2022 International Conference on Electrical Engineering, Computer and Information Technology
University of Jember, Jember, Indonesia, November 22-23 2022

2022 International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT) | 978-1-6654-9352-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICEECIT55908.2022.10030306

K-Means Algorithm to Identify the Optimal Production Period of a Photovoltaic Array Installed at Any Point in a Very Large Area

LAMIDI Taohidi A.
Departement of Electrical Engineering
EPAC-UAC
Abomey-Calavi, Benin
alamou26@gmail.com

NOUNANGNONHOU C. Télesphore
Departement of Electrical Engineering
EPAC-UAC
Abomey-Calavi, Benin
nocteles2000@gmail.com

AGBOMAHENA B. Macaire
Departement of Electrical Engineering
EPAC-UAC
Abomey-Calavi, Benin
agbomac@yahoo.fr

DIDAVI Audace B. K.
Departement of Electrical Engineering
EPAC-UAC
Abomey-Calavi, Benin
didavia@gmail.com

***Abstract*— This work presents a technique to identify the optimal production periods of a PV generator installed at any geographical point located in a very large area. This method is based on artificial intelligence techniques of time series clustering. Irradiation, temperature and precipitation were determined as influencing parameters of the PV power. As a study area, the entire Beninese territory was chosen, where 418 geographical points were strategically selected. influencing parameters were downloaded from the NASA database from January 1st 2012 to December 31st 2021. Using the k-means clustering algorithm, irradiation, temperature and precipitation clusters were formed. We arrived at 19 irradiation clusters, 41 temperature clusters and 41 precipitation clusters. Using these clusters, 78 sub-zones were then formed within the study area by gathering points with the same irradiation, temperature and precipitation clusters. The clustering models were developed in Python 3 using the tslearn library. The Silhouette score was used to evaluate the quality of the clusters and a minimum of 0.99 was obtained within the clusters as Silhouette score. The optimal production periods depend not only on the potential of the location where it is installed, but also on the evolution over time of the parameters impacting its power output.**

Keywords— K means Algorithm - Clustering - Very Large Area - Irradiation - Temperature - Precipitation – PV Optimal Production Period.

I. INTRODUCTION

Research in the field of electrical energy has always been of particular interest due to its importance in the evolution of the modern world and its necessity. In recent years, particular emphasis is placed on research and exploitation of new energy sources, especially renewable, given the prognosis of depletion of natural resources currently exploited (gas, oil, uranium and others). These renewable energy sources (solar, wind, biomass ...) have the advantage of being more evenly distributed on the earth. The yearly production of hydroelectric, solar photovoltaic and wind energy in Africa, for example, is estimated at 1,478 TWh/year, 1,449,742 TWh/year and 978,066 TWh/year respectively [1]. But the other important remark to make beside these strong potentials is that in sub-Saharan Africa the access rate to electricity estimated by the World Bank in 2020 was 48.4% while its potential in renewable energy should allow the whole population to access it [2]. We can then ask ourselves enough questions, in particular what is slowing down the integration and the exploitation of these renewable sources to satisfy the energy needs of the populations.

Looking closely at this situation, if we put aside the financial aspect which is not negligible in the installation of production systems based on renewable sources, these sources mainly solar photovoltaic and wind have many flaws including the intermittency and random variability of their output power. For solar photovoltaic in particular, the intermittency is due to the natural movement of the earth on itself and around the sun. Its random variability is due to the variation of many weather parameters such as irradiation, temperature, wind speed. As some authors have shown in [3]–[8], the variation of the PV power is linked not only to the variations of ambient temperature but also to the cloud cover, in other words the sky clarity index. The mastery of the behavior of these weather parameters would be a big step towards a good integration of solar photovoltaic in the energy mix. But generally, these parameters vary in time and space: depending on the geographical location they do not have the same variations at the same time and vary from one moment to another. For each site where a PV system is to be installed, an appropriate study must therefore be carried out in order to master the optimal production periods but also to claim efficient forecasting of the PV power.

In this work, we propose a way to identify the optimal production periods of a PV system installed at any geographical point located within a very large area based on artificial intelligence clustering techniques.

II. METHODS

A. Study area presentation

As mentioned in the introduction, the objective of this study is to determine for any point located within a very large area, the optimal production periods of a photovoltaic system. The first methodological step is therefore to delimit an area for the study. In order to make the results realistic, we have chosen the Beninese territory with a total area of 114,763 km², whose map and solar photovoltaic potential are presented in Fig. 1.

As can be seen in Fig. 1, the overall daily irradiation over the entire territory varies between 4.8kWh/m^2 and 5.8kWh/m^2, a good range of potential that justifies the taking of the entire territory as a study area for possible electrical energy production with photovoltaics.

B. Identification of influencing parameters for the study

The control of the variability of the PV output power obviously requires the control of the weather parameters that influence it. As pointed out by many authors in their studies, the estimation of the PV power output is mainly influenced by temperature and irradiation variations, which are also influenced by other phenomena such as cloud cover, cloud movement or seasons of the year [3]–[10]. Thus, we have retained for this study as parameters, the irradiation (Ir), temperature (T), precipitation (P) and the sky clearness index (Kt).

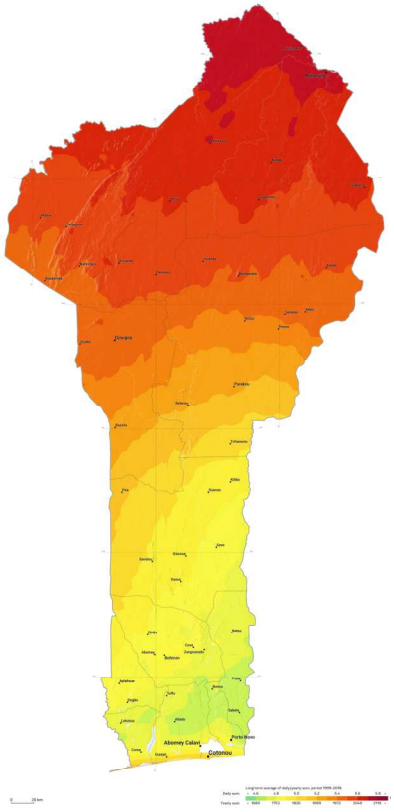

Fig. 1. Benin Global horizontal irradiation

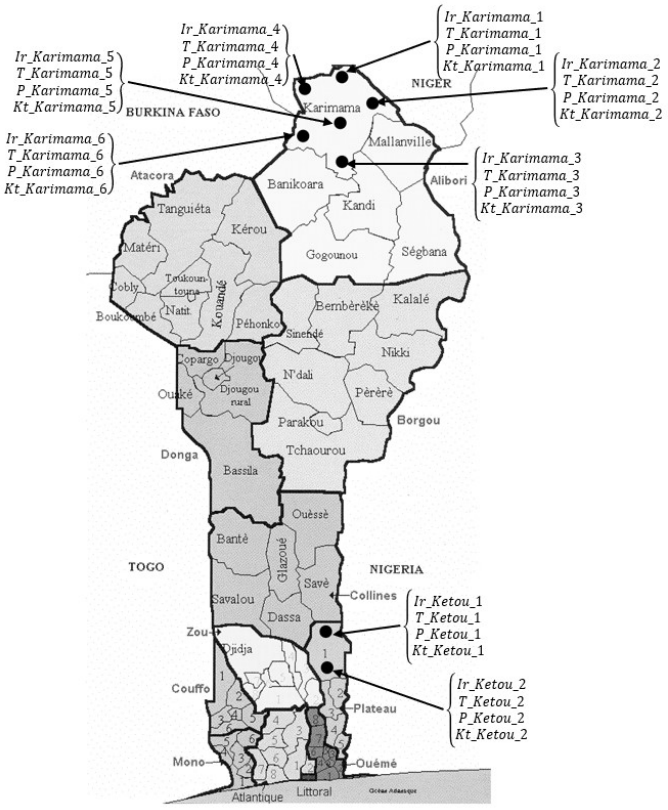

Fig. 2. Study area grid and sampling

C. Study area grid and sampling.

Once the parameters have been identified, the study area must be gridded and samples of these parameters taken. As our study area here takes into account the Beninese territory, we have chosen to keep the division according to the townships as shown in Fig. 2. Then within each township and according to their size we take a given number of samples.

D. Data and data preprocessing

Once the geographic sampling points have been defined, the data for each parameter must then be downloaded for each point. For this study, we used data from the NASA database, whose characteristics follow:

- **Database**: Power Larc NASA ;
- **Satelite**: NASA/POWER CERES/MERRA2 ;
- **Resolution**: 0.5 x 0.625 degrees (lat/lon) ;
- **Region**: 30.02 meters ;
- **Period** : January 1, 2012 to December 31, 2021 ;
- **Hourly parameters**:
  - **Ir** : All-sky surface shortwave downward irradiance (Wh/m^2) ;
  - **T** : Temperature at 2 meters (C) ;
  - **P** : Corrected precipitation (mm/hour) ;
  - **Kt** : All-sky insolation brightness index (dimensionless).

Once the data is downloaded, it must be cleaned and then pre-processed. The cleaning consists in removing the outliers, this can be done by observing them or by calculating statistical

data such as the mean, the standard deviation, the maximum or the minimum.

Once the data is cleaned, it must be brought back to the same scale so that the algorithm can interpret them in the same way. For this, two essential techniques are used: normalization and standardization. Here we opt for the normalization to an interval [a;b] whose equation is presented in (1).

$$x_{norm} = a + \frac{(x - \min(x))(b-a)}{\max(x) - \min(x)} \quad (1)$$

E. Clustering of parameters by similarity of variation

Since the idea is to be able to define the optimal production periods of a PV system installed at any geographical point located in the study area and that this production depends on the variation of the identified parameters, the next step is to develop clustering models thanks to which we will not only group these parameters by similarity of variation but also to be able to determine how these parameters will vary for another point other than the samples. The idea here is to be able to cluster the irradiations, temperatures and precipitations collected on the samples by variation and amplitude similarity. For a specific point x, silhouette coefficient s(x) allows to evaluate if this point belongs to the "good" cluster: is it close to the points of the cluster to which it belongs? Is it far from the other points? The principle of clustering is presented in Fig.3.

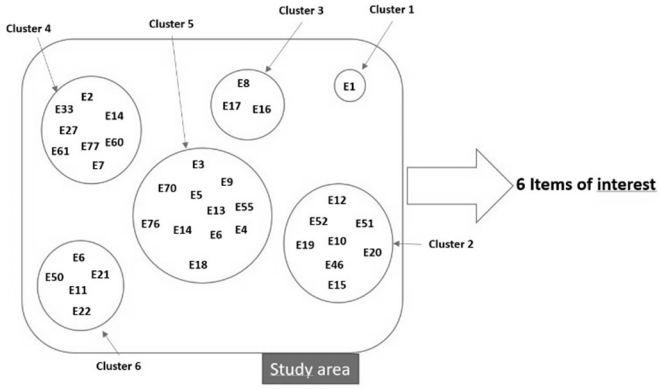

Fig. 3. Principle of clustering for one parameter

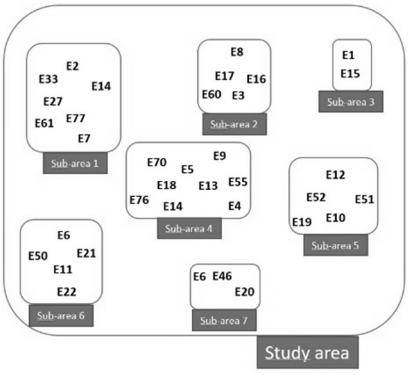

Fig. 4. Principle of sub-areas formation

As all study parameters vary over time, a time series clustering algorithm must be identified. In the literature, several clustering algorithms exist :

- Hierarchical clustering algorithms ;
- Partitioning clustering algorithms ;
- Model-based clustering algorithms ;
- Density-based clustering algorithms ;
- Hybrid clustering algorithms.

We opted for clustering with the k-means algorithm implemented with the tslearn library under Python 3. To evaluate the clustering quality, we use the silhouette score whose equation is presented in (2).

$$S(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \quad (2)$$

Where:
$$a(x) = \frac{1}{|C_k| - 1} \sum_{u \in C_k, u \neq x} d(u, x) \quad (3)$$

$$b(x) = \min_{l \neq k} \frac{1}{|C_l|} \sum_{u \in C_l} d(u, x) \quad (4)$$

a(x) represents the average distance of x from all other points of the cluster C_k to which it belongs and b(x) the smallest value that a(x) could take, if x were assigned to another cluster.

F. Optimal production periods identification

Once the irradiance, temperature, and precipitation clusters are formed, we define geographic sub-areas within the study area with the geographic points that have the same irradiance, temperature, and precipitation clusters as shown in Fig. 4.

Thus in each sub-area, irradiation, temperature and precipitation vary in the same way. In each zone, it is sufficient to select the periods for which the precipitation is low, the temperature acceptable (not too high) and the irradiation high enough for a good production. As for the clarity index, it will allow to identify within the sub-areas the non-rainy periods for which the sky is enough cloudy, this is why it is not used for the formation of clusters.

III. Results and Discussion

A. Grid and sampling results

By placing the sample points in the 77 communes of Benin, we obtained a total of 418 sample points for which irradiation, temperature and precipitation data are downloaded. Data (geographical coordinates and weather data) for all 418 samples can be found at:

https://drive.google.com/drive/folders/12upFhUx38TexXe4YWZhaIYGNAAx7IOey?usp=sharing

B. Sampling data

Temperature, irradiance and precipitation datasets are available at:

https://drive.google.com/drive/folders/1vWSgFo4AfBL9vF31e5VwG4o3WUo0r6l2?usp=sharing

Fig. 5, 6, and 7 show the headers for the main features of the three parameters for the 418 points. We can see that there are no apparent anomalies. The minimum and maximum values are within acceptable ranges for each parameter.

The mean values and standard deviations show that there are no outliers (negative values, values that are too large, or replacement values such as '-999 or NA' when the data is not available for a given time).

	count	mean	std	min	25%	50%	75%	max
Irradiation_Abomey1	87672.0	198.557816	264.166877	0.0	0.0	8.165	399.7800	956.55
Irradiation_Abomey2	87672.0	205.778458	273.153425	0.0	0.0	7.095	422.1075	957.94
Irradiation_Abomey3	87672.0	205.778458	273.153425	0.0	0.0	7.095	422.1075	957.94
Irradiation_Abomey4	87672.0	198.557816	264.166877	0.0	0.0	8.165	399.7800	956.55
Irradiation_Abomeycalavi1	87672.0	201.813452	267.974151	0.0	0.0	7.005	415.8325	953.26
...	...	...	...	...	...	...	...	...
Irradiation_Zogbodomey2	87672.0	198.557816	264.166877	0.0	0.0	8.165	399.7800	956.55
Irradiation_Zogbodomey3	87672.0	198.557816	264.166877	0.0	0.0	8.165	399.7800	956.55
Irradiation_Zogbodomey4	87672.0	193.500048	259.267531	0.0	0.0	7.395	393.0500	957.99
Irradiation_Zogbodomey5	87672.0	193.500048	259.267531	0.0	0.0	7.395	393.0500	957.99
Irradiation_Zogbodomey6	87672.0	198.557816	264.166877	0.0	0.0	8.165	399.7800	956.55

418 rows × 8 columns

Fig. 5. Header of Irradiation Data Statistics

	count	mean	std	min	25%	50%	75%	max
T2M_Abomey1	87672.0	26.619111	3.487386	12.58	23.98	25.84	29.01	40.55
T2M_Abomey2	87672.0	26.619111	3.487386	12.58	23.98	25.84	29.01	40.55
T2M_Abomey3	87672.0	26.619111	3.487386	12.58	23.98	25.84	29.01	40.55
T2M_Abomey4	87672.0	26.619111	3.487386	12.58	23.98	25.84	29.01	40.55
T2M_Abomeycalavi1	87672.0	26.971846	1.440564	20.12	25.98	27.12	28.01	31.53
...	...	...	...	...	...	...	...	...
T2M_Zogbodomey2	87672.0	26.102342	3.011694	11.40	23.85	25.52	28.34	38.68
T2M_Zogbodomey3	87672.0	26.102342	3.011694	11.40	23.85	25.52	28.34	38.68
T2M_Zogbodomey4	87672.0	26.102342	3.011694	11.40	23.85	25.52	28.34	38.68
T2M_Zogbodomey5	87672.0	26.102342	3.011694	11.40	23.85	25.52	28.34	38.68
T2M_Zogbodomey6	87672.0	26.102342	3.011694	11.40	23.85	25.52	28.34	38.68

418 rows × 8 columns

Fig. 6. Header of Temperature Data Statistics

	count	mean	std	min	25%	50%	75%	max
Precipitation_Abomey1	87672.0	0.129278	0.304607	0.0	0.01	0.04	0.14	20.12
Precipitation_Abomey2	87672.0	0.129278	0.304607	0.0	0.01	0.04	0.14	20.12
Precipitation_Abomey3	87672.0	0.129278	0.304607	0.0	0.01	0.04	0.14	20.12
Precipitation_Abomey4	87672.0	0.129278	0.304607	0.0	0.01	0.04	0.14	20.12
Precipitation_Abomeycalavi1	87672.0	0.133902	0.336261	0.0	0.01	0.05	0.16	36.02
...	...	...	...	...	...	...	...	...
Precipitation_Zogbodomey2	87672.0	0.157221	0.358340	0.0	0.01	0.06	0.18	16.98
Precipitation_Zogbodomey3	87672.0	0.157221	0.358340	0.0	0.01	0.06	0.18	16.98
Precipitation_Zogbodomey4	87672.0	0.157221	0.358340	0.0	0.01	0.06	0.18	16.98
Precipitation_Zogbodomey5	87672.0	0.157221	0.358340	0.0	0.01	0.06	0.18	16.98
Precipitation_Zogbodomey6	87672.0	0.157221	0.358340	0.0	0.01	0.06	0.18	16.98

418 rows × 8 columns

Fig. 7. Header of Precipitation Data Statistics

C. *Clustering results*

Tables I, II and III show the clustering performance as well as the number of elements in each cluster for irradiation, temperature and precipitation respectively.

TABLE I. IRRADIATION CLUSTERS AND SILHOUETTE SCORE

Cluster Irradiation N°	Silhouette score	Cluster Irradiation N°	Silhouette score
1	1	11	0,999999332
2	0,999998029	12	1
3	1	13	0,999999515
4	0,999998316	14	0,999998893
5	1	15	0,999999312
6	1	16	1
7	1	17	0,999998951
8	1	18	0
9	0,999999001	19	0
10	0,999998406		

TABLE II. TEMPERATURE CLUSTERS AND SILHOUETTE SCORE

Cluster Temperature N°	Silhouette score	Cluster Temperature N°	Silhouette score
1	1	22	0,99999837
2	0,999998855	23	1
3	1	24	1
4	0,999996828	25	1
5	0,999998966	26	0,999997622
6	0,99999851	27	1
7	1	28	0,999997109
8	0,999999072	29	1
9	1	30	0,999997523
10	1	31	1
11	0,999997817	32	1
12	1	33	0,999999178
13	0,999997131	34	1
14	0,99999891	35	0
15	1	36	1
16	0,999998879	37	0,999996682
17	1	38	0,99999605
18	0	39	1
19	0,999996967	40	0
20	0,999998296	41	0
21	0,999996755		

TABLE III. PRECIPITATION CLUSTERS AND SILHOUETTE SCORE

Cluster Precipitation N°	Silhouette score	Cluster Precipitation N°	Silhouette score
1	1	22	0,999999855
2	1	23	1
3	0,999999487	24	0,999998923
4	1	25	0,999999093
5	0,999999494	26	0,999999654
6	1	27	0,99999968
7	1	28	0,999999641
8	0,999999029	29	1
9	0,999999407	30	0,999999522
10	1	31	1
11	0,999999467	32	0
12	1	33	1
13	0,999999187	34	0,999999683
14	0,999999118	35	1
15	0,999999687	36	0,999999743
16	1	37	0,999999513
17	1	38	0
18	1	39	0
19	0,999999454	40	0,999998219
20	0,999999223	41	0
21	0,999998953		

We thus obtained 19 clusters for irradiation, 41 clusters for temperature and 41 clusters for precipitation to represent the behavior of these parameters in the study area (Beninese territory for this study). All clustering results are available at:

https://drive.google.com/drive/folders/1JsMCBaq1lQEdha_TjUAx8ZzVYzrWufNJ?usp=sharing

As we can see, the weather parameters of the elements in sub-area 3 are all found in the clusters whose curves have been plotted here.

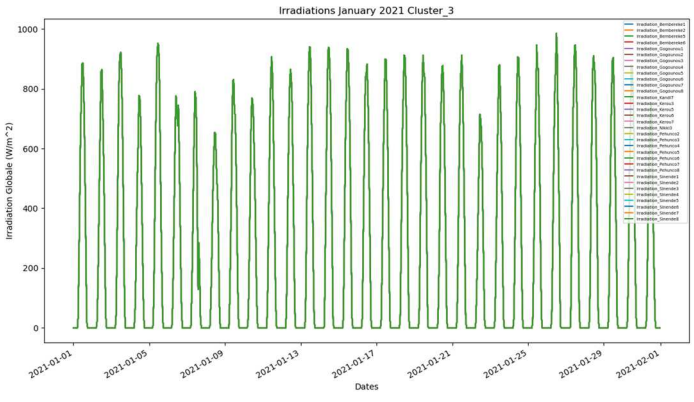

Fig. 8. Cluster 3 Irradiation for January 2021

Thus for any other point than the 418 chosen to form the clusters, the variation of its weather parameters (irradiation, temperature and precipitation) should be close to the variations within one of the formed clusters. To verify this hypothesis, we took a point with coordinates other than the 418. The performances and the clusters in which its parameters were found are presented in Table IV.

Once this verification was done, sub-areas according to the principle of Fig. 4 (elements of the same sub-area have the same irradiance, temperature and precipitation clusters) were formed. The details of the sub-area elements can be found at: https://drive.google.com/drive/folders/1LCz1H0EYVqGSsktzasTtgnsm3_nPiQBU?usp=sharing.

We obtained a total of 79 sub-zones. In the following we only present the results related to sub-area 3 due to the number of sub-areas obtained. Fig 8, 9 and 10 show the representative curves of irradiation, temperature and precipitation within the clusters of sub-area 3. Table V presents the details related to sub-area 3.

Fig. 11 shows the representative curves of monthly averages of irradiance, precipitation and clearness index.

TABLE IV. CLUSTERING TEST RESULTS

Test Point (Longitude: 2.7391, Latitude: 7.0506)					
Ir Cluster N°	Silhouette Score	T Cluster N°	Silhouette Score	P Cluster N°	Silhouette Score
4	0.99	5	0.99	12	0.99

TABLE V. SUB-AREA 1 DETAILS

Sub-area elements	Sub-area Irradiation cluster N°	Sub-area Temperature cluster N°	Sub-area Precipitation cluster N°
Kerou5, Kerou6, Kerou7, Pehunco3, Pehunco4, Pehunco6, Pehunco7, Pehunco8	3	2	26

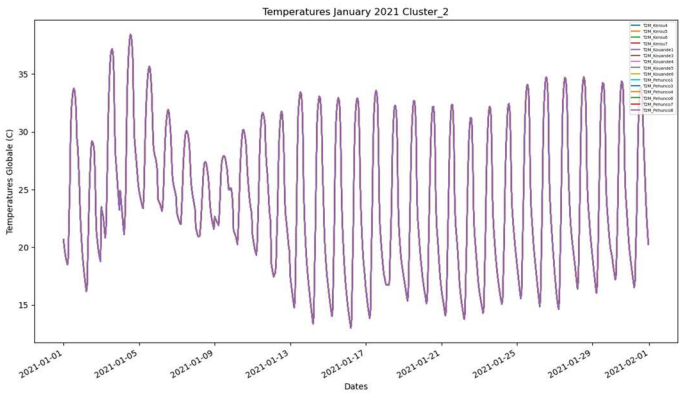

Fig. 9. Cluster 2 Temperature for January 2021

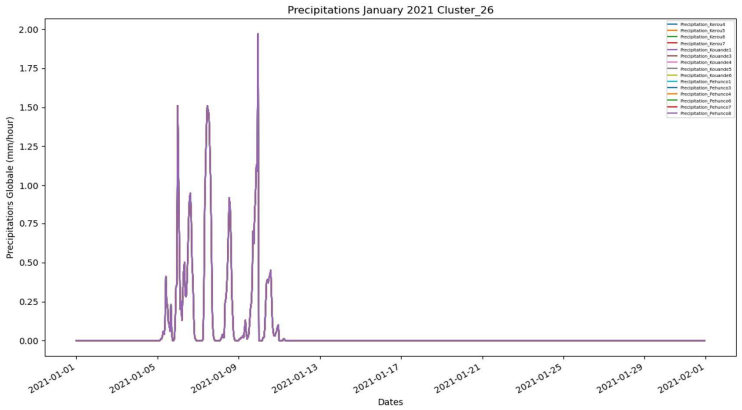

Fig. 10. Cluster 26 Precipitation for January 2021

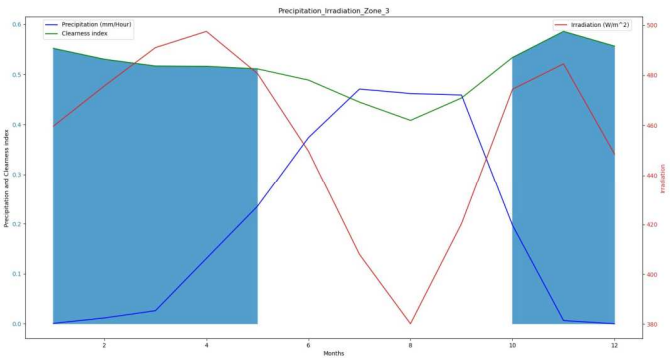

Fig. 11. Monthly averages of irradiance, precipitation and clearness index

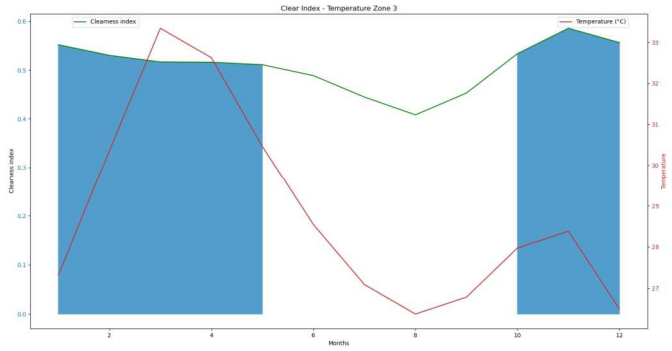

Fig. 12. Monthly averages of temperature clearness index

As we can see in Fig. 11, the values of the clearness index are high in this area only when the precipitation values are low, i.e. outside the rainy seasons. It can be concluded that the sky in this sub-area is highly cloudy only in the rainy season. We can also notice that in these periods of low precipitation, the irradiation is also very high. By setting the minimum value of the lightness index to 0.5 for a good solar production and an effective forecast, we can define the optimal production period as January to May and October to December.

Fig. 12 shows the representative curves of the monthly average temperature and the luminosity index as well as the delimited period. It can be noticed that during this period the temperature varies between 26°C and 34°C which would not affect too much the performances of most of the PV generators. Indeed, the solar panels are tested at a temperature of about 25°C and are designed to work optimally between 15°C and 35°C.

We can therefore retain for this sub-zone as the optimal production period from January to May and then from October to December. This period will be valid for all the geographical sites that have the same irradiation, temperature and precipitation clusters.

IV. Conclusion

The production performance of a PV generator depends not only on the potential of the place where it is installed but also on how the parameters influencing its output power evolve over time. In a few years, facing the increasing needs in electrical energy, the ability to find any exploitable space for the integration of renewable energy will be essential. The results of this work show that a global but precise and accurate view can be obtained by using artificial intelligence clustering techniques. We can note convincing results, in particular silhouette coefficients of at least 0.99. This assurance of a good production in a well-defined period will allow not only a good planning of the operation but also an efficient forecasting of the output power of the generator.

References

[1] IRENA, GIZ, and KFW, "The Renewable Energy Transtion in Africa," May 2021. Accessed: Aug. 11, 2022. [Online]. Available: https://africa-energy-portal.org/sites/default/files/2021-05/The%20Renewable%20Energy%20Transtion%20in%20Africa.pdf

[2] "Access to electricity (% of population) - Sub-Saharan Africa | Data." https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=ZG (accessed Sep. 10, 2022).

[3] F. H. Gandoman, F. Raeisi, and A. Ahmadi, "A literature review on estimating of PV-array hourly power under cloudy weather conditions," *Renew. Sustain. Energy Rev.*, vol. 63, pp. 579–592, 2016.

[4] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Appl. Energy*, vol. 259, p. 114216, 2020.

[5] Z. Zhen *et al.*, "SVM based cloud classification model using total sky images for PV power forecasting," in *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2015, pp. 1–5.

[6] J. Liu, W. Fang, X. Zhang, and C. Yang, "An Improved Photovoltaic Power Forecasting Model With the Assistance of Aerosol Index Data," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 434–442, Apr. 2015, doi: 10/ggznrk.

[7] J. Kou *et al.*, "Photovoltaic power forecasting based on artificial neural network and meteorological data," in *2013 IEEE International Conference of IEEE Region 10 (TENCON 2013)*, 2013, pp. 1–4.

[8] H. Aprillia, H.-T. Yang, and C.-M. Huang, "Short-term photovoltaic power forecasting using a convolutional neural network–salp swarm algorithm," *Energies*, vol. 13, no. 8, p. 1879, 2020.

[9] H. Han, L. Qu, L. Zhu, and N. Chen, "Study on the influence of PV power generation model parameter variation on power system," in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, Nov. 2017, pp. 1–5. doi: 10.1109/EI2.2017.8244421.

[10] 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," in *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, Jun. 2016, pp. 3159–3164. doi: 10.1109/PVSC.2016.7750248.