustering site-selection factors for PV installations (Saracoglu *et al.*, 2018). The authors established the framework by merging grey systems theory with fuzzy logic theory.

A dynamic hierarchical clustering algorithm was used to characterize and evaluate the long-term spatiotemporal GHI variability (Salguero *et al.*, 2022). The authors used three-stage methods using hierarchical cluster analysis to address the drawbacks of static clustering modelling of the GHI dataset, using Spanish territory as a case study, 22-year period data, and numerous observations from an online satellite database. The authors also examined and evaluated the temporal variation in monthly and annual average clusters. Although the proposed methodology appears to be applicable to the evaluation of solar attributes in other jurisdictions, the focus of this study is on long-term GHI data clustering.

Although previous clustering studies have provided useful insights into possible approaches for selecting suitable locations for solar PV plant deployment, contributions involving long-term temporal periods and different solar attributes to explore PV power output variability are generally minor. There is a scarcity of literature that discusses long-term multi-attribute clustering of solar energy resources for potential solar PV plant site selection.

The detection of variations in annual and monthly clustering results can be extremely useful for planners and policymakers in formulating appropriate strategies, leading to the optimal allocation of investments in solar PV power plant deployment within the study area because of the potential impact of PV integration into the grid. This research contributes to enhancing the existing clustering studies on solar PV power plant site selection by incorporating three solar attributes that influence solar PV plant power output. This research uses hourly temporal-based direct irradiation, diffuse irradiation, and ambient temperature as clustering factors rather than GHI or DNI values alone.

The primary goal of this research is to present an early-stage methodology for evaluating potential solar PV generation sites that can be used by planners and policymakers in developing countries or jurisdictions aiming to add more solar PV capacity to their electricity generation portfolios. This paper focuses on the long-term variability of direct and diffuse irradiation as well as temperature, which affects the PV power output across a large spatial boundary. While focusing on the Java-Bali region of Indonesia, all datasets used in this study are freely available online for cases from other countries and jurisdictions.

2. Case study and dataset

This study considers all areas of the Java-Bali region in Indonesia as a case study. The Java-Bali region has the largest interconnected electricity grid in Indonesia. It is also the country's economic powerhouse, with more than 60% of the total population and consumes approximately 70% of the total national electricity generation. As shown in Fig. 1 (Wikipedia, 2023), the region contains several provinces, from west to east,

Fig. 1 Map of Indonesia (top) and the Java-Bali region (down).

including Banten, Jakarta, West Java, Central Java, Yogyakarta, East Java, and Bali. This study uses long-term hourly temporal-based datasets of the solar PV output model from 2005–2022 obtained through API access on the Renewables Ninja (RN) website (Renewables.ninja, 2023; Pfenninger, Staffell, 2016) for clustering analysis.

The dataset includes hourly temporal-based solar attributes of direct irradiation, diffuse irradiation, and ambient temperature based on NASA MERRA2 satellite images, in addition to modelled hourly solar PV output data. This study calculated the solar PV plant output at each location using a 1 MW solar PV capacity to obtain modelled solar PV output data as well as the three solar attributes. This study employed gridded rectangular geographical coordinates of 0.05° latitude and longitude, or 5 km × 5 km, around the Java-Bali region, as shown in Fig. 1 (in red). In one year, there were 8,760 hours of data for each attribute at each location. Data preprocessing was performed to reduce the size of the dataset. The data preprocessing steps are as follows.

1) Exclude any data before 9 a.m. and after 3 p.m. in all locations.
2) Exclude locations outside the Java-Bali land border, such as the sea.
3) Calculate the annual average values of each attribute at each location based on the hourly data obtained in Steps 1 and 2.

Following data preprocessing, 4,510 locations along the Java-Bali land border were identified, each with a single solar attribute value (direct irradiation, diffuse irradiation, temperature, and solar PV output).

3. Methods

This study uses *k*-means and DBSCAN clustering to obtain areas with different clustering attributes and compare the results between the two methods. This study applies single-attribute and three-attribute clustering approaches under the two methods to examine solar attributes individually and collectively, using both clustering techniques in terms of the yearly average values of hourly temporal-based direct irradiation, diffuse irradiation, and temperature.

Moreover, this study performs single-attribute clustering on the solar PV power output dataset obtained from a solar PV output model within the same study area and compares the results with those obtained from three-attribute clustering. The clustering approach is demonstrated using a 2005–2022 long-term dataset of gridded hourly temporal-based direct and diffuse irradiation, ambient temperature, and solar PV output for the Java-Bali region of Indonesia.

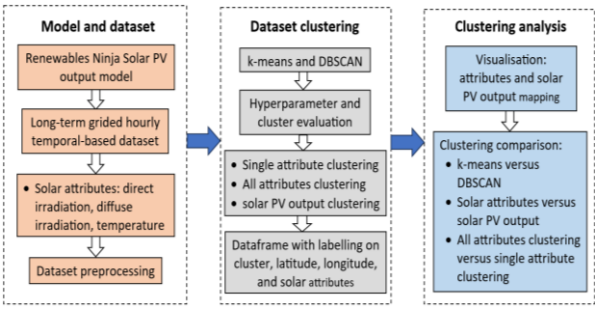

Fig. 2 Framework for assessing solar attributes clustering.

3.1 The methodological framework

This study develops a framework for assessing potential solar PV plant sites in the Java-Bali region of Indonesia by clustering areas based on three solar attributes, as described in Section 2, either individually or together, in one analysis. The framework employs the *k*-means and DBSCAN clustering techniques to obtain clusters. In this study, the clustering results of the solar attributes were obtained using these two techniques, the clustering results of the solar PV output, and the clustering results of the three attributes clustered together. Fig. 2 depicts the conceptual framework of the study.

3.2 K-means and DBSCAN clustering

K-means is one of the unsupervised machine learning (ML) algorithms for data clustering. *K*-Means divides the data into k groups or clusters. The algorithm is intended to group data with a high degree of similarity and separate data with significant differences.

The data are clustered using k-means technique according to the following steps:

1) Select the number of clusters (*k*), which indicates the number of clusters initiated at the start of the clustering process.
2) Determine the centroids by selecting *k* random points from the data or using the k-means++ algorithm (Arthur and Vassilvitskii, 2007) to select the initial values to avoid changing centroids owing to random picking.
3) All data points (or locations) are grouped to the closest centroids by calculating the distance of each data point to the centroid. Each data point is grouped into clusters with the closest centre using the Euclidean distance formula (Faizah, 2020).
4) Recalculate the centroids of newly formed clusters (cluster centres updation).
5) Repeat steps 3 and 4, or re-assign data points, until the computation reaches the maximum number of iterations.

Fig. 3 (Guidotti, 2019) shows the visual description of the clustering steps with *k*-means, given the number of clusters (*k*) equal to 2. DBSCAN is also an unsupervised ML algorithm for data clustering. The DBSCAN algorithm aims to group data based on spatial density, which enables the grouping of data with high density in the same area while ignoring data with low

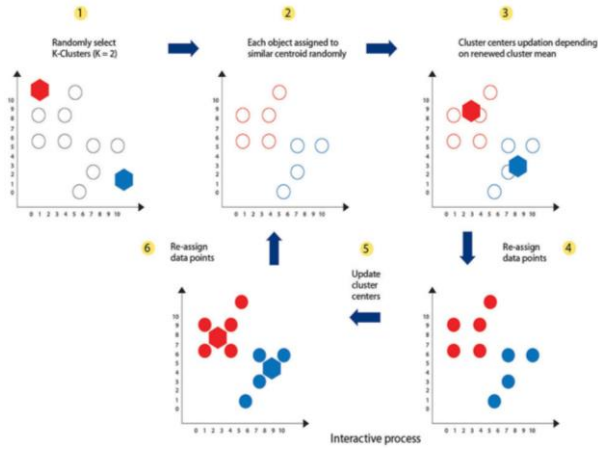

Fig. 3 Example of k-means clustering process with k = 2.

density or as 'noise' (Ester *et al.*, 2023). The algorithm employs parameters such as core points, border points, noise points, and direct density-reachable and density-connected points.

Brief definitions of these parameters are provided below.

- Core points are those inside a cluster with a specific number of neighbours within a predetermined radius, also known as *epsilons* (*Eps*). The minimum number of neighbours within the *Eps* radius is known as *MinPts*. The core points have the same number or more MinPts within the *Eps* radius.
- Border points have fewer *MinPts* than neighbours in the *Eps* radius. These points are on the border of the cluster and therefore remain within the *Eps* radius of the cluster.
- The noise points have insufficient *MinPts* in the *Eps* and are outside the *Eps* radius of the other core points. This indicates that these points are not associated with any group.

Fig. 4 (Farahnakian *et al.*, 2023) depicts the visual description of the DBSCAN clustering process with *MinPts* equals 3. It particularly shows a formed cluster, noise points, border points, *Eps*, and *MinPts* of a DBSCAN process.

3.3 Davies-Bouldin index and Silhouette score

The Davies–Bouldin (DB) index (Davies, Bouldin, 1979) is a validation metric for clustering models and other unsupervised machine learning algorithms. The index was calculated by measuring the average similarity of each cluster and comparing it with the most similar cluster. In the DB index, similarity is defined as the ratio of inter-cluster and intra-cluster distances. This index assigns a better score (i.e., closer to zero) to well-separated clusters with less dispersion. This study searched for the best clustering results for each year mainly based on the DB index.

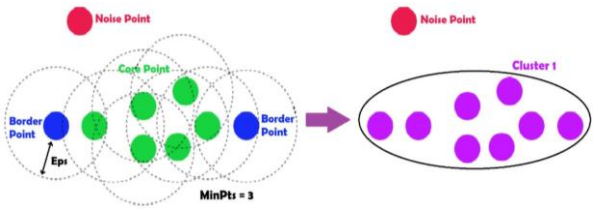

Fig. 4 Example of DBSCAN clustering process with *MinPts* = 3.

The Silhouette score (SS) is an evaluation metric used in cluster analysis and other unsupervised machine learning algorithms. The SS assesses the quality of clustering by comparing the data in one cluster with those in other clusters. The Silhouette score ranges from −1 to 1, with 1 indicating better-defined and separated clusters (Shahapure, Nicholas, 2020).

A value approaching 1 indicates that the data in the formed clusters approach each other within the cluster and are distinct or well separated from the other clusters. A value close to zero indicates that the data in that cluster are nearly identical to those in other clusters. A value close to −1 indicates that data in one cluster cannot be in that cluster and must be in another cluster.

3.4 Single-attribute and three-attribute clustering

Single-attribute and three-attribute clustering analyses were performed independently for each year from 2005–2022 using both *k*-means and DBSCAN techniques. In single-attribute clustering, the steps presented in Fig. 3 and Fig. 4 are used to cluster direct irradiation, diffuse irradiation, temperature, and PV output separately.

The analysis begins by retrieving the pre-processed datasets, which include the yearly average values of direct irradiation, diffuse irradiation, temperature, and PV output data for all locations. The region for the retrieved dataset is then specified, as shown in Fig. 5, followed by hyper-parameter search and clustering analysis using k-means and DBSCAN. Finally, the Silhouette score and Davies-Bouldin index are used to estimate the optimal cluster number.Three-attribute clustering seeks to generate cluster areas that contain direct irradiation, diffuse irradiation, and temperature in a single clustering process. The stages for three-attribute clustering are the same as for single-attribute clustering, with the addition of a step to normalise the data using the Minmax Scaler (Scikit-learn, 2023) before clustering. Due to the significant discrepancy in the magnitude of values between the irradiation and temperature datasets, data normalisation is required to scale the values of irradiations and temperature.

3.5 Hyper-parameters

Hyper-parameters have values that cannot be estimated directly from the training process and must therefore be specified before training (Yang, Shami, 2020). Hyper-parameters are frequently used in machine learning applications to develop specific models or algorithms (Howland, 2006). The *k*-means clustering has one hyperparameter, the number of clusters k (Sinaga, Yang, 2020), while the DBSCAN algorithm has two parameters: *Eps* and *MinPts* (Wiharto *et al.*, 2021).

This study searches for the k-means hyper-parameter by trying *k*-values in succession, starting from 2 up to 20 and evaluating the clustering performance using the DB index. If the current DB index is higher than the previous index, the current k is saved and used in the clustering analysis; otherwise, the previous k is stored and used along with the corresponding DB

Fig. 5 A defined Java-Bali region of the retrieved dataset.

index. For verification purposes, the SS was then calculated based on the best clustering result for each year.

In DBSCAN, *MinPts* is searched from 2 to 100, while *Eps* is searched from 0.001 to 1, with 0.001 increments in all clusters except for solar PV output clustering, which is up to 10. Appropriate *MinPts* and *Eps* values were selected by testing them using the DB index. The *MinPts* and *Eps* values with lower DB indexes were saved and used in the clustering stage.

4. Results and Discussions

4.1 K-means clustering results

The analyses using k-means presented the clustering results obtained from three-attribute clustering and single-attribute clustering while considering the optimal *k* values, that is, the optimal number of clusters applied to the dataset. Table 1 presents the optimal number of clusters k for k-means clustering from 2005–2022.

Except for the three-attribute clustering in 2018 and 2019, the DB index was relatively low each year (i.e., less than one). This indicated a relatively high dispersion of data points between clusters, implying that the similarity between clusters was generally low. In other words, the established k-means clusters had high intra-cluster similarity, but low inter-cluster similarity. In all years, a SS greater than zero indicates good inter-cluster dispersion of the data points. This indicates that the clusters were well separated and distinct. Small DB and SS standard deviations across all years indicate that each year's clustering results were consistent in terms of high intra-and low inter-cluster similarity, density, and dispersion.

Meanwhile, Fig. 6 to Fig 11 shows the results of the three-attribute clustering using the *k*-means technique. Using the pre-processed datasets of direct irradiation, diffuse irradiation, and temperature, *k*-means clustering generated areas or clusters with data points displaying a range of values for these attributes. As shown in Fig. 6, the three-attribute clustering in 2019 using *k*-means divided the Java-Bali region into four clusters, each with three attributes and their own range of values.

Table 1
Number of optimal clusters and evaluation metrics of *k*-means three-attribute and single-attribute clustering in 2005–2022

Year	Three-attribute clustering: direct irradiation, diffuse irradiation, temperature			Single-attribute clustering: Solar PV output		
	Number of *k* clusters	Evaluation metrics		Number of *k* clusters	Evaluation metrics	
		DB	SS		DB	SS
2022	3	0.91	0.41	2	0.49	0.64
2021	5	0.81	0.46	2	0.49	0.64
2020	6	0.76	0.47	2	0.51	0.62
2019	4	0.73	0.48	3	0.47	0.6
2018	2	1.03	0.4	2	0.55	0.59
2017	2	1.04	0.4	2	0.53	0.61
2016	4	0.75	0.44	3	0.53	0.58
2015	4	0.76	0.47	2	0.49	0.63
2014	2	0.75	0.52	2	0.51	0.62
2013	3	0.88	0.4	5	0.5	0.55
2012	4	0.73	0.47	4	0.49	0.57
2011	5	0.82	0.44	2	0.57	0.58
2010	6	0.84	0.42	2	0.46	0.65
2009	5	0.78	0.48	2	0.47	0.65
2008	4	0.69	0.5	2	0.57	0.58
2007	4	0.7	0.49	2	0.48	0.64
2006	5	0.75	0.46	2	0.52	0.62
2005	2	0.79	0.49	3	0.46	0.63

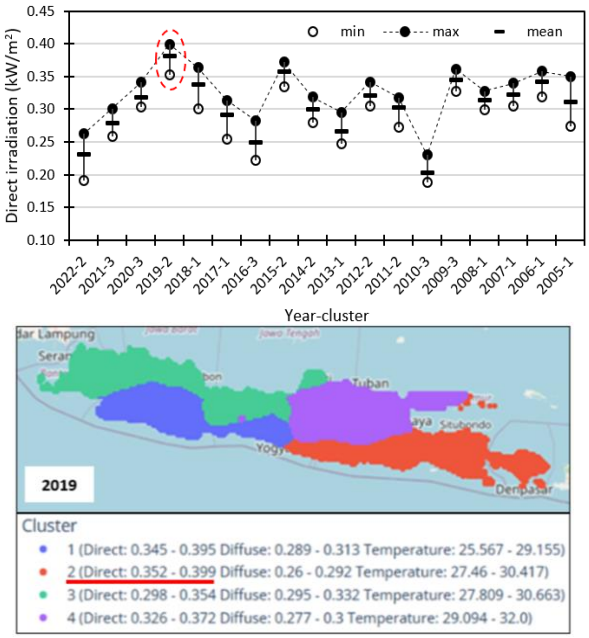

Fig. 6 Clustering results of direct irradiation obtained from the *k*-means three-attribute clustering: (Top) the years and clusters showing a range of values among clusters with the highest value of yearly average direct irradiation in each year; and (Down) the 2019 Java-Bali's clustering map showing the corresponding area (in Red).

Fig. 6 on the top shows the yearly average range of values in terms of the maximum, minimum, and mean among clusters, with the highest yearly average values of direct irradiation (in kW/m^2) in each year. The figure shows the 2005–2022 yearly average range of direct irradiation for every location assigned by a specific cluster. The highest range of values was obtained in 2019, particularly for cluster 2. Among the clusters with the highest value over the 18-year period, the figure also highlights fluctuations in terms of the highest yearly average of direct irradiation values from 2005–2022, with the lowest value obtained in the 2010-cluster 3 at 0.230 kW/m^2 and the highest one in the 2019-cluster at 0.399 kW/m^2.

In other words, there was a 0.169 kW/m^2 difference between the highest and lowest maximum of the yearly average direct irradiation value over the 18-year period. The down-side of Fig 6 shows the corresponding area with the highest annual average direct irradiation values in 2019 (in red). The area mostly includes the eastern and southern parts of East Java Province, Bali Province, and some islands in the north. Areas with the highest direct irradiation are potential locations for PV plants because they have the greatest potential for solar PV output power.

Like Fig 6, Fig 7 depicts the clustering results of direct irradiation obtained from the three-attribute clustering in terms of the yearly average range of values among clusters with the minimum values and the mapping areas of clusters in the Java-Bali region in 2010. The figure on the top shows the annual average range of direct irradiation from 2005–2022 for each location among the clusters with minimum values. The lowest minimum value of yearly average direct irradiation over the 18-year period was obtained in 2010, particularly in Cluster 5. The down-side of Fig 7 shows the corresponding area of Cluster 5 (orange), with a direct irradiation range of 0.122–0.161 kW/m^2, as shown in Fig 7 on the top. The area primarily depicts the western and northern parts of Java, with a small portion in the south comprising West Java, Jakarta, and Banten.

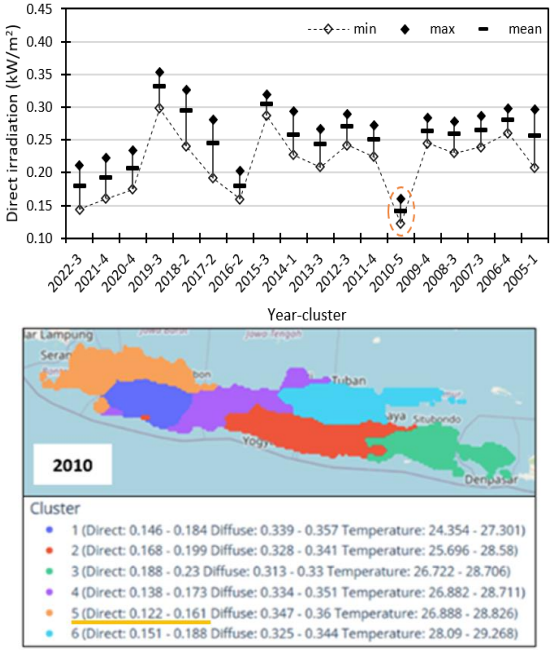

Fig. 7 Clustering results of direct irradiation obtained from the *k*-means three-attribute clustering: (Left) the years and clusters showing a range of values among clusters with the lowest value of yearly average direct irradiation in each year; and (Right) the 2010 Java-Bali's clustering map showing the corresponding area (in Orange).

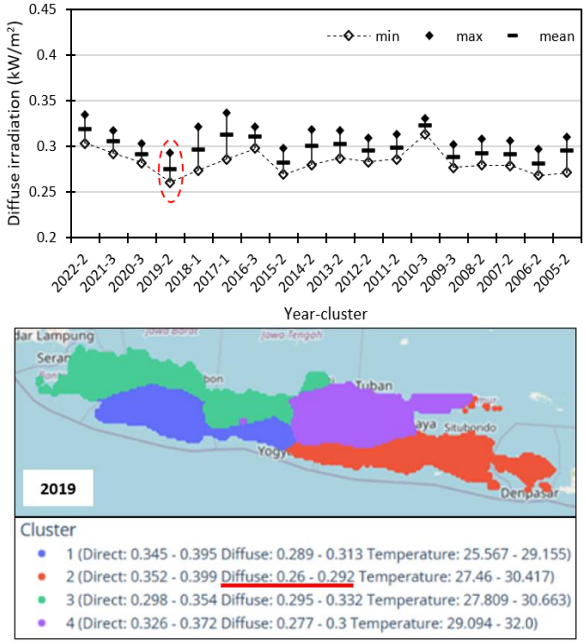

Fig. 9 Clustering results of diffuse irradiation obtained from the *k*-means three-attribute clustering: (Top) the years and clusters showing a range of values among clusters with the lowest value of yearly average diffuse irradiation in each year; and (Down) the Java-Bali's clustering map in 2010 showing the corresponding area (in Red).

Fig. 8 and Fig. 9 depict the clustering results of diffuse irradiation obtained from three-attribute clustering, with figures on the top depicting the maximum-mean-minimum ranges of the yearly average numerical results of clusters with the maximum values (in Fig. 8) and with the minimum values (in Fig. 9), and figures on the down-side visualising areas according to diffuse irradiation clusters in a corresponding year.

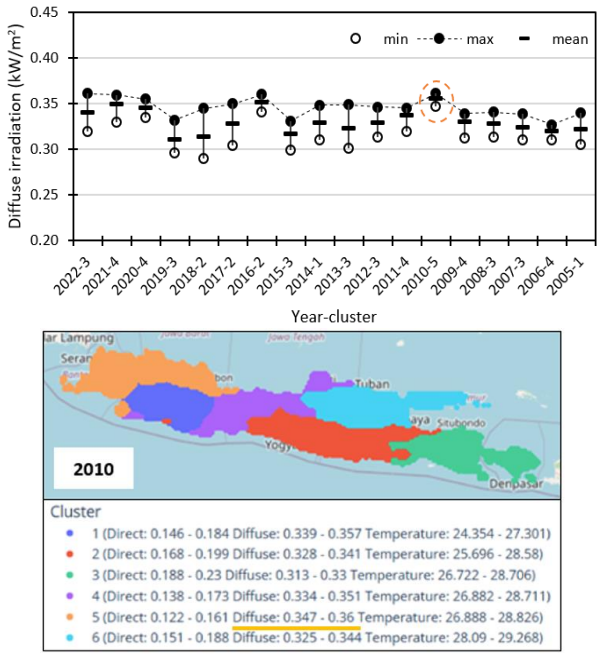

Fig. 8 Clustering results of diffuse irradiation obtained from the *k*-means three-attribute clustering: (Top) the years and clusters showing a range of values among clusters with the highest value of yearly average diffuse irradiation in each year; and (Down) the 2010 Java-Bali's clustering map showing the corresponding area (in Orange).

Fig. 8 shows that, over the 18-year period, the mean values of yearly average diffuse irradiation have mostly ranged between 0.32 kW/m^2 and 0.35 kW/m^2, except for 2010, which was slightly higher than this range. This finding indicates a relatively consistent annual weather situation in the Java-Bali region over the 18-year period, particularly in terms of cloud distribution across the region.

The highest annual average value of diffuse irradiation over the 18-year period was obtained in 2010, as shown in Fig. 8 on the top, with values ranging from 0.347 kW/m^2 to 0.36 kW/m^2. Located in the same area as the lowest range of the annual average direct irradiation value (see Fig. 7, down), this finding confirms a crucial relationship between direct and diffuse irradiation in a specific area or cluster, in terms of the typical opposite direction of these two attributes. While the western and northern parts of West Java, with a small portion in the south, received the lowest annual average direct irradiation in 2010, the same area received the highest annual average diffuse irradiation. This typical relationship between direct and diffuse irradiation across the Java-Bali region is confirmed in this study for most of the other years.

In comparison with the results shown in Fig. 6, Fig. 9 shows another example of this relationship. While the eastern and southern parts of East Java and Bali had the highest range of annual average direct irradiation, the same cluster had the lowest range of annual average diffuse irradiation, ranging from 0.26 kW/m^2 to 0.292 kW/m^2. A high diffuse irradiation for a specific area indicates a lower potential of PV output power in that specific area compared with another area with low diffuse irradiation, making that area less favourable for PV plant placement.

Fig. 10 and Fig. 11 on the top show the clustering results of the yearly average ambient temperature from to 2005–2022

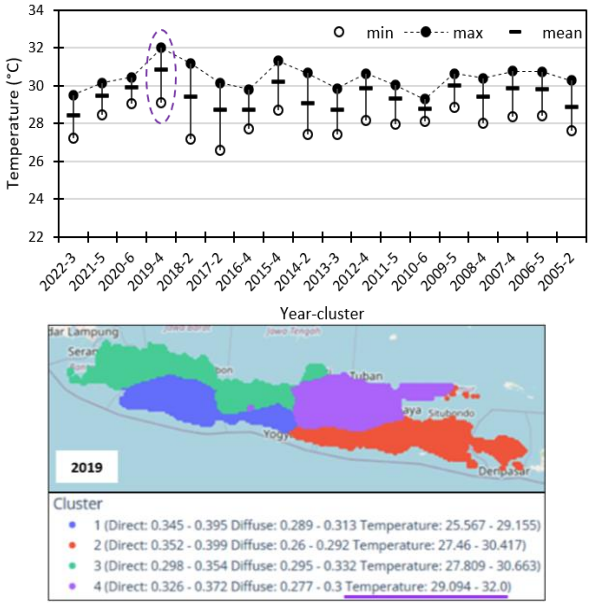

Fig. 10 Clustering results of ambient temperature obtained from the *k*-means three-attribute clustering: (Top) the years and clusters showing the range of values among clusters with the highest value of yearly average temperature in each year; and (Down) the Java-Bali's clustering map in 2010 showing the corresponding area (in Purple).

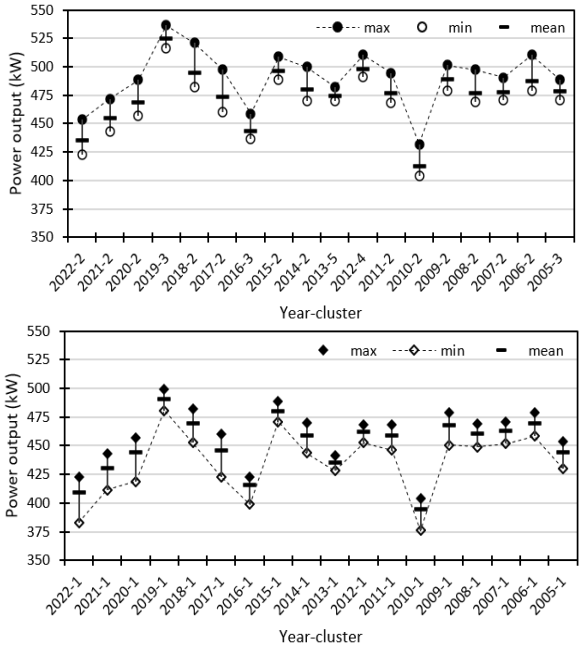

Fig. 12 Clustering results of solar PV output obtained from the *k*-means single-attribute clustering: (Top) the years and clusters displaying a range of yearly average output values among clusters with the highest PV outputs in each year; and (Down) the years and clusters showing a range of yearly average output values among clusters with the lowest PV outputs.

across all locations in the Java-Bali region, considering clusters with the highest and lowest yearly average temperature values, respectively. The figures on the down-side depict the cluster mapping for the corresponding years. The Java–Bali region, which is located near the equator, has a relatively high maximum annual average temperature during the day. From 2005–2022, the maximum annual average temperature reached 32 °C in 2019, and over this 18-year period, the temperature exceeded 30 °C, except in 2022 and 2010. The ambient temperature impacts the efficiency of a PV system.

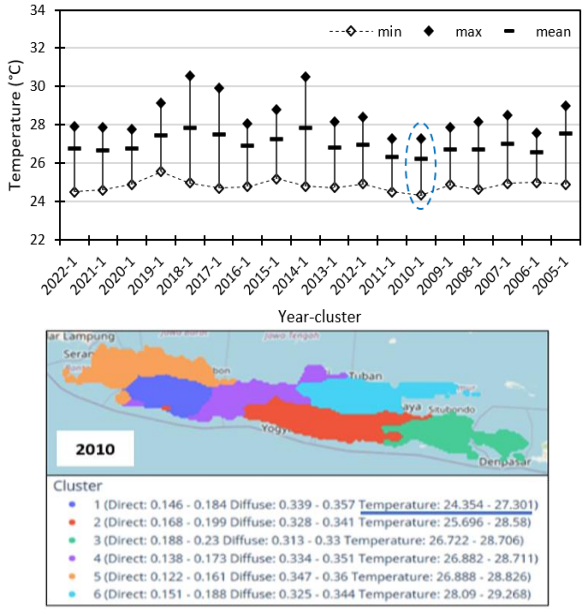

Fig. 11 Clustering results of ambient temperature obtained from the *k*-means three-attribute clustering: (Top) the years and clusters showing a range of values among clusters with the lowest value of yearly average temperature in each year; and (Down) the Java-Bali's clustering map in 2010 showing the corresponding area (in Blue).

While the rated efficiency of a PV module is commonly tested at 25 °C, higher ambient temperatures reduce the efficiency. Thus, policymakers and planners in the Java–Bali region can better estimate the potential of PV energy generation in different clusters by considering and comparing temperature attributes across clusters. Fig. 10 shows a 2019 map of the Java–Bali region with the highest yearly average temperature (cluster 4 in purple), ranging from 29.09 °C to 32.0 °C. The area encompasses the northern and western parts of the East Java Province (see Fig. 10, purple), indicating that the PV modules installed in that area may have lower efficiency. In 2010, the southern and eastern parts of West Java province had the lowest range of yearly average temperature in terms of minimum and mean values, reaching 24.3 °C and 26.3 °C, respectively (see Fig. 11, blue).

Fig. 12 depicts the single-attribute clustering results for solar PV output power in terms of yearly average values from 2005–2022. Fig. 12 shows the cluster with the highest output each year on the top, and the cluster with the lowest output on the down-side. Notably, the results in Fig. 12 follow a pattern like those shown in Fig. 6 and Fig. 7.

The results in Fig 12 indicate that the yearly average PV outputs in the Java-Bali region are substantially dependent on and proportional to direct irradiation. Meanwhile, minor changes in annual average diffuse irradiance and temperature appear to have minimal effect. Fig. 12 shows the historical range of the yearly average PV output power over the 18-year period, planners and policymakers could arguably anticipate a similar future range of potential PV output power for high-level planning, given the 375–530 kW/MW PV capacity, considering the lowest and highest yearly average output power between 2005 and 2022. In addition, the historical range of the yearly average PV output power shown in Fig. 12 implies the potential impact of PV plant

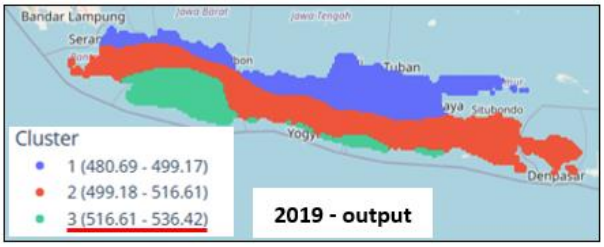

Fig. 13 Solar PV output clustering map that corresponds to the results shown in Figure 9: year and cluster with the highest range of yearly average PV output (2019-3).

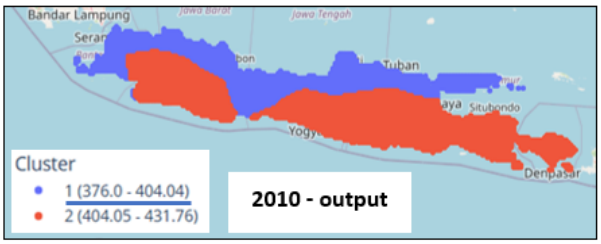

Fig. 14 Solar PV output clustering map that corresponds to the results shown in Figure 12: year and cluster with the lowest range of yearly average PV outputs (2010-1).

output on the annual ramp-up and ramp-down requirements of the system in the Java–Bali region.

Fig. 13 depicts the maps of the PV output power clustering corresponding to the results shown in Fig. 12. The years with the highest range in the annual average PV output were 2019, 2018, and 2006. The southern parts of Java, from west to east, and Bali appear to have greater potential than the northern part of the island. The northern part of Java had the lowest annual average PV output power in 2010, 2022, and 2016.

Fig. 13–15 show the single-attribute clustering results of direct irradiation, diffuse irradiation, and temperature, and the corresponding clusters of PV output power. In this context, planners and policymakers can further analyse and compare the clustering of each solar attribute year after year, learn the potential impact of the attributes on PV plant output power over a broad spatial coverage, and thus gain some insights into future generation planning decision-making.

As expected, the lowest range of yearly average diffuse irradiation, that is, in Cluster 1, ranging from 0.260–0.287 kW/m^2 (in blue), appeared to reduce the conversion effectiveness of

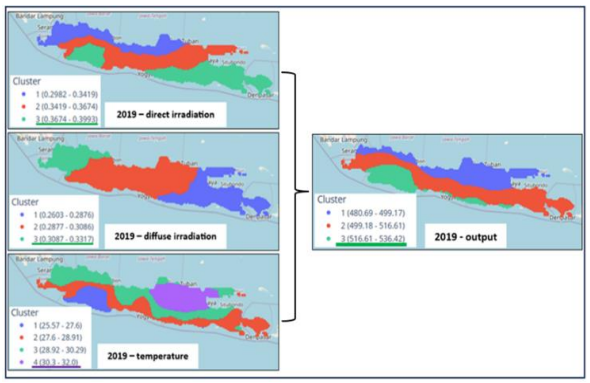

Fig. 15 *K*-means single-attribute clustering results of direct irradiation, diffuse irradiation, and temperature, and the corresponding clusters of PV output power in 2019.

direct irradiation values (in green) into PV output power in those areas. The affected areas are presented in the corresponding PV output clusters. Cluster 2 (red), particularly in the eastern and southern parts of East Java and Bali, is not a cluster with the highest yearly average PV output power.

As shown in Fig. 15, the southern part of Java Island and Bali are better locations for solar PV power generation than the northern part of Java, as resulted in the single-attribute clustering of solar PV output power. However, more detailed results from the three-attribute clustering suggest the most suitable locations across the region as the eastern and southern parts of East Java, the southern part of West Java, and Bali, owing to the comparatively highest range of annual average direct irradiation and the lowest range of diffuse irradiation, as well as moderate temperature ranges in comparison to other areas.

Therefore, implementing three-attribute clustering enables stakeholders to better understand the characteristics of a cluster by collectively presenting and identifying all attributes in an area or cluster. The presence of all three attributes, which compose the clusters, indicates that these attributes have different impacts on the solar PV output power in different clusters.

4.2 DBSCAN clustering results

DBSCAN analysis, like *k*-means three-attribute clustering, considers direct irradiation, diffuse irradiation, and temperature altogether in the clustering. Fig. 16 shows the range of yearly average direct irradiation over the 18-year period, focusing on clusters with the highest yearly average direct irradiation.

The clustering result shows a pattern similarity between DBSCAN and k-means, as shown in Fig. 6. DBSCAN also produced relatively similar areas or clusters as *k*-means, with the eastern and southern parts of East Java and Bali having the highest annual average values in 2019 (red). The presence of so-called noise clusters in the DBSCAN algorithm distinguishes

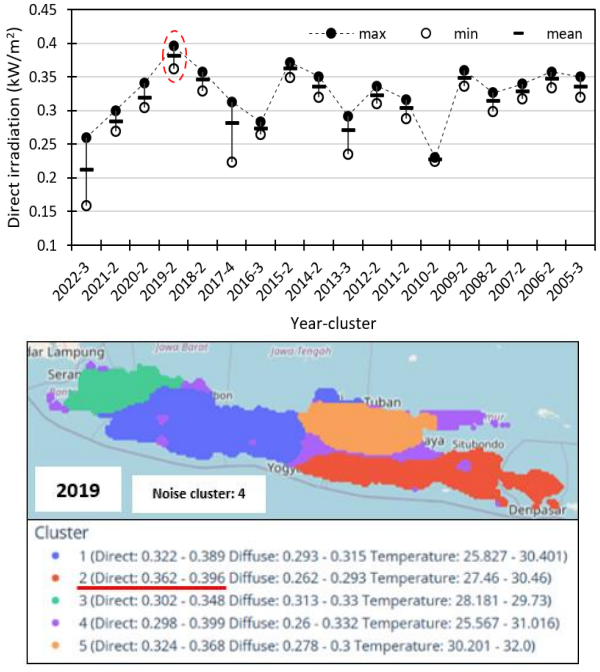

Fig. 16 Clustering results of direct irradiation obtained from the DBSCAN three-attribute clustering: (Top) the years and clusters showing a range of values among clusters with the highest value of yearly average direct irradiation in each year; and (Down) the 2019 Java-Bali's clustering map showing the corresponding area (in red).

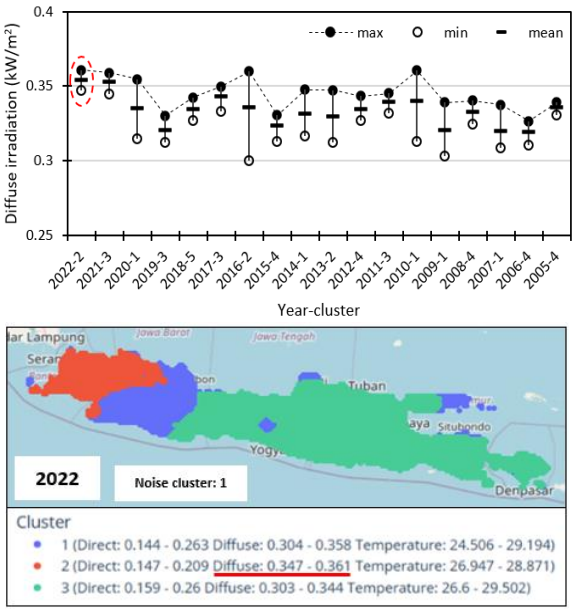

Fig. 17 Clustering results of diffuse irradiation obtained from the DBSCAN three-attribute clustering: (Top) the years and clusters showing a range of values among clusters with the highest value of yearly average direct irradiation in each year; and (Down) the 2019 Java-Bali's clustering map showing the corresponding area (in Red).

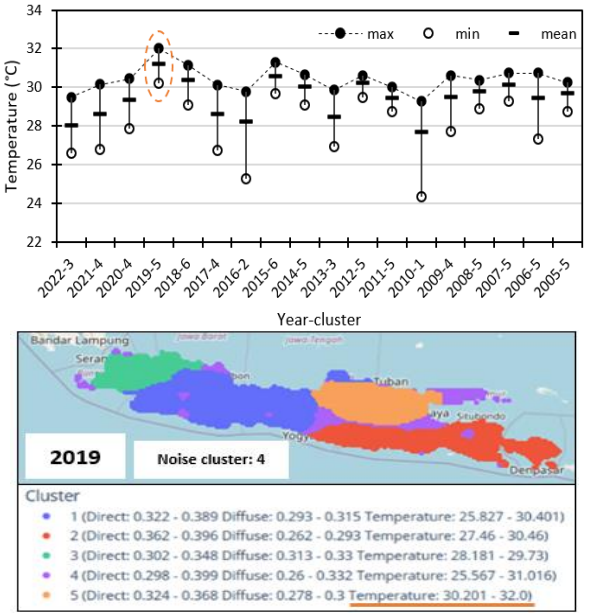

Fig. 18 Clustering results of direct irradiation obtained from the DBSCAN three-attribute clustering: (Top) the years and clusters showing a range of values among clusters with the highest value of yearly average direct irradiation in each year; and (Down) the 2019 Java-Bali's clustering map showing the corresponding area (red).

DBSCAN from *k*-means clustering. In 2019, noise clusters (purple) were visible in a few locations across Java. These clusters represent areas with less data density and thus separate neighbouring clusters. Here, the areas in purple separate other areas in red, orange, and blue.

The DBSCAN three-attribute clustering results for diffuse irradiation are shown in Fig. 17. The highest range of annual average diffuse irradiation is obtained in 2022 in Cluster 2. The top of Fig. 17 shows the corresponding area (in red) which includes Banten, Jakarta, the western parts of West Java, and its southern portion. While DBSCAN, like *k*-means, identified the western and northern parts of Java Island as having higher diffuse irradiation, the algorithm demonstrated a more year-to-year dynamic in terms of the range of clusters, with the highest yearly average diffuse irradiation. Owing to the nature of DBSCAN in clustering high-density data points, including the presence of noise clusters, some differences exist in the cluster size compared with *k*-means.

Finally, Fig. 18 shows the clustering results of the temperature obtained from DBSCAN three-attribute clustering. The top side of Fig 18 shows the highest range of yearly average temperatures among the clusters, with the maximum values of temperature in each year. Meanwhile, the down-side of Fig. 18 depicts a map of the Java–Bali region showing an area of the temperature cluster in the northern and north-western parts of East Java, that is, Cluster 5 shown in orange. Like diffuse irradiation, DBSCAN clustering produced a smaller area for the temperature cluster compared to that in k-means (see Fig. 10).

Clustering results reported in Section 4.1 (*k*-means) and Section 4.2 (DBSCAN) also help stakeholders understand the long-term potential of solar resources in the Java-Bali region. In this context, the 18-year mean value of the annual range of values of solar attributes produced from the clustering analyses can be recognised and used as one of the bases for electricity industry planning studies as well as a more extensive assessment of technological potential.

In the case of three-attribute clustering of direct irradiation with *k*-means and DBSCAN, the 18-year mean value of clusters with the highest yearly average value is achieved at very similar values of 0.305 kW/m^2 and 0.310 kW/m^2, respectively. It turns out that only six years of direct irradiation had an annual mean value of less than 0.305 kW/m^2. This finding implies that in the long run, the solar resources in terms of direct irradiation will typically surpass 0.3 kW/m^2/MW PV installed capacity over all areas suitable for solar PV power plant placement.

The results presented above, whether obtained from single- or three-attribute clustering, show that using *k*-means and DBSCAN methods allows planners and policymakers to further explore and study potential locations for placing solar PV plants across a large area or region. Although *k*-means appears to be an effective method for clustering the yearly average solar attributes, DBSCAN provides the user with an alternative method for achieving a similar goal.

The implementation of three-attribute clustering enables stakeholders to better understand the characteristics of a cluster by collectively identifying three solar attributes as well as the magnitude of each attribute in an area or cluster. The presence of all three attributes, as well as their magnitudes, which compose the clusters, indicates that these attributes have different impacts on the solar PV output power in different clusters. This research recognizes that planners and policymakers must make careful decisions on the appropriate location of solar PV power plants. Nevertheless, this research does not consider factors such as local ecosystems and technical information other than solar attributes.

5. Conclusion

This study presented the methodology and application of *k*-means and DBSCAN clustering techniques to investigate the potential of solar PV plant sites in the Java–Bali region of Indonesia using a satellite-based dataset of solar attributes. The analyses considered hourly temporal-based solar attributes,

such as direct irradiation, diffuse irradiation, and ambient temperature from 2005–2022. This study employed three-attribute clustering to obtain clusters containing the yearly average data ranges of the three solar attributes combined and single-attribute clustering to obtain clusters containing the yearly average data ranges for each attribute. The latter method was used to cluster the solar PV output power.

Based on the Davies–Bouldin index and Silhouette score evaluation metrics, the results show that both k-means and DBSCAN methods can be used to create clusters of yearly average solar attribute(s) across the Java–Bali region, including three-attribute clustering. The combinations of the three attributes in the clusters as well as their magnitudes were proportionally correlated with the solar PV output power in the corresponding clusters in the case of three-attribute clustering. Although *k*-means appears to be an effective method for clustering the yearly average solar attributes, DBSCAN provides the user with an alternative method for achieving a similar goal. Owing to the nature of DBSCAN in clustering high-density data points, including its ability to illustrate the presence of noise clusters, DBSCAN clustering produced different cluster sizes, especially smaller cluster areas compared to *k*-means, and demonstrated a more year-to-year dynamic in terms of the ranges of values in clusters.

Considering the 18-year period of the dataset, the single-attribute clustering results showed that the southern part of Java Island and Bali are better locations for solar PV power generation than the northern part of Java. More detailed results from the three-attribute clustering suggest the most suitable locations across the region as the eastern and southern parts of East Java, the southern part of West Java, and Bali, owing to the comparatively highest range of annual average direct irradiation and the lowest range of diffuse irradiation, as well as moderate temperature ranges in comparison to other areas.

In addition to obtaining clusters of solar attributes and PV output power for the Java–Bali region, the analyses in this study provide insights for planners and policymakers in planning the variable renewable energy electricity industry, particularly high-penetration solar PV, and especially regarding the long-term historical performance of solar resources and the potential impact of hourly temporal-based solar attributes and output power on system reliability and flexibility. Although the Java-Bali region of Indonesia was used as a case study, the method and findings appear to have broader relevance to stakeholders involved in solar PV-based electricity generation planning and policy-making communities considering their options for a more sustainable energy future.

As this study does not consider factors other than the three solar attributes that may influence the selection of PV location candidates, future work should seek to improve the methodology by incorporating these factors into the analysis and adding the methodology around daily to yearly forecasting of solar PV output power over a wide spatial coverage of the solar PV investment decision-making framework.

Acknowledgments

The authors would like to thank Dick Jovian and Jimlee Christanto Widjaya, both were students at Informatics Department, Petra Christian University, for helping with data collection.

Author Contributions: Y.T.: Conceptualization, Methodology, Formal analysis, Writing – original draft, Resources, Supervision, Funding acquisition, Writing - Review & Editing, G.S.B.: Conceptualization, Methodology, Software, Formal analysis, Validation, Supervision, Writing - Review & Editing, S.F.M.: Software, Validation, Visualization, Data Curation.

Funding: This research was funded by the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia, under the Competitive Fundamental Research Scheme Contract No. 183/E5/PG.02.00. PL/2023 and 002/SP2H/PT/LL7/2023, 11/SP2H/PT/LPPM-UKP/2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

Agyekum, E.B., Amjad, F., Shah, L., Velkin, V.I. (2021). Optimizing photovoltaic power plant site selection using analytical hierarchy process and density-based clustering – Policy implications for transmission network expansion, Ghana. *Sustainable Energy Technologies and Assessments* 47, 101521; https://doi.org/10.1016/j.seta.2021.101521

Amjad, F. & Shah, L. (2020). Identification and assessment of sites for solar farms development using GIS and density-based clustering technique – A case of Pakistan. *Renewable Energy* 155, 761-769; https://doi.org/10.1016/j.renene.2020.03.083

Arthur, D. & Vassilvitskii, S. (2007). *k-means++: the advantages of careful seeding*. In Proc of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics Philadelphia, PA, USA.

Azimi, R., Ghayekhloo, M., Ghofrani, M. (2016). A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. *Energy Conversion and Management* 118, 331-344; https://doi.org/10.1016/j.enconman.2016.04.009

Cheng, K., Guo, L.M., Wang, Y.K., Zafar, M.T. (2017). Application of clustering analysis in the prediction of photovoltaic power generation based on neural network. *IOP Conf. Ser.: Earth Environ. Sci.* 93, 012024; https://doi.org/10.1088/1755-1315/93/1/012024

Davies, D.L. & Bouldin, D.W. (1979). A cluster separation measure. *IEEE Transaction on pattern Analysis and Machine Intelligence*. PAMI-1(2), 224-227; https://doi.org/10.1109/TPAMI.1979.4766909

Ester M, Kriegel HP, Sander J, Xu X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. https://www.dbs.ifi.lmu.de/Publikationen/Papers/KDD-96.final.frame.pdf. Accessed on 10 September 2023.

Faizah, N.M., Surohman, Fabrianto, L., Hendra, Prasetyo, R. (2020). Unbalanced data clustering with k-means and euclidean distance algorithm approach case study population and refugee Data. *J. Phys.: Conf. Ser.* 1477, 022005; https://doi.org/10.1088/1742-6596/1477/2/022005

Farahnakian, F., Nicolas, F., Farahnakian, F., Nevalainen, P., Sheikh, J., Heikkonen, J., Baka, CR. (2023). A Comprehensive Study of Clustering-Based Techniques for Detecting Abnormal Vessel Behavior. *Remote Sensing* 15(6), 1447; https://doi.org/10.3390/rs15061477

Feng, C., Cui, M., Hodge, B.M., Lu, S., Hamann, H.F., Zhang, J. (2019). Unsupervised clustering-based short-term solar forecasting. *IEEE Transaction on Sustainable Energy* 10(4), 2174-2185; https://doi.org/10.1109/TSTE.2018.2881531

Fu, L., Yang, Y., Yao, X., Jiao, X., Zhu, T. (2019). A Regional photovoltaic output prediction method based on hierarchical clustering and the mRMR criterion. *Energies* 12, 3817; https://doi.org/10.3390/en12203817

Guidotti, E. (2019). *Text Mining*. https://storage.guidotti.dev/course/text-mining-unimi-2019-2020/lect05-clustering.html#8. Accessed on 26 February 2024.

Gutierrez, L.G., Voyant, C., Notton, G., Almorox, J. (2022). Evaluation and Comparison of Spatial Clustering for Solar Irradiance Time Series. Applied Sciences 12(17), 8529; https://doi.org/10.3390/app12178529

Howland, P., Wang, J., Park, H. (2006). Solving the small sample size problem in face recognition using generalized discriminant analysis. *Pattern Recognition* 39(2), 277-287; https://doi.org/10.1016/j.patcog.2005.06.013

IEA. (2023). *Solar PV still dominates renewable energy capacity additions.* https://www.iea.org/energy-system/renewables/solar-pv. Accessed on 7 October 2023.

IRENA. (2023). *Renewable capacity statistics.* https://www.irena.org/Publications/2023/Mar/Renewable-capacity-statistics-2023. Accessed on 10 October 2023.

Kumar, D.S., Maharjan, S., Albert, Srinivasan, D. (2022). Ramp-rate limiting strategies to alleviate the impact of PV power ramping on voltage fluctuations using energy storage systems. *Solar Energy* 234, 377-386; https://doi.org/10.1016/j.solener.2022.01.059

Laldjebaev, M., Isaev, R., Saukhimov, A. (2021). Renewable energy in Central Asia: An overview of potentials, deployment, outlook, and barriers. *Energy Reports* 7, 3125-3136; https://doi.org/10.1016/j.egyr.2021.05.014

Li, S., Liu, W., Hu, S., Xu, B. (2022). A method for determining the applicable geographical regions of PV modules field reliability assessment results based on regional clustering of environmental factors and their weights. *Sustainable Energy Technologies and Assessments* 53, 102620; https://doi.org/10.1016/j.seta.2022.102620

Liu, W., Jiang, X., Li, S., Luo, J., Wen, G. (2020). Photovoltaic module regional clustering in mainland China and application based on factors influencing field reliability. *Renew Sustain Energy Rev* 133, 110339; https://doi.org/10.1016/j.rser.2020.110339

Mahmoud, S.R., Gan, C.K. (2022). Classification of solar variability using k-means method for the evaluation of solar photovoltaic systems performance. *International Journal of Renewable Energy Research* 12, 692-702; https://doi.org/10.20508/ijrer.v12i2.12929.g8457

Munshi, A.A. & Mohamed, Y.A.R.I. (2016). Photovoltaic power pattern clustering based on conventional and swarm clustering methods. *Solar Energy* 124, 39-56; https://doi.org/10.1016/j.solener.2015.11.010

Pfenninger, S. & Staffell, I.L. (2016). Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. *Energy* 114, 1251-1265; https://doi.org/10.1016/j.energy.2016.08.060

Renewables.ninja. (2023). https://renewables.ninja/. Accessed on 1 September 2023.

Salguero, P.M., Sánchez, M.C.B., García, A.M., Lozano, J.M.S. (2022). Spatio-temporal dynamic clustering modeling for solar irradiance resource assessment. *Renewable Energy* 200, 344-359; https://doi.org/10.1016/j.renene.2022.09.113

Saracoglu, B.O., Ohunakin, O.S., Adelekan, D.S., Gill, J., Atiba, O.E., Okokpujie, I.P., Atayero, A.A. (2018). A framework for selecting the location of very large photovoltaic solar power plants on a global/supergrid. *Energy Reports* 4, 586-602; https://doi.org/10.1016/j.egyr.2018.09.002

Scikit-learn. (2023). Minmax. https://scikit-learn.org/0.15/modules/generated/sklearn.preprocessing.MinMaxScaler.html#sklearn.preprocessing.MinMaxScaler.fit_transform. Accessed on 10 August 2023.

Shahapure, K.R. & Nicholas, C. (2020). *Cluster quality analysis using silhouette score.* In Proc. IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, Australia. https://doi.org/10.1109/DSAA49011.2020.00096

Shubbak, M.H. (2019). Advances in solar photovoltaics: Technology review and patent trends. *Renewable and Sustainable Energy Reviews* 115, 109383; https://doi.org/10.1016/j.rser.2019.109383

Sinaga, K.P. & Yang, M.S. (2020). Unsupervised k-means clustering algorithm. *IEEE Access* 8, 80716-80727; https://doi.org/10.1109/ACCESS.2020.2988796

Sreenath, S., Azmi, A.M., Dahlan, N.Y., Sudhakar, K. (2022). A decade of solar PV deployment in ASEAN: Policy landscape and recommendations. *Energy Reports* 8(10), 460-469; https://doi.org/10.1016/j.egyr.2022.05.219

Tanoto, Y., Haghdadi, N., Bruce, A., MacGill, I. (2020). Clustering based assessment of cost, security and environmental tradeoffs with possible future electricity generation portfolios. *Applied Energy* 270, 115219; https://doi.org/10.1016/j.apenergy.2020.115219

Tanoto, Y., Haghdadi, N., Bruce, A., MacGill, I. (2021). Reliability-cost trade-offs for electricity industry planning with high variable renewable energy penetrations in emerging economies: A case study of Indonesia's Java-Bali grid. *Energy* 227, 120474; https://doi.org/10.1016/j.energy.2021.120474

Tanoto, Y., MacGill, I., Bruce, A., Haghdadi, N. (2017). *Photovoltaic Deployment Experience and Technical Potential in Indonesia's Java-Madura-Bali Electricity Grid.* https://www.ceem.unsw.edu.au/sites/default/files/documents/PhotovoltaicDeploymentExprience.pdf. Accessed on 1 September 2023.

Tlhalerwa, K. & Mulalu, M. (2019). Assessment of the concentrated solar power potential in Botswana. *Renewable Sustainable Energy Reviews* 109, 294-306; https://doi.org/10.1016/j.rser.2019.04.019

Tsafarakis, O., Sinapsis, K., Sark, W.G.J.H.M. (2018). PV system performance evaluation by clustering production data to normal and non-normal operation. *Energies* 11(4), 977; https://doi.org/10.3390/en11040977

Wang, Z., Koprinska, I., Rana, M. 92016). *Clustering based methods for solar power forecasting.* In Proc. 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada. https://doi.org/10.1109/IJCNN.2016.7727374

WEF. (2023). *Africa is leading the way in solar power potential.* https://www.weforum.org/agenda/2022/09/africa-solar-power-potential/. Accessed on 5 October 2023.

Wiharto, Wicaksana, A.K., Cahyani, D.E. (2021). Modification of a density-based spatial clustering algorithm for applications with noise for data reduction in intrusion detection systems. *International Journal of Fuzzy Logic and Intelligent Systems* 21(2), 189-203; https://doi.org/10.5391/IJFIS.2021.21.2.189

Wikipedia. (2023). *Indonesia provinces map (in Malgache).* https://en.wikipedia.org/wiki/Provinces_of_Indonesia#/media/File:Indonesia,_administrative_divisions_-_en_-_monochrome.svg. Accessed on 4 October 2023.

Yang, L., Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 415, 295-316; https://doi.org/10.1016/j.neucom.2020.07.061

Zagouras, A., Pedro, H.T.C., Coimbra, C.F.M. (2014). Clustering the solar resource for grid management in island mode. *Solar Energy* 110, 507-518; https://doi.org/10.1016/j.solener.2014.10.002