Classification of Solar Variability Using K-Means Method for The Evaluation of Solar Photovoltaic Systems Performance

Sara Ragab Mahmoud*‡, Chin Kim Gan*

*Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka

(sara.r.farghaly@gmail.com, ckgan@utem.edu.my)

‡Corresponding Author; Sara Ragab Mahmoud, sara.r.farghaly@gmail.com

Received: 29.03.2022 Accepted: 25.04.2022

Abstract- This paper presents a classification of solar tilt irradiance using the k-means clustering method, and an evaluation of the impact of different solar variabilities on monocrystalline and thin-film photovoltaic (PV) systems. The variability index and clearness index were implemented to quantify five years of solar datasets to assist in clustering solar variabilities. The elbow method was used to validate the k-clustering for solar variabilities. Due to the compact solar datasets, the Silhouette Coefficient and Gap Statistic were utilized to validate the k-cluster numbers. The PV performance was evaluated using the generated power, energy, and performance ratio for solar datasets from 2015 and 2019. Equal number of samples was taken from each PV system to analyse the average calculated values. The results showed that the elbow method was inaccurate for clustering solar variabilities, although it showed a weak elbow at K2 that was inaccurate for grouping solar variabilities. However, the k-means validation methods detected K3, K4, and K5 as the best k-cluster numbers. Among them, K4 was compatible for separating four types of solar variabilities, namely, overcast, moderate, mixed (clear/mild), and high variability. Based on the average performance values of the monocrystalline and thin-film PV systems for 2015 compared to 2019, similar degradation values were detected, especially for the performance ratio (0.77) under overcast. The thin-film showed degraded generated power and energy under the moderate type. The degraded generated power and performance ratio for the monocrystalline were due to the high passing clouds under the mixed and high variability types.

Keywords Solar Tilt Irradiance; Variability Index; Clearness Index; K-means clustering methods; Monocrystalline; Thin-Film.

1. Introduction

The United Nations (UN) has persistently encouraged consumers to use eco-friendly energy sources to generate clean electricity. The UN has also gradually established sustainable development agenda plans [1], as renewable energy systems are expected to be more dependable and affordable to every consumer in both urban and rural areas by 2030 [2]. Despite the existence of several natural renewable energy sources, solar energy, particularly solar irradiance (electromagnetic radiation emitted by the sun), is the most exploited and utilized source for producing electricity [3-5]. The Malaysian government has taken advantage of the abundantly available electromagnetic radiation that reaches the surface of their region. Since Malaysia is a tropical country, the sun provides plenteous amount of solar irradiance in the range of 4000–5000 kWh/m^2 [6].

Although this country geographically receives adequate irradiance to operate any solar PV system, passing clouds can cause significant fluctuations in the solar irradiance expected to be received. Passing clouds can also cause shading while passing above PV panels, thus, causing more solar variabilities and less generated power [7, 8]. As shown in Fig. 1, the dynamic shading impacts the P-V curve, and when more shading occurs, the system will generate lower power [9].

Small-scale PV panel systems in Malaysia are assembled in a tilted angle so they could better absorb the Solar tilt Irradiance (G_{tilt}). PV performance degradation could also be evaluated more accurately when the PV modules are assembled on the rooftop with a tilt angle [10,11].

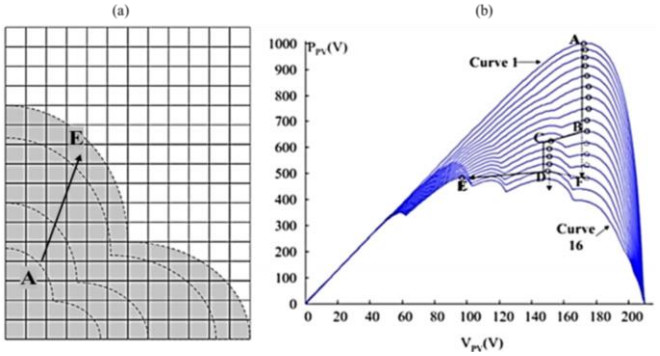


Fig. 1. Different partial shading mechanisms: (a) PV module; and (b) P-V curve on the same module [9].

Furthermore, solar irradiance is compartmentalized into different types due to the different types of passing clouds [12]. Jiang et al. [13] classified solar irradiance into three types, namely, overcast, mostly clear sky, and cloudy day in Singapore, which is a tropical country. Meanwhile, Elkholy et al. [14] classified solar irradiance in Egypt during the winter into two types, namely, sunny day and cloudy day. Both studies have classified PV system performance under different types depending on the passing clouds and the location.

Quantification methods were found to be more accurate for categorizing the stochastic components of solar irradiance [15]. Blaga et al. [16] clarified that no method could accurately quantify solar irradiance [17]. Previous studies [18–21] have used the Variability Index (VI) and the Clearness Index (CI) in order to categorize the quantified solar irradiance for PV performance evaluation. The VI is the ratio between the measured solar irradiance time interval for reaching the PV system and the ideal clear sky during the same time interval. Meanwhile, the CI is the ratio between the measured solar irradiance on a given surface and the ideal clear sky model, which is obtained from the extra-terrestrial level. Stein et al. [18] classified the Global Horizontal Irradiance (GHI) into four types for two different locations in the USA, as shown in Fig. 2 depending on the geographical location, the figure indicates different clear sky waveforms, along with different VI scales.

Trueblood et al. [19] implemented the VI and CI to classify the Plane of Array (POA), which is similar to G_{tilt} in the USA. Figure 3 shows that the categorization depends on the VI and CI for different solar variabilities. Their categorization of VI values was different, as the VI was lower than 2 for clear sky, while CI was higher than 0.5. As for the VI values under overcast conditions, they should be lower than 2, while CI values should be lower than 0.5, indicating low solar variability. Furthermore, the mild, moderate, and high variability levels were determined depending on the VI range values, thus, neglecting the CI values.

Moreover, Lai and McCulloch [20, 21] implemented different clustering methods, including the k-means method, on CI values to detect the solar behaviour of the clear sky. Li et al. [22] used the k-means method to analyse the clustering behaviour over the solar radiation (kWh/m^2), along with clustering the historical CI data that were collected from four different regions in the USA. Bhola and Bhardwaj [23] used the k-means method on solar irradiance data from India to extract the monthly average performance ratio in order to evaluate PV performance under Indian weather conditions.

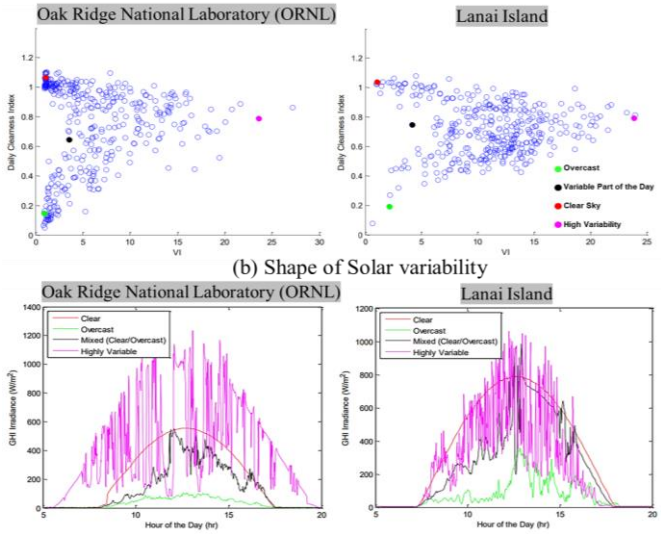


Fig. 2. Quantification of GHI at the Oak Ridge National Laboratory (ORNL) and Lanai Island in the USA: (a)Scatter plot for CI versus VI; and (b)Solar variability behaviour [18].

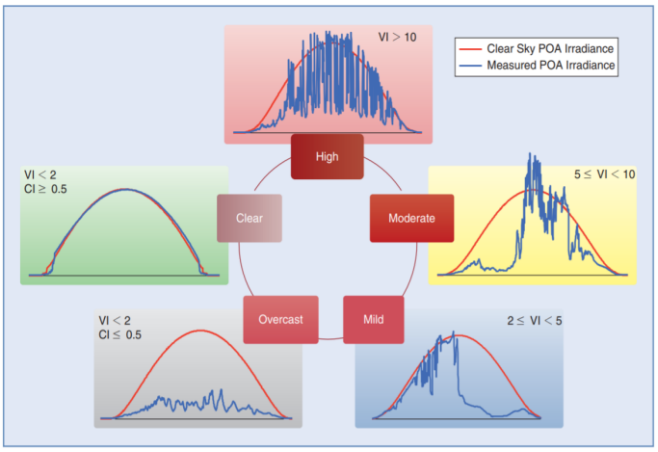


Fig. 3. POA classification by using VI and CI values [19].

The degradation performance of different types of PV panels was evaluated using the standard code, IEC 61724 [24, 25]. It was used to assess the performance reliability of PV systems by monitoring and investigating PV parameters. The performance ratio (PR) method is mainly used to evaluate the degradation performance of energy production using daily, monthly, or annual data, as defined in IEC 61724 [26]. Researchers worldwide use the PR method in standard assessment techniques, as it is accessible via the International Energy Agency under Task number 13 [27]. Previous research studies [28–32] have evaluated and compared different types of PV module, including monocrystalline (c-Si) and thin-film (TF), in tropical weather conditions, including in Malaysia. However, performance degradation is evaluated either annually or monthly, but not under each type of solar variability classifications for each type of PV modules. Previous studies conducted in different states in Malaysia [33–35] have further explained the varying performances of c-Si and TF panels, as follows:

- Under the plentiful amount of solar irradiance with fewer clouds, the c-Si panels performed well compared to TF

panels in producing energy. Both systems generated significantly high power

- With more passing clouds resulting in lesser solar irradiance, TF panels performed better in generating power compared to c-Si panels.
- Under the solar variability of Malaysian weather conditions, c-Si panels suffered from high temperatures, unlike TF panels that were the best in maintaining the lowest panel temperatures.

The c-Si panels have been used for many years, and are considered the oldest and most efficient panels. This type of system belongs to the initial generation of PV modules made from silicon with a crystalline structure. TF panels are thinner than c-Si panels by 10 μm and are mostly recommended to be used during summer, as their production process uses little or no silicon [36]. Thus, by comparing and analysing c-Si and TF panel systems under different solar variability types using IEC 61724, more details could be obtained regarding their degradation performance.

It is essential to demonstrate the importance of evaluating PV system performance under different solar variabilities according to tilted solar irradiance instead of the global horizontal irradiance. The latter is considered as an uncertainty assessment for measuring and evaluating the convergence of solar tilt irradiance and PV performance of tilting panels [37]. Furthermore, most of the small-scale PV systems in Malaysia are mounted on tilted rooftops, and roof slopes are widespread in residential areas.

In this study, the quantification method was used on tilted solar irradiance and classified by using the k-means method. The validation method for the k-means method was used to pick out suitable k-cluster numbers. Next, the degradation performance of c-Si and TF PV systems was evaluated under different solar variabilities by using Active Power, Production Energy, and Performance Ratio. The datasets for 2015 and 2019 were focused on the PV system evaluation. The grid-connected PV systems of c-Si and TF panels are located on the rooftop of the Faculty of Electrical Engineering (FKE) building inside the Universiti Teknikal Malaysia Melaka (UTeM) campus. These systems were used to evaluate performance degradation under solar variability conditions in Melaka.

2. Methodology

The geographical coordinates of the c-Si and TF PV systems are 2°18'52.3"N 102°19'11.3"E, and Fig. 4. shows the single line diagram (SLG) for each system [38]. The total numbers of c-Si and TF modules in the system were 24 and 48 modules, respectively. Each of the panel systems were connected in string connections with three single-phase inverters, separately.

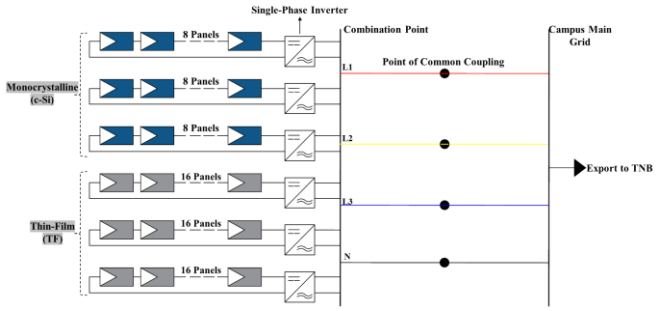


Fig. 4. Single line diagram connection for on-grid c-Si and TF PV systems.

Regarding the collection of measured data, Fig. 5 shows two different points that the datasets were collected from. The data for 2019 were collected from Siemens Sentron PAC4200 monitoring devices, while the data for 2015 were collected from the SMA Sunny Boy (SB 2000HF-30) solar inverter. These datasets were filtered and the hours between 7:18 a.m. until 7:00 p.m. were used as the duration between sunrise and sunset. Additionally, two pyranometer devices were used to collect the measured data of tilted solar irradiance. Both devices are located next to the c-Si and TF systems, as shown in Fig. 6 specification parameters for the c-Si and TF panels are shown in Table 1.

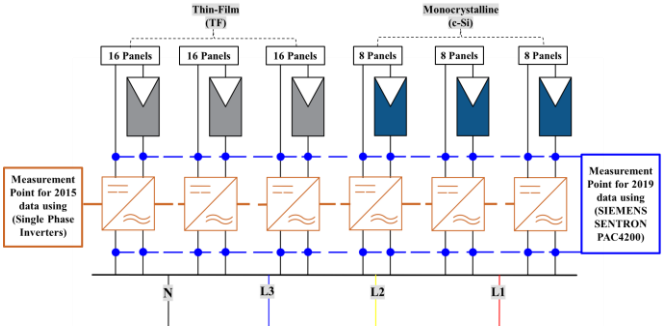


Fig. 5. Measurement points for the PV system.

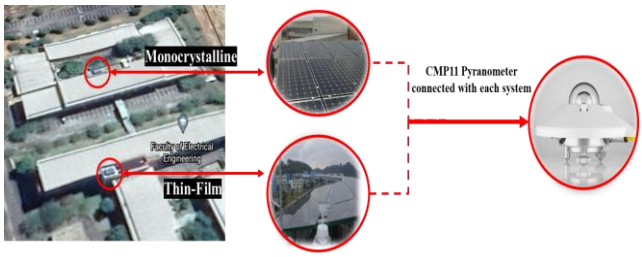


Fig. 6. C-Si and TF modules with pyranometer device at FKE's rooftop.

To obtain accurate classification for the quantified G_{tilt}, a 5-year dataset was used to avoid massive missing data, because the PV monitoring systems often face data loss, or prolonged periods that generated major uncertainty data [5]. The following Eq. (1) and Eq. (2) can be used to quantify the solar tilt irradiance by using the VI and CI values, respectively [39].

$$VI = \sqrt{\frac{\sum\left(G_{tilt}(k) - G_{tilt}(k-1)\right)^2}{\sum\left(G_{ideal}(k) - G_{ideal}(k-1)\right)^2}} \quad (1)$$

$$CI = \frac{G_{tilt}(k)}{G_{iso}(k)} \quad (2)$$

Table 1. Specification parameters for the c-Si and TF panels

Specification Parameters	c-Si	TF
Electrical Datasheet under STC		
Maximum Power (P_{max})	255 Wp	130 Wp
Open Circuit Voltage (V_{oc})	37.8 V	60.4 V
Short Circuit Current (I_{sc})	8.66 A	3.41 A
Maximum Power Point Voltage (V_{pmax})	31.4 V	46.1 V
Maximum Power Point Current (I_{pmax})	8.15 A	2.82 A
Module Efficiency (Ƞ)	15.51%	9.3%
Power Tolerance	+ / -2%	+7% / -2%
PV Panels at FKE's rooftop		
Area per Module (m^2)	1.68	1.40
Number of Modules	24 modules	48 modules
Array Maximum Power ($P_{max(rated)}$)	6.12 kWp	6.24 kWp
Total Panels Area (m^2)	40.32	67.2
Fixed Tilt Angle	10°	10°

where G_{tilt} is the actual tilted solar irradiance, G_{ideal} is the reference of solar irradiance, G_{iso} is the reference of solar isolation, and k is time. Then, a computational programming language called Python was used for the k-means method to partition the combination of VI and CI scattering datasets. Python is a programming language that adheres to the Object-Oriented Programming paradigm; it is frequently used to create websites, applications, and conduct data analysis. in this paper it was used to cluster the solar variability.

Figure 7 shows a flowchart of the steps taken to classify the VI and CI values by using python code. The k-means++ function was used as the initializing scheme, because the VI and CI scattering uniform waveforms could reduce the k-means efficiency. By using the k-means function, the k-cluster range was chosen between 1 and 12, following the number of k-clusters in a previous study [23], with extension to the k-cluster range by one extra number. The iteration number was equal to the total counted input data from the five years of VI and CI data.

The Elkan k-means algorithm was used as it is efficient on data with well-defined clusters by using triangle inequality [40]. It is one of k-means functions that accelerate the clustering process without sacrificing computational time against solution quality [41]. Also, it eliminates excessive distance by putting a set of lower and upper limitations on distances between points and selected centroids. After that, clustering was performed on the data using validation methods for the best number of k-clusters which are as follows:

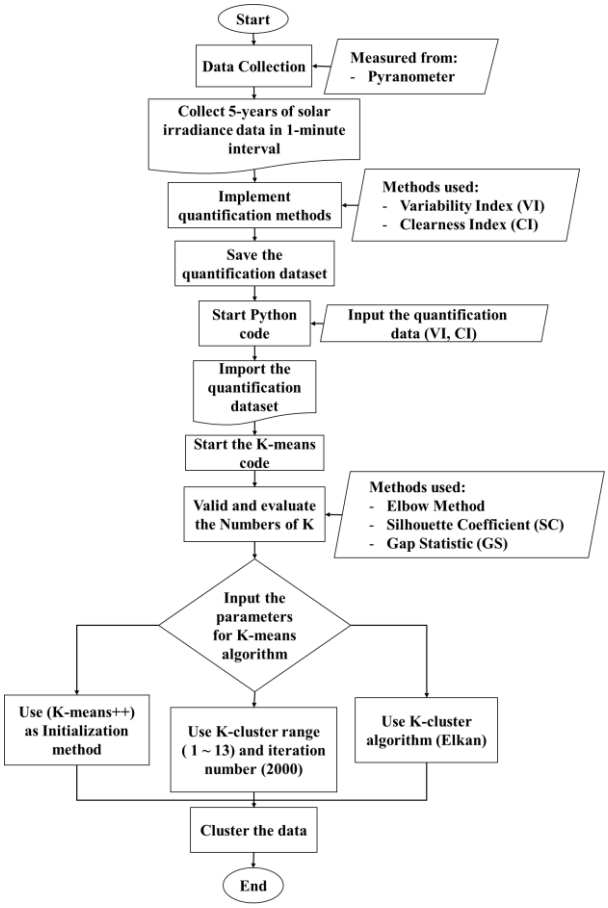


Fig. 7. Flowchart of computation steps for k-means algorithm and estimation methods for k-clusters.

- The elbow method was chosen for k-means because it can utilize the range of k-cluster numbers and detects the optimal number using the distance in k-means clustering. In some rare cases, the elbow points are not clearly shown due to the sensitivity of the k-means method [42]. However, when the elbow method did not function well, the Silhouette Coefficient (SC) and Gap Statistic (GS) were used to validate the best k-cluster numbers.
- The SC compares the similarities in each group of scattering data, with the resolution of other clusters' separation [43]. The range of SC is between 1 to -1; thus, if the SC values reach the negative side, this means that the cluster method is not qualified due to the over-compactness in the data. Furthermore, if the SC value is close to 0, the data holds multiple grouping numbers of the k-clusters. Equation (3) is the calculation method of SC where a is the average distance between each sample at the same k-cluster, and b is the samples average distance to the sample in the nearest cluster.

$$SC = \frac{b-a}{max\{a,b\}} = \begin{cases} 1 & a < b \\ 0 & a = b \\ -1 & a > b \end{cases} \quad (3)$$

- The GS was designed by Tibshirani et al. [44], as a modern method for the implementation with any clustering

method [45]. It compares the distraction during data partition with an assumption of null distribution reference. If the null reference is valid and the dispersion within a cluster is reduced, then, the shown k-cluster value is the best cluster [46]. Additionally, the highest GS value from the smallest k-cluster number will be utilized. Equation (4) represents GS where K is the number of k-cluster, and E is the standard Error which is the standard deviation of its sampling distribution, and $WCSS$ is within cluster sum of square.

$$GS(K) = (E_n^*\{log\, WCSS_K\}) - log\, WCSS_K \quad (4)$$

The previous steps for the k-means clustering algorithm were conducted to find the best numbers of k-clusters to classify the tilted solar irradiance. Next, the three main types of solar variabilities mentioned by Baharin et al. [47] were quantified from the same PV systems at the FKE. They were used as a reference with the k-clusters to consider the good ones in grouping the different solar variabilities with the compatible ones. The three solar variability types are clear sky, high variability, and overcast, as shown in Fig. 8 with regards to the clear sky condition, which is quite rare in Melaka, Malaysia, light fluctuations with less than 50% variability in the solar irradiance of solar tilt was used as a mixed, clear sky condition.

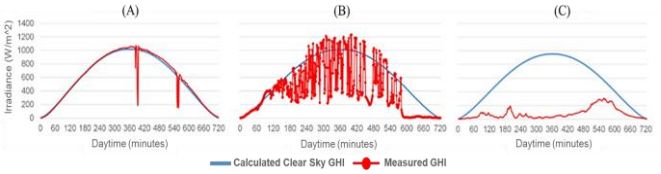


Fig. 8. Graph samples of different solar variabilities: (a) Clear Sky; (b) High Variability; and (c) Overcast [47]

Lastly, using the IEC 61724, the degradation performance was evaluated using random samples taken from the 2015 and 2019 datasets. These random samples were chosen daily depending on the lower number of days for different solar variability types. As shown in Fig.9., active power and energy values were extracted from the monitoring systems. The evaluation did not depend on the same specific day from both years, but on the average values, along with the waveforms generated from the random samples under each type of solar tilt irradiance. Equation 5 was used to calculate the PR for the c-Si and TF systems [48]. Where E_{ac} is the Active Energy $P_{(max(rated))}$ is the array maximum power for C-Si and TF, H is the Total in-plane solar irradiance, G_{STC} is the reference of solar irradiance from STC.

$$PR = \frac{E_{ac}}{P_{max\,(rated)}} \Big/ \frac{H}{G_{STC}} \quad (4)$$

3. Results

Fig. 10 shows an exponentially decreasing curve for the elbow method that is used to find the optimal number for the k-means algorithm. The curve shows a smooth flow with significant dim and weak elbow points at K2, while the compactness of the VI and CI scattering data are apparent in the weak elbow point.

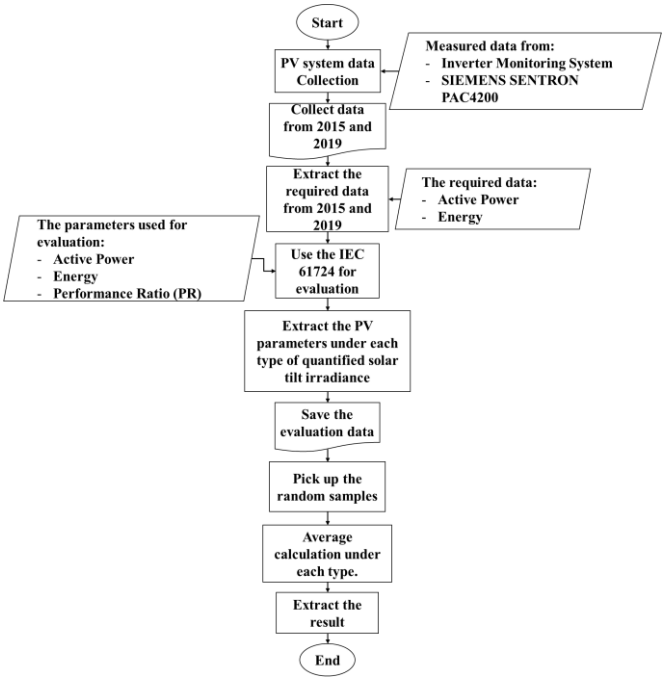


Fig. 9. Flowchart of the build-up relationship between the PV data and the solar irradiance data.

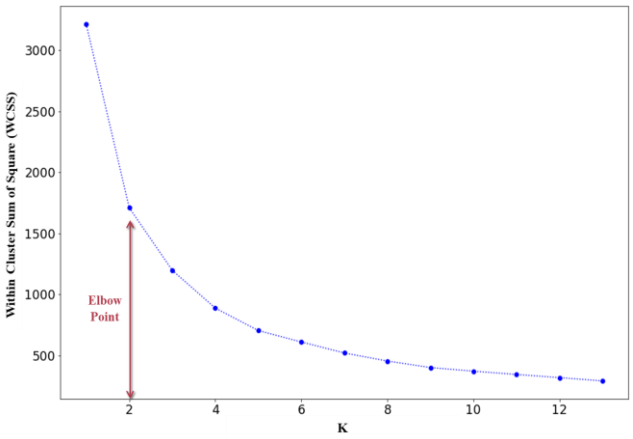


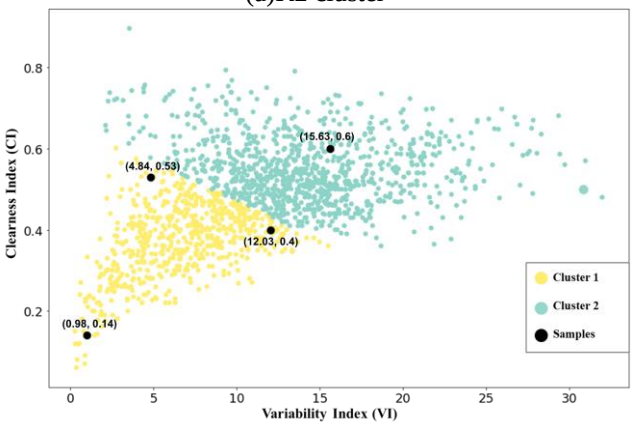
Fig. 10. The Elbow method for the VI and CI data set.

It is improper to use K2 as the optimal k-cluster for solar classification, because different solar variability types have been incorrectly grouped, as shown in Fig.11. The figure also shows five randomly picked points, as visualized by the solar tilt irradiance. These points showed that three solar variability plots at Cluster 1 behaved differently, as they gave low, mild, and semi-clear intensity solar conditions. Additionally, if K2 had been implemented during the evaluation process for the PV systems, it would have been inefficient for evaluating the impact of different solar variability types on the PV system performance degradation.

The SC in Fig. 12(a) shows that the k-cluster values ranged between 0.4 and 0.3, which clarifies that the scattering VI and CI data have more than one k-cluster number as the best one for partitioning. The K4 was higher than K3 in SC, while K5 was slightly equal to K3. However, in the GS plot (Fig. 12(b)), K3 is higher than K4 and K5. K6 was lower than K5 under the SC and GS, while K7 upwards started to cause sudden changes compared to K3, K4, and K5. The SC has also detected that there were more than one

k-clusters and that K3 was equal to K5. Meanwhile, the GS showed the best k-cluster in the smallest k-cluster values. K3, K4, and K5 are chosen as the best numbers of k-clusters, and Fig. 13 shows the plot and values under each method.

(a)K2 cluster



(b)Samples

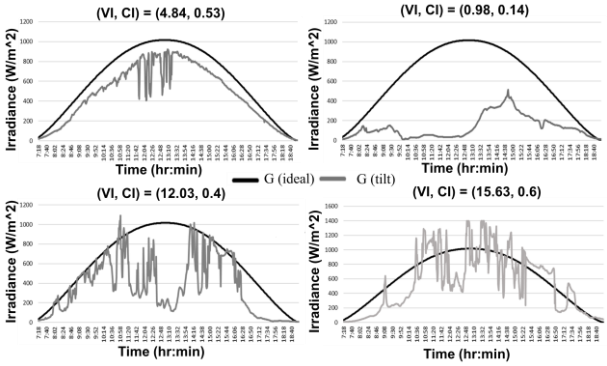
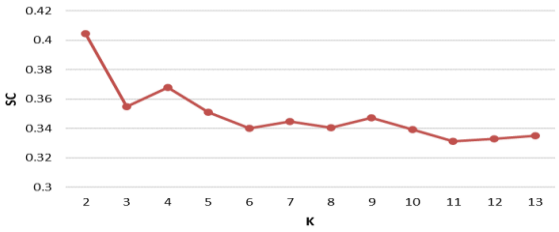


Fig. 11. The difference in solar tilt irradiance waveform under (a) K2 cluster by using (b) Samples.

(a)Silhouette Coefficient (SC)



(b)Gap Statistic (GS)

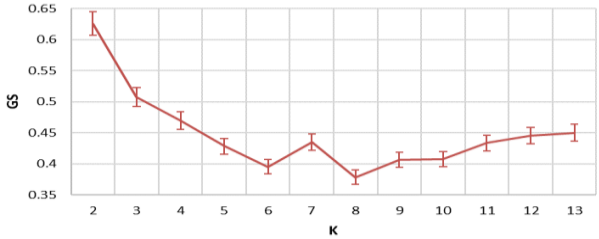


Fig. 12. Evaluating the numbers of k-clusters: (a) Silhouette Coefficient (SC); and (b) Gap Statistic (GS)

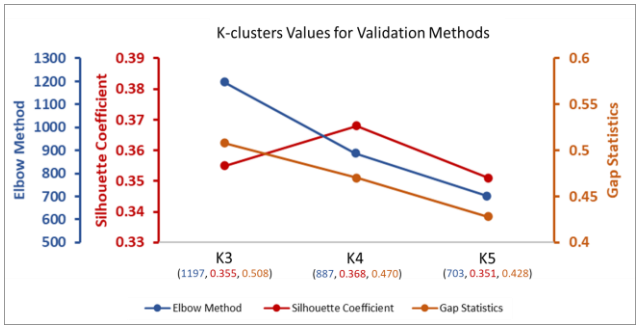
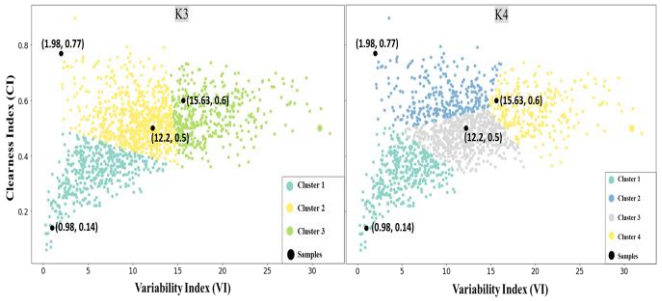


Fig. 13. Variance in the k-cluster performance for K3, K4, and K5.

Figure 14 and Figure 15 show a comparison between K3 and K4, and K4 and K5, respectively. In order to evaluate the performance of the PV systems, one of the most compatible k-clusters was chosen for grouping each type of solar variability.

Figure 14 shows randomly picked points for the solar variability performance between K3 and K4. In the K3, Cluster 2 region, two solar variability waveforms were acting differently, which showed that moderate solar fluctuation can occur when the CI has lower values while the VI has higher values. Meanwhile, Fig. 15 shows a comparison between K4 and K5, with different randomly picked points. This figure is focused on picking solar variabilities close to the decision boundary lines and merging the clear sky type with the mixed sky type. Significantly, this result showed that some mixed sky data were dropped at the new Cluster 2 region in K5, while they were kept in the same Cluster 2 region in K4. The same observation was true for the behaviour of the low solar irradiance in the Cluster 1 and Cluster 2 regions in K4 and K5, respectively.

(a)K3 versus K4



(b)Samples

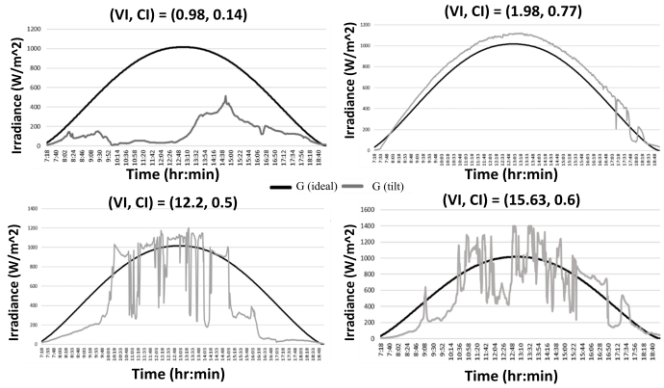


Fig. 14. Comparison between scattering plot for (a) K3 versus K4 by using (b) Samples.

(a)K5 versus K4

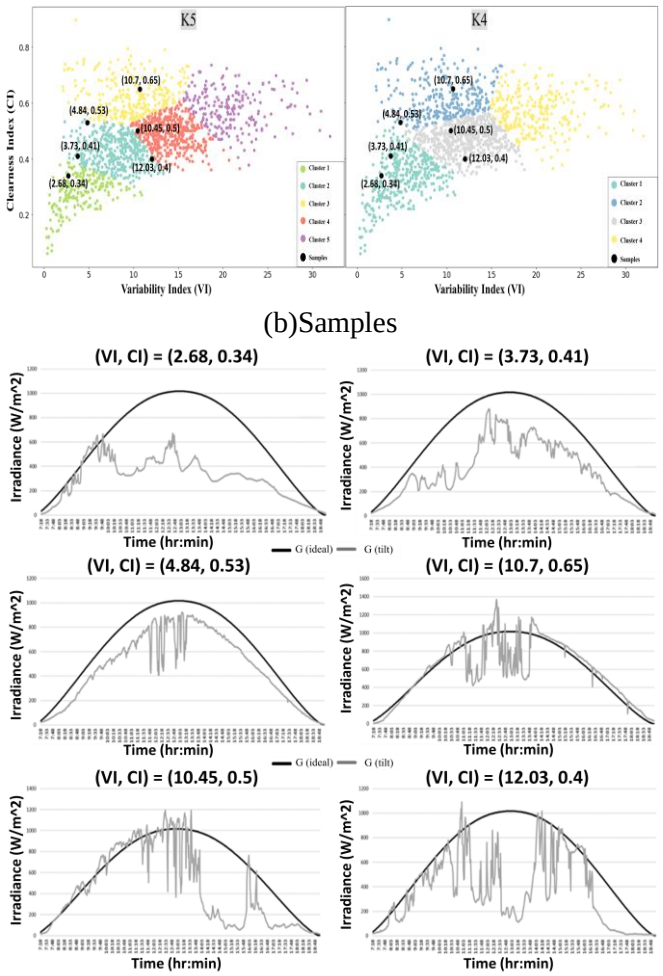


Fig. 15. Comparison between scattering plot for (a) K5 versus K4 by using (b) Samples.

Clearly, the solar variability results showed that K4 was compatible for grouping the same solar irradiance variability. Figure 15 and Table 2 show more details of the values for each cluster under K4 for the five-year VI and CI data. The centroid of the full scatter data was similar to the Cluster 2 centroid, as the VI and CI values for both were 12.04 and 0.47, and 11.33 and 0.48, respectively. Furthermore, the range of VI (5.40–18.68) and CI (0.34–0.56) values for Cluster 2 retained the highest number of days (626) compared to the other clusters. These results showed that most of the tilted irradiance variabilities at the FKE building have mostly behaved as the moderate solar irradiance type for the last five years.

Cluster 1 was located in the VI range of 0.27 and 10.92, and the CI range of 0.06 and 0.50, which indicated that the overcast type was located in the lowest region of the scattering plot. The ranges of the VI (2.11–16.15) and CI (0.51–0.90) values for Cluster 3 took on an outstanding partition place to merge between the rare appearance cases, the clear sky, and mild types, under the mixed (clear/mild) type. Cluster 4 can be linked to the highest VI range (14.90–31.98) and in the middle to nearly the high value range of the CI (0.36–0.74), showing a massive solar fluctuation in this zone. Based on Fig.8, it can be identified as the high variability type.

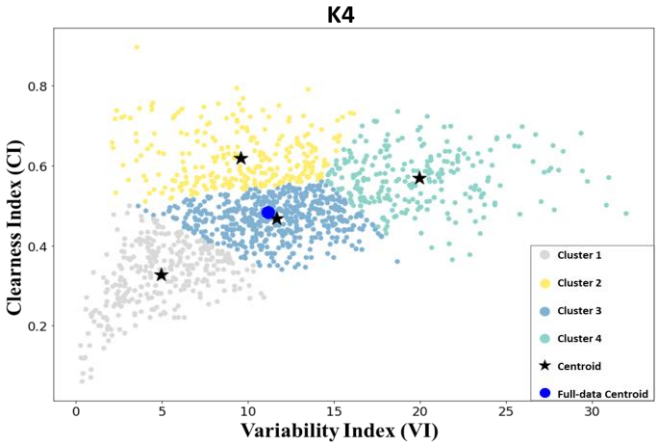


Fig. 16. K4 cluster for VI versus CI

Table 2. The K4 centroids and ranges for the clustering for both VI and CI

Number of Days		VI		CI	
		Centroid	Range	Centroid	Range
Full Data Centroid	1608	11.33		0.48	
Cluster 1	383	5.19	0.27 ~ 10.92	0.33	0.06 ~ 0.50
Cluster 2	626	12.04	5.40 ~ 18.68	0.47	0.34 ~ 0.56
Cluster 3	318	9.71	2.11 ~ 16.15	0.61	0.51 ~ 0.90
Cluster 4	281	20.03	14.90 ~ 31.98	0.57	0.36 ~ 0.74

Clusters 1–4 have been identified as overcast, moderate, mixed (clear/mild), and high variability, as shown in Fig. 17 the figure for overcast shows a low G_{tilt} performance with some fluctuations. The performance of this type of solar variability was below that of the G_{ideal} waveform. The moderate type made the G_{tilt} start to interact with G_{ideal} at several minutes of intervals, with medium fluctuation occurring simultaneously. In the mixed (clear/mild) type, the G_{tilt} followed the ideal G_{ideal} mixed with above medium fluctuation that occurred after midday. For the last type, which was high variability, the G_{tilt} overlapped and surpassed the G_{ideal} waveform with lots of fluctuation that happened multiple times per hour.

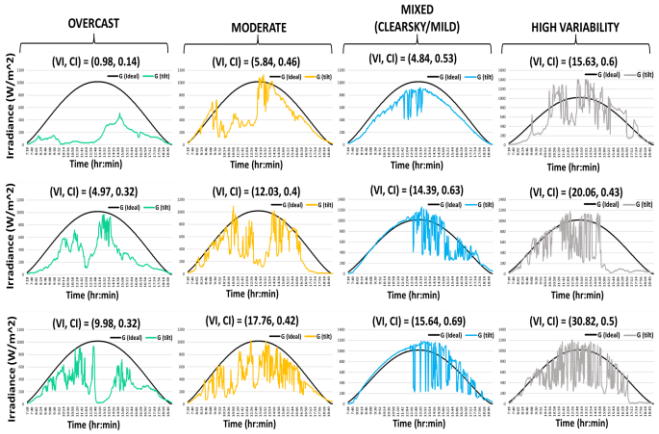


Fig. 17. The four types of Solar tilt irradiance variabilities

Figure 18 shows the percentage of solar variability types in 2015 and 2019. Both years showed that the majority of the population was in the moderate category since their

percentages of 35% and 49% represented the total days of 127 and 125, respectively. This result indicated that more than a quarter of the year had a lot of passing clouds that caused rain. The high variability type had the same percentage of 18% for both years, but occurred on different total numbers of days, which were 64 days in 2015 and 45 days in 2019. The mixed type had a higher percentage of occurrence in 2015 than in 2019, at 24% with 85 days and 12% with 31 days. Meanwhile, the overcast type occurred 23% of the year with 85 days and 21% with 54 days. The different percentages between types indicated that most of the datasets were missing.

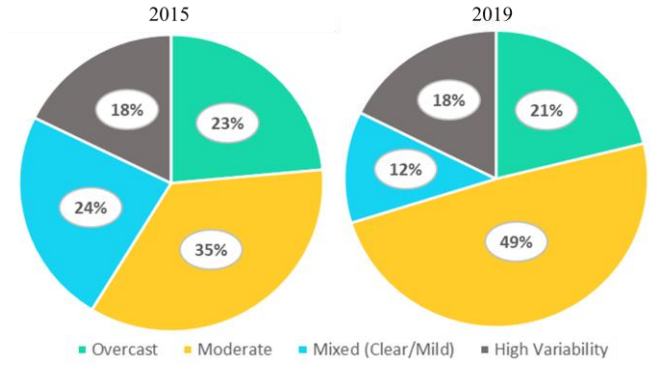


Fig. 18. Percentage of tilted solar irradiance in 2015 and 2019.

The dissection of 30 random samples was made to comprehend the analyses of c-Si and TF systems performed for the 2015 and 2019 datasets. Figure 18 clarifies that the minimum number of days is located in the mixed type by 12%. Then, a simple average calculation is applied, as shown in Table 3. The power, energy, and performance ratios of the c-Si and TF systems have been denoted as P_{ac1}, E_1, and PR_1, and P_{ac2}, E_2, and PR_2, respectively. Based on the average calculation for these samples, the quantifying values for VI and CI gave values that are close to the centroid values, as shown in Table 2.

When the c-Si and TF degradation values for the 2015 and 2019 datasets were compared, the same degradation impact was found under overcast conditions. Even with the slightly high values generated from the TF system in 2015, both panels have given the same average PR values in 2019. The sequence drops were clear, which occurred due to the heavy clouds. Figure 19 shows a clear vision of the drop.

The degradation of the TF system under the moderate type was more obvious based on the E_2 and PR_2 values (see Fig. 20). P_{ac2} was noticeably reduced in the c-Si system compared to in the TF system. Thus, degradation performance still occurred under this type of solar condition, but it varied between the still holding power generation from the TF system and the high injected energy from the c-Si system.

The mixed type resulted in the same degradation performance as the moderate type for both panel systems. However, the TF system had shown high values in 2015 and 2019, while the c-Si system had shown low values between both years, but highly generated values for the mixed type compared to the moderate type. Furthermore, the generated power and energy for both systems were considered to be the best when compared between the overcast and the moderate types, yet the TF system was performing better than the c-Si system under the mixed type (see Fig. 21).

The active power and energy values were highly impacted for the c-Si system under the high variability condition, which led to a reduction in the PR2 values. Additionally, Fig. 22 shows a high fluctuation drop for the c-Si system compared to the TF system. This result clarified the inability for a good electrical generation by the c-Si modules during highly intensive passing of clouds.

Table 3. Average values for c-Si and TF performances under different solar variability types

	Overcast	Moderate	Mixed	HighVariability
2015				
VI	4.70	12.34	10.42	19.96
CI	0.35	0.47	0.61	0.54
P_{AC1}	1.46	1.94	2.40	2.13
P_{AC2}	1.55	2.19	2.79	2.39
E_1	17.79	23.70	29.14	25.88
E_2	18.57	26.16	33.40	28.69
PR_1	0.81	0.80	0.75	0.76
PR_2	0.83	0.86	0.83	0.83
2019				
VI	5.58	12.81	10.69	19.61
CI	0.32	0.46	0.60	0.55
P_{AC1}	1.37	1.88	2.18	1.94
P_{AC2}	1.38	1.95	2.51	2.31
E_1	16.04	22.07	26.52	22.48
E_2	16.28	23.13	30.60	26.95
PR_1	0.77	0.74	0.69	0.65
PR_2	0.77	0.76	0.78	0.76

4. Conclusion

This paper presents the classification of quantified solar tilt irradiance using k-means method to evaluate the performance of PV systems. The k-means method was implemented using three validation methods, which were the Elbow Method, Silhouette Coefficient, and Gap Statistic. The degradation performance was also evaluated by picking random samples from different solar variability types. The results showed that the elbow point was convincing in showing that K2 was the best number of k-clusters. However, it was considered too weak to cluster the data into two types, thus, it was rejected as an optimal number. The k-clusters for SC indicated that there were three best k-cluster numbers, as their values were lower than 0.5, including the GS values that assisted in choosing the smallest k-clusters. It validated that the best k-clusters were K3, K4, and K5. In a comparison between the three k-clusters, the k-means decision boundary performed well with K4, as it managed to partition and grouped the four types of tilt solar irradiances,

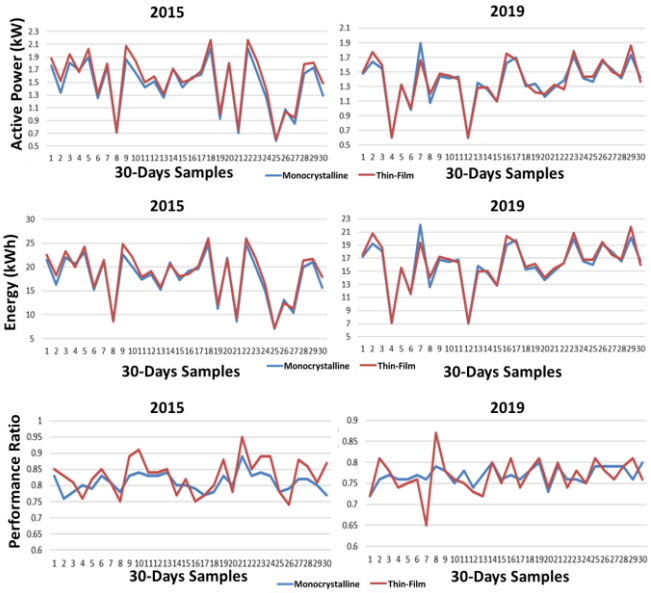


Fig. 19. The random samples of c-Si and TF systems' performance under overcast type

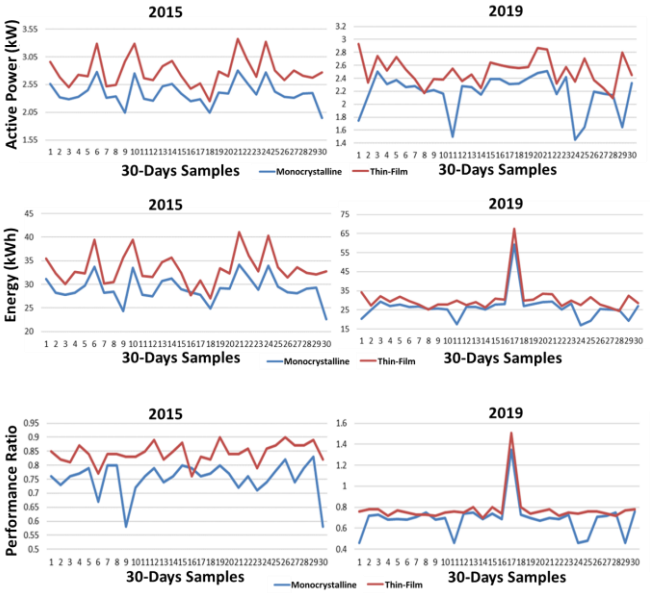


Fig. 21. The random samples of c-Si and TF systems' performance under mixed type

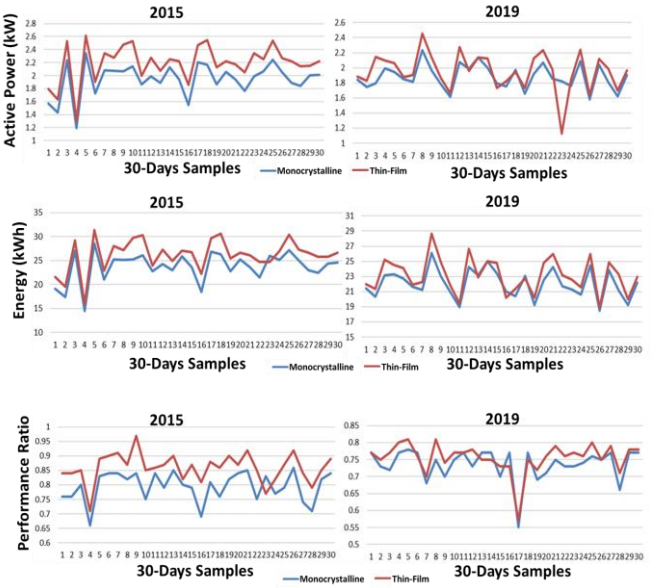


Fig. 20. The random samples of c-Si and TF systems' performance under moderate type

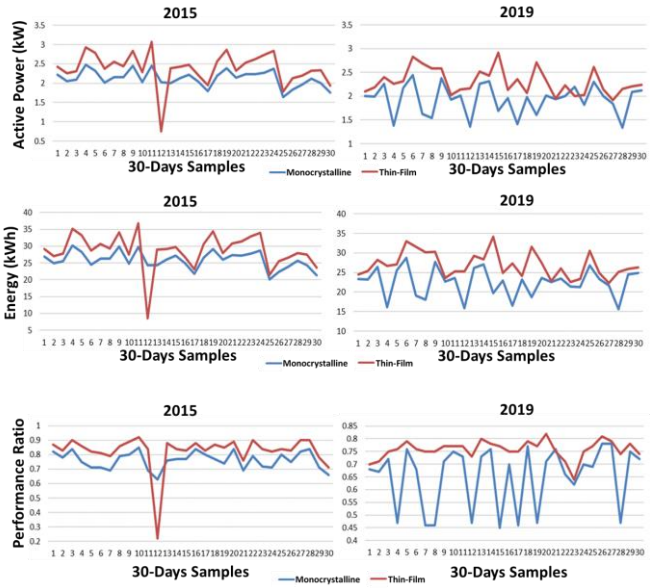


Fig. 22. The random samples of c-Si and TF systems' performance under high variability

which were overcast, mixed (clear/mild), moderate, and high variability. Then, the average values calculated from 30 random samples were overviewed with the degradation performance of PV panel systems under each type, and the following points were established:

- In 2015, the TF system generated higher active power than the c-Si system. However, in 2019, the generated active power by the TF system was reduced, especially under overcast and moderate solar variables. Meanwhile, the generated active power by the c-Si system in 2015 under the same conditions was low, but the degradation was not as high as in the TF system.
- Under mixed conditions, the generated power and energy from both systems were higher than under other solar variability types. Furthermore, passing clouds can highly impact the performance of the c-Si system, and the degradation performance detected from the low generation of power and energy compared to the TF system.
- The degradation performance was more obvious in the c-Si system than in the TF system under the high variability type, which indicated that more passing clouds could negatively impact the c-Si system's generation performance.

Overall, the TF showed higher degradation performance than c-Si under each type except the high variability type, as the degradation was obvious in 2019 for the c-Si system.

This study has confirmed the relationship between G_{tilt} variability and high impacts on PV performance. The importance of implementing meteorological methods to assess and classify G_{tilt} variability was also studied to find compatible solar variability types for the Melaka region.

References

[1] J. UN (2022) [Online]. "THE 17 GOALS | Sustainable Development", Department of Economic and Social Affairs. Available at: https://sdgs.un.org/goals (Accessed: 06 December 2020).

[2] R. William, "Transforming Our World: The 2030 Agenda for Sustainable Development", A New Era in Global Health, ISBN-13: 9780826190123, pp. 618, 2017.

[3] G. Rottman, "Measurement of total and spectral solar irradiance", Space Science Review, vol. 125, no. 1–4, pp. 39–51, 2006.

[4] M. Kumar, "Social, Economic, and Environmental Impacts of Renewable Energy Resources", Wind Solar Hybrid Renewable Energy System. London, United Kingdom: IntechOpen, 2020 [Online]. Available at: https://www.intechopen.com/chapters/70874.(Accessed: 06 December 2020).

[5] K. Okedu, A. Salem AlSenaidi, I. Al Hajri, I. Al Rashdi, and W. Al Salman, "Real Time Dynamic Analysis of Solar PV Integration for Energy Optimization", Intarnationl Journal of Smart grid, vol.4, no.2, pp. 68-79, 2020.

[6] S. T. Mohammad, H. H. Al-Kayiem, M. A. Aurybi, and A. K. Khlief, "Measurement of global and direct normal solar energy radiation in Seri Iskandar and comparison with other cities of Malaysia", Case Studies in Thermal Engineering, Vol. 18, pp. 100591, 2020.

[7] M. A. Al Mamun, M. Hasanuzzaman, and J. Selvaraj, "Experimental investigation of the effect of partial shading on photovoltaic performance", IET Renewable Power Generation, vol. 11, no. 7, pp. 912–921, 2017.

[8] K. Tsuboi, T. Matsuoka and T. Yachi, "An output degradation of photovoltaic module by fine particles deposition", 2012 International Conference on Renewable Energy Research and Applications (ICRERA), pp. 1-5, 2012.

[9] K. Ishaque and Z. Salam, "A deterministic particle swarm optimization maximum power point tracker for photovoltaic system under partial shading condition", IEEE Transactions on Industrial Electronic, vol. 60, no. 8, pp. 3195–3206, 2013.

[10] C. A. Gueymard, "From global horizontal to global tilted irradiance: How accurate are solar energy engineering predictions in practice?", American Solar Energy Society - SOLAR 2008, Including Proc. of 37th ASES Annual Conf., 33rd National Passive Solar Conf., 3rd Renewable Energy Policy and Marketing Conf.: Catch the Clean Energy Wave, vol. 3, pp. 1434–1441, 2008.

[11] J. Polo, S. Garcia-Bouhaben, and M. C. Alonso-García, "A comparative study of the impact of horizontal-to-tilted solar irradiance conversion in modelling small PV array performance", Journal of Renewable and Sustainable Energy, vol. 8, no. 5, pp. 053501, 2016.

[12] Weatherwise, "Clouds: A Primer on the Most Common Types and What They Forecast", Weatherwise, vol. 73, no. 1, pp. 34–39, 2020.

[13] L. L. Jiang, D. L. Maskell, R. Srivatsan, and Q. Xu, "Power variability of small scale PV systems caused by shading from passing clouds in tropical region", Conference Record of the IEEE Photovoltaic Specialists Conference, pp. 3159–3164, 2016.

[14] A. Elkholy, F. H. Fahmy, A. A. Abou El-Ela, A. E.-S. A. Nafeh, and S. R. Spea, "Experimental evaluation of 8kW grid-connected photovoltaic system in Egypt", Journal of Electrical Systems and Information Technology, vol. 3, no. 2, pp. 217–229, 2016.

[15] R. Blaga and M. Paulescu, "Quantifiers for the solar irradiance variability: A new perspective", Solar Energy, vol. 174, pp. 606–616, 2018.

[16] R. Blaga, A. Sabadus, N. Stefu, C. Dughir, M. Paulescu, and V. Badescu, "A current perspective on the accuracy of incoming solar energy forecasting", Progress in Energy and Combustion Science, vol. 70, pp. 119–144, 2019.

[17] M. U. Yousuf and S. M. R. Hussain, "Performance evaluation of independent global solar radiation estimation models for different climatic zones: A case study", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2021, DOI: 10.1080/15567036.2021.1958955.

[18] J. S. Stein, C. W. Hansen, and M. J. Reno, "The variability index: A new and novel metric for quantifying irradiance and pv output variability", World Renewable Energy Forum, WREF 2012, Including World Renewable Energy Congress XII and Colorado Renewable Energy Society (CRES) Annual Conferen, vol. 4, pp. 2764–2770, 2012.

[19] C. Trueblood, S. Coley, T. Key, L. Rogers, A. Ellis, C. Hansen, and E. Philpot, "PV Measures up for fleet duty: Data from a tennessee plant are used to illustrate metrics that characterize plant performance", IEEE Power and Energy Magazine, vol. 11, no. 2, pp. 33–44, 2013.

[20] C. S. Lai, X. Li, L. L. Lai, and M. D. McCulloch, "Daily clearness index profiles and weather conditions studies for photovoltaic systems", Energy Procedia, vol. 142, pp. 77-82, 2017.

[21] C. S. Lai, Y. Jia, M. D. McCulloch, and Z. Xu, "Daily Clearness Index Profiles Cluster Analysis for Photovoltaic System", IEEE Transactions on Industrial Informatics, vol. 13, no. 5, pp. 2322–2332, 2017.

[22] S. Li, H. Ma, and W. Li, "Typical solar radiation year construction using k-means clustering and discrete-time Markov chain", Applied Energy, vol. 205, pp. 720–731, 2017.

[23] P. Bhola and S. Bhardwaj, "Clustering-based computation of degradation rate for photovoltaic systems", Journal of Renewable and Sustainable Energy, vol. 11, no. 1, p. 014701, 2019.

[24] International Electrotechnical Commission (IEC), "Standard IEC 61724: Photovoltaic system performance monitoring – Guidelines for measurement, data exchange and analysis", IEC 61724-1, 2017.

[25] K. P. Satsangi, G. S. Sailesh Babu, D. B. Das and A. K. Saxena, "Performance Evaluation of Grid Interactive Photovoltaic System", 2018 International Conference on Computing, Power and Communication Technologies (GUCON), pp. 691-695, 2018.

[26] R. Srivastava, A. N. Tiwari, and V. K. Giri, "An overview on performance of PV plants commissioned at

different places in the world", Energy for Sustainable Development, vol. 54., pp. 51–59, 2020.

[27] K. P. Satsangi, D. B. Das, and A. K. Saxena, "Performance analysis of 40kWp solar photovoltaic plant", IEEE Region 10 Humanitarian Technology Conference 2016, (R10-HTC 2016), IEEE Publisher, doi: 10.1109/R10-HTC.2016.7906831, 2017.

[28] H. Salah Mohammed, C. Kim Gan, and K. Azmi Baharin, "Performance Evaluation of Various Solar Photovoltaic Module Technologies under Tropical Climate Conditions at Melaka, Malaysia", Journal of Engineering and Applied Sciences, vol. 14, no. 2, pp. 336–341, 2019.

[29] A. M. Humada, M. Hojabri, H. M. Hamada, F. B. Samsuri, and M. N. Ahmed, "Performance evaluation of two PV technologies (c-Si and CIS) for building integrated photovoltaic based on tropical climate condition: A case study in Malaysia", Energy and Buildings, vol. 119, pp. 233–241, 2016.

[30] D. A. Quansah, M. S. Adaramola, G. K. Appiah, and I. A. Edwin, "Performance analysis of different grid-connected solar photovoltaic (PV) system technologies with combined capacity of 20 kW located in humid tropical climate", International Journal of Hydrogen Energy, vol. 42, no. 7, pp. 4626–4635, 2017.

[31] A. K. Shukla, K. Sudhakar, and P. Baredar, "Simulation and performance analysis of 110 kWp grid-connected photovoltaic system for residential building in India: A comparative analysis of various PV technology", Energy Reports, vol. 2, pp. 82–88, 2016.

[32] J. Y. Ye, T. Reindl, A. G. Aberle, and T. M. Walsh, "Performance degradation of various PV module technologies in tropical Singapore", IEEE Journal of Photovoltaics, vol. 4, no. 5, pp. 1288–1294, 2014.

[33] S. Mekhilef, A. Safari, W. E. S. Mustaffa, R. Saidur, R. Omar, and M. A. A. Younis, "Solar energy in Malaysia: Current state and prospects", Renewable and Sustainable Energy Reviews, vol. 16, no. 1, pp. 386–396, 2012.

[34] A. Ghazali M. and A. M. Abdul Rahman, "The Performance of Three Different Solar Panels for Solar Electricity Applying Solar Tracking Device under the Malaysian Climate Condition", Energy and Environment Research, vol. 2, no. 1, 2012.

[35] L. Y. Seng, G. Lalchand, and G. M. Sow Lin, "Economical, environmental and technical analysis of building integrated photovoltaic systems in Malaysia", Energy Policy, vol. 36, no. 6, pp. 2130–2142, 2008.

[36] N. K. Kasim, H. H. Hussain, and A. N. Abed, "Performance Analysis of Grid-Connected CIGS PV Solar System and Comparison with PVsyst Simulation Program", International Journal of Smart grid, vol.3, no.4, pp. 73-79, 2019.

[37] V. Annathurai, C. K. Gan, K. A. Ibrahim, K. A. Baharin, and M. R. A. Ghani, "A review on the impact of distributed energy resources uncertainty on distribution networks", International Review of Electrical Engineering, vol. 11, no. 4, pp. 420–427, 2016.

[38] C. K. Gan, S. R. Mahmoud, K. A. Baharin, and M. H. Hairi, "Influence of single-phase solar photovoltaic systems on total harmonic distortion: A case study", Indonesian Journal of Electrical Engineering and Computer Science, vol. 12, no. 2, pp. 620–624, 2018.

[39] G. M. Lohmann, "Irradiance variability quantification and small-scale averaging in space and time: A short review", Atmosphere, vol. 9, no. 7, p. 264, 2018.

[40] G. Hamerly, "Making k-means even faster", Proceedings of the 10th SIAM International Conference on Data Mining, SDM 2010, pp. 130–140, doi: 10.1137/1.9781611972801.12, 2010.

[41] A. H. Fadaei and S. H. Khasteh, "Enhanced K-means re-clustering over dynamic networks", Expert Systems with Application, vol. 132, pp. 126–140, 2019.

[42] T. M. Kodinariya and D. P. R. Makwana, "Review on determining of cluster in K-means clustering review on determining number of cluster in K-means clustering", International Journal, vol. 1, pp. 90–95, 2016.

[43] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm", J, vol. 2, no. 2, pp. 226–235, 2019.

[44] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic", Journal of the Royal Statistical Society. Series B: Statistical Methodology, vol. 63, no. 2, pp. 411-423, 2001.

[45] S. K. Kingrani, M. Levene, and D. Zhang, "Estimating the number of clusters using diversity", Artificial Intelligence Research., vol. 7, no. 1, p. 15, 2017.

[46] J. Yang, J. Y. Lee, M. Choi, and Y. Joo, "A New Approach to Determine the Optimal Number of Clusters Based on the Gap Statistic", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12081 LNCS, pp. 227–239, 2020.

[47] K. A. Baharin, H. A. Rahman, M. Y. Hassan, C. K. Gan, and M. F. Sulaima, "Quantifying Variability for Grid-connected Photovoltaics in the Tropics for Microgrid Application", Energy Procedia, vol. 103, pp. 400–405, 2016.

[48] S. F. Zulkifli, H. Abdul, and M. Yusri, "Difference of PV Solar Farm Performance between Simulated Results with Actual Measurement under Climate Condition at Eastern Peninsular", Journal of Science and Technology, vol. 9, no. 1, pp. 1–6, 2017.