

Classification of Hourly Clearness Index of Solar Radiation in the District of Yamoussoukro

Siaman Paule Carine Yeboua^{1*}, Yao N’Goran², Kouakou Konan¹

¹Laboratory of Electricity and Energy Conversion (ECEN), Polytechnic National Institute Felix Houphouet-Boigny (INP-HB), Yamoussoukro, Côte d’Ivoire

²Laboratory of Physics of Condensed Matter and Technology (LPMCT), UFR SSMT, University Félix Houphouet-Boigny, Abidjan, Côte d’Ivoire

Email: *pauleyboua@yahoo.fr

How to cite this paper: Yeboua, S.P.C., N’Goran, Y. and Konan, K. (2019) Classification of Hourly Clearness Index of Solar Radiation in the District of Yamoussoukro. *Energy and Power Engineering*, 11, 220-231. <https://doi.org/10.4236/epe.2019.115014>

Received: April 3, 2019

Accepted: May 28, 2019

Published: May 31, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The exploitation of systems using solar energy as a source of energy is not fluctuations free because of short passage of clouds on solar radiation. The amplitude, the persistence and the frequency of these fluctuations should be analyzed with appropriate tools, instead of focusing on their location over time. The analysis of these fluctuations should use the instantaneous clearness index whose distribution is given as a first approximation which is independent not only of the season but also of the site. It is important to evaluate the potential solar energy in a region. Indeed such evaluation helps the decision-makers in their reflections on agricultural or photovoltaic solar projects. Then this study was conducted for a predictive purpose. The method used in our work combines the classification method which is the hierarchical ascending classification and two partitioning methods, the principal component analysis and the K-means method. The partitioning method enabled to achieve a number of well-known situations (in advance) that are representative of the day. The study was based on the data of a climatic weather station in the district of Yamoussoukro located in the center region of Côte d’Ivoire during the 2017 year. Using the clearness index, the study allowed the classification of the solar radiation in the region. Thus, it showed that only 346 days of the 365 days in 2017 were classified (95%). We identified three clusters of days, the cloudy sky (29%), the partly cloudy sky (32%) and the clear sky (39%). The statistical tests used for the characterization of these clusters will be detailed in a future study.

Keywords

Clearness Index, Hierarchical Clustering, Principal Component Analysis,

1. Introduction

A previous study on the climatology of the solar radiation carried out in 2011 [1] in Côte d'Ivoire showed the importance of knowing the solar radiation arriving on the ground to better understand the performance of solar energy systems. However, the solar radiation measured on the ground is not fluctuations free because of short passage of clouds on solar radiation [2]. This intermittency of solar radiation is one of the major disadvantages of the large-scale application of solar photovoltaic (PV) and other forms of solar energy. Thus, short fluctuations in solar irradiance can lead to unpredictable variations in voltage and power if the electricity produced is to be injected into the electrical network [3]. In addition, in order to provide adequate measures to reinforce the network over time while avoiding overly cautious and costly measures, network operators need tools for a realistic estimation of these disturbances [4].

In solar energy applications, the analysis of these fluctuations should focus on instantaneous or hourly clearness index [5]. The probability distribution for an average clearness index was achieved as a first approximation which is independent not only of the season but also of the site. Therefore, for effective dimensioning of energy conversion systems and for a predictive purpose, it appears important to characterize the solar energy resource. This characterization is justified by the fact that hourly and daily solar irradiation data do not take into account the fluctuations of local weather conditions. For this, it is necessary to classify the different states of the sky based on the clearness index. The clearness index as a classification criterion has been described in previous studies [6] [7], namely by the use of the fractal dimension and a mixture of Dirichlet distributions. The method used in our work combines several classification methods as the factorial method (PCA) having an exploratory purpose and reducing data, the supervised classification method (HCA) and the partitioning method (K-means). Generally, the authors used a single classification method [8] [9]. In our work, the combination of several methods allowed to classify with high accuracy the different states of sky. This article presents the results of the classification of hourly clearness index of solar radiation obtained by measurements made in 2017. The data are from one of the weather-climate observation network stations and solar monitoring in Côte d'Ivoire (ROSSCI) located in Yamoussoukro.

2. Data and Methods

2.1. Constitution of Database

The data of the weather-climate station are recorded in steps of one minute 24h/24h during the 365 days [1] of the year 2017. Eight (08) climatic weather parameters (Temperature, Relative humidity, Barometric pressure, Rainfall,

Speed and Direction wind, solar and UV radiation) are thus measured and others are calculated from them. A total of 12 text files (xxxx.txt) representing the 12 months of 2017 are saved and constitute the initial database (IDB) in text format (xxxx.txt). But only the values of the instantaneous global solar irradiance (E_i in W/m^2) are exclusively exploited within the framework of this study. Also, a pretreatment of the data of this database allows a reorganization into an organized database (ODB) and transferred in Excel format (xxxx.xls) corresponding to 525,600 values ($60 \times 24 \times 365$) for the year. Then, from the BDO database, the values of the global instantaneous solar irradiance from 6 am to 5 pm are extracted using a custom filter under the Excel software, to give a new solar database (SDB). Thus, this SDB contains only 12 hours of recordings per day, a total of 262,800 values ($60 \times 12 \times 365$) of the instantaneous global solar irradiance for the year 2017 to be processed. But this number is not reached because there are missing or lost data in the IDB. These data losses are due to connection problems, power cuts or software transmission errors.

From the solar database (SDB), an array of 4380 values (12×365) of hourly global solar radiation (IGH Wh/m^2) is obtained by integrating the measured values of instantaneous global solar irradiance (E_i in W/m^2) over an hour according to the following formula:

$$IGH = \frac{\sum_{i=1}^{60} EG_i}{60} \text{ (Wh/m}^2\text{)} \quad (1)$$

Then, Similarly, a table of 4380 values of hourly global extraterrestrial radiation (IGH_0 Wh/m^2) is calculated according to formula [10] as follows:

$$IGH_0 = I_0 \times \left(\frac{r}{r_0}\right)^2 \times \sin(h) \quad (2)$$

with

- the solar constant $I_0 = 1367 \text{ W} \cdot \text{m}^{-2}$ [11]
- Earth-sun distance given by the expression [12]:

$$\left(\frac{r}{r_0}\right)^2 = 1 + 0.03344 \times \cos\left(\frac{360 \cdot j}{365.2} - 0.0049\right) \quad (3)$$

j is the day number in the year from 1 to 365 in a calendar year;

- solar height h [13]:

$$\sinh = \sin \theta \times \sin \delta + \cos \theta \times \cos \delta \times \cos \omega \quad (4)$$

with

- the latitude $\theta = 6.8692^\circ$;
- the declination δ of the place expressed in degrees:

$$\delta = 23.45 \times \sin\left(\frac{360(284 + j)}{365}\right) \quad (5)$$

- the hour angle ω expressed in degrees:

$$\omega = 15 \times (TST - 12) \quad (6)$$

- The expression of ω contains the true solar time (TST):

$$TST = LST + TE \quad (7)$$

where LST is the local solar time and TE , the time equation.

LST is given by [14]:

$$LST = UT + \frac{\varphi}{15} \quad (8)$$

φ is the longitude of the place, $\varphi = -5.2396^\circ$, and UT is the universal time expressed in hours.

TE is given by [15]:

$$TE = 9.87 \times \sin(2B) - 7.35 \times \cos(B) - 1.5 \times \sin(B) \quad (9)$$

where

$$B = 360 \frac{j-81}{365} \quad (10)$$

The calculation is developed in an algorithm in MATLAB software and the results are arranged as arrays.

Knowledge of hourly global solar radiation (IGH Wh/m²) and hourly global extraterrestrial radiation (IGH_0 Wh/m²) allows to determine the clearness index (K_t) which is defined as the ratio between arriving IGH on the earth's surface and IGH_0 following formula [16] [17]:

$$K_t = \frac{IGH}{IGH_0} \quad (11)$$

Finally, the results of the calculation of hourly clearness index profiles (4380 values) are summarized in the form of boxplot shown in **Figure 1**. It is noted that the values of the clearness index profiles are between 0 and 1 from 7 am to 5

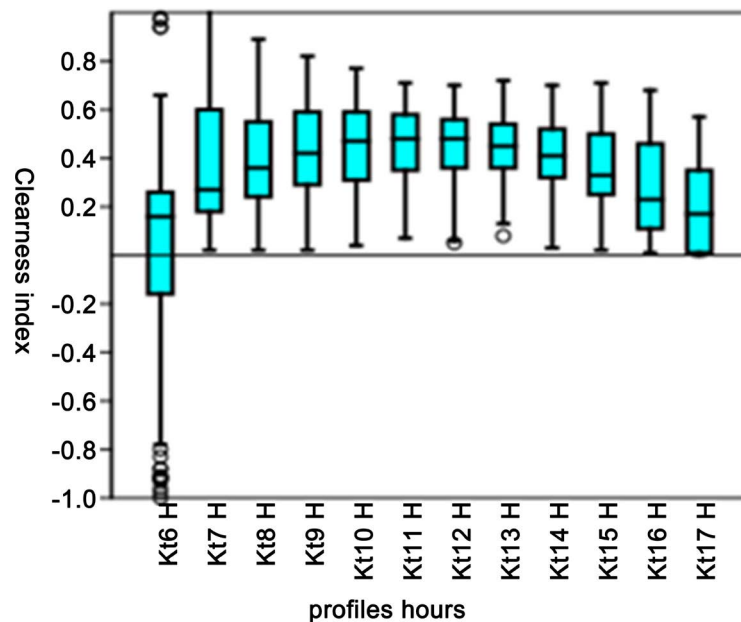


Figure 1. Boxplot of calculated hourly clearness index profiles.

pm with two outliers (bias) at the profiles 12 pm and 1 pm. For 6 am profile, the values are between -1 and 0.7 with more outliers that is to say remote data values of other data values. The overall analysis of boxplots illustrates asymmetric clearness index data. These are hourly clearness index values (K_t) to be used in the classification.

2.2. Methods

Since the accuracy of the measurements is fine and that the number n observation is large ($365 \times 12 = 4380$), it may happen that the number of distinct values observed are relatively high. The observed distribution univariate (DO1) that flows from this:

- A large number of lines in the staffing table;
- Many low amplitude workforces.

This situation does not allow to easily identify the essential characteristics of the observed distribution. One solution to this problem is to adopt a more comprehensive approach to data by carrying a group of the latter, that is to say by bringing together in one category (called a class) values observed relatively close to each other.

In the classification process, two approaches are operated: (i) saying nonparametric approaches (hierarchical clustering, mobile centers method); and (ii) the so-called probabilistic approaches (clustering) using an assumption about the distribution of individuals to classify. The purpose of classification is to group the data defined by a set of variables in homogeneous and distinct clusters together. To classify must: 1) perform a principal component analysis (PCA) on the whole data base containing the values of hourly clearness index profiles 2) retain the main components that explain a significant proportion of variance, 3) applying a hierarchical clustering using euclidean distances and 4) consolidation of clusters by the k-means method for better partition. The k-means method requires to know first knowledge of the number of classes to be determined [18]. In order to know this number, we used previously the PCA to reduce the number of variables and to visualize as much as possible on a plane the observations described by several variables. As the PCA does not distinguish enough the class number, we applied the ascending hierarchical classification using Ward's method. This procedure, presented in **Figure 2**, is implemented in the FactomineR

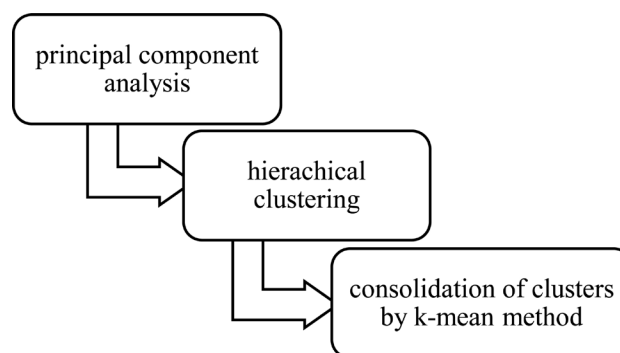


Figure 2. Classification process.

package of the R software [19].

3. Results and Discussion

3.1. Classification Results

3.1.1. Principal Component Analysis

1) Eigenvalues and percentage of variance

The PCA helped to highlight the affinities between different hours (variables) and to deduct distributions of clearness index profiles during the year. The first three components express 81.51% of the total variance, with 45.23% for the first factor, 30.00% for the second factor and 6.27% for the third factor. The analysis is restricted to the first two factors, the eigenvalues greater than unity, say more than 75% of the initial variance. **Table 1** shows the values, the percentage of variance explained and that of the cumulative variance each axis.

2) Main variables correlated-component

The correlation matrix (**Table 2**) shows that the axis I is very well correlated positively to clearness index profiles from 7 am to 12 pm (Kt7h to Kt12h) and negatively with the profiles of 3 pm to 5 pm (Kt15h to Kt17h). The axis II by cons, has very good positive correlation with the clearness index profiles from 11 am to 5 pm (Kt11h to Kt17h) (**Table 2**).

3) Results of variables and days graphs (PCA)

The correlation circle shows that the axis I positively door profiles from 7 am

Table 1. Own values and percentage of variance obtained from PCA.

Axis	Own values	Variance (%)	Cumulative variance (%)
I	4.98	45.23	45.23
II	3.30	30.00	75.23
III	0.69	6.27	81.51

Table 2. Correlations between the variables and main axis.

Variables	axis I	axis II
Kt7h	0.89	-0.08
Kt8h	0.89	0.11
Kt9h	0.86	0.23
Kt10h	0.81	0.34
Kt11h	0.68	0.49
Kt12h	0.54	0.62
Kt13h	0.15	0.80
Kt14h	-0.21	0.81
Kt15h	-0.48	0.75
Kt16h	-0.65	0.62
Kt17h	-0.72	0.50

to 1 pm and negatively profiles from 2 pm to 5 pm, this reflects a contrast between these profiles and analyzing the averages of each profile, we see that the profiles of 7 am to 1 pm have relatively higher average than those from 2 pm to 5 pm. However, the axis 2 carries positively all clearness index profiles except 7 am profile (Figure 3(a)).

The plan of statistical units (days), highlights the spatial distributions of individuals (days) appears to favor three (3) groups. (Figure 3(b)).

3.1.2. Results of Hierarchical Clustering Analysis (HCA)

The hierarchical clustering analysis (HCA), by calculation of euclidean distances between individuals provides a dendrogram accompanied by inertial gains graph (Figure 4). From this dendrogram with the cutoff level proposed by the FactoMineR

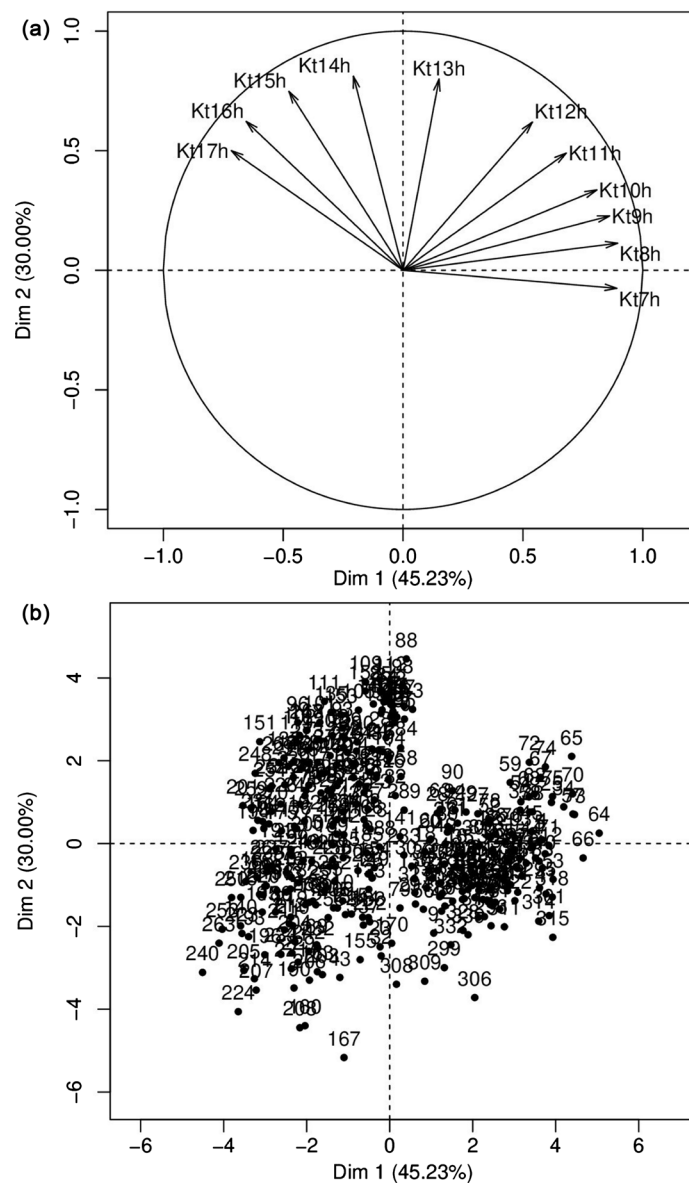


Figure 3. Outputs PCA: projection variables (profiles hours) on the factorial axis I and II (a) showing individuals (days) on the factorial axis I and II (b).

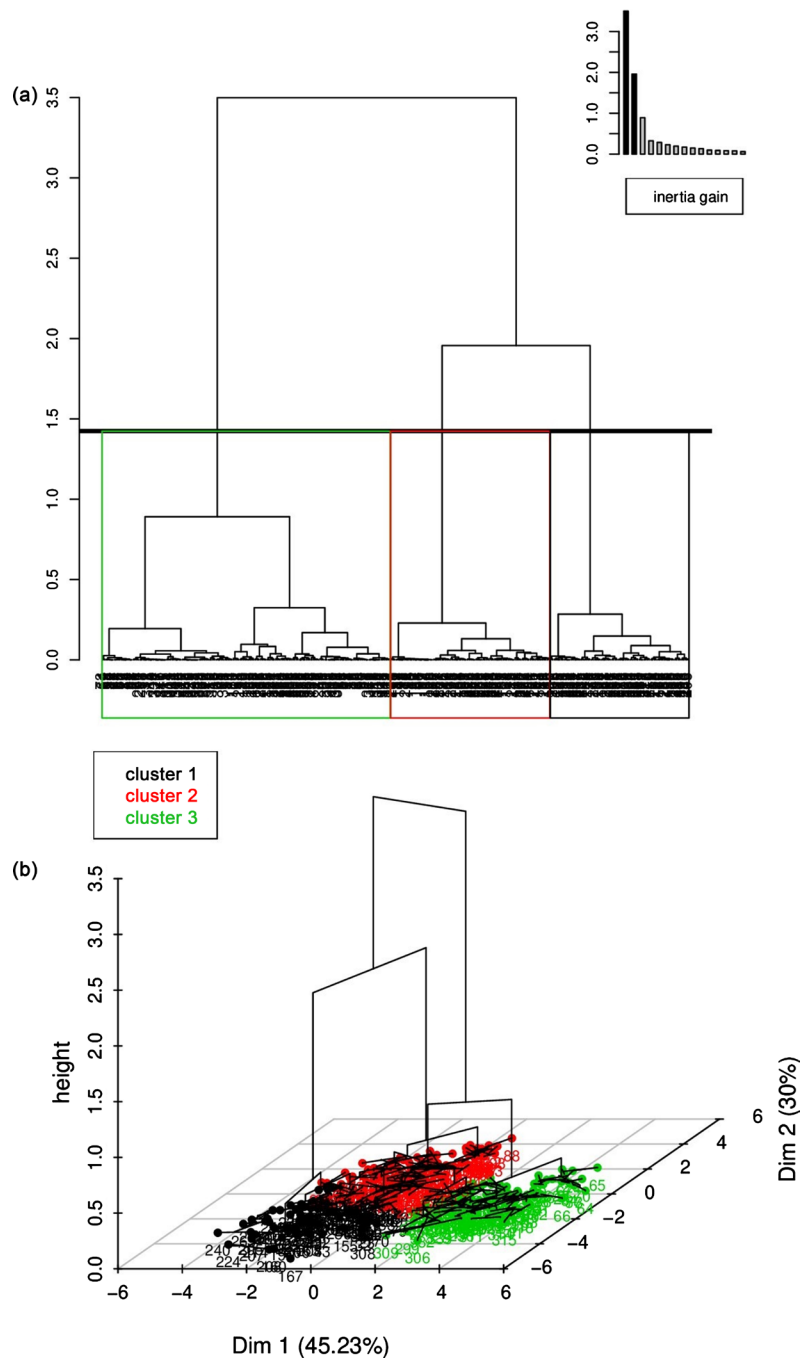


Figure 4. Dendrogram showing the different class.

package, there are three clusters of individuals close to each other (**Figure 4(a)**):

- class (or cluster) 1 Black
- class (or cluster) 2 red
- class (or cluster) 3 green

Thus, class 1 is more distant from class 2 which is more distant from class 3.

This dendrogram is then projected in 3D on the factorial axes of the PCA for a better visualization. In this graph (**Figure 4(b)**), individuals (days) are colored according to their membership in each class.

3.1.3. Consolidation of Clusters by K-Means

This procedure led to the production of a map of days based on their membership in each cluster. This map is shown in the factorial axes (Figure 5). On this map we distinguish three groups divided into functions of the similarities of each individual. Thus, the individuals present in cluster 1, in black, possess sufficiently similar characteristics to be considered as a single individual and thus forming a single group. It is the same for cluster 2, in red and 3 in green.

3.2. Discussion

The results of calculations showed asymmetric data values of clearness index. Indeed, this asymmetry indicates that the data may not be normally distributed throughout the year. These results show that the values of global extraterritorial radiation (IGH_0) are negative at 6 am. In fact, in the calculations we used the Equation of Time (ET), or it is related to the timekeeping of the sun [10]. Indeed, as the sun makes a complete revolution around the Earth in 24 hours observer, the actual time is just after the angle of the sun with a fixed direction. According [20], if we consider the sun as a clock, it begs the question of precision timekeeping. This chronometer error can be directly observed by noting the position of the sun in the sky in the same calendar time for each day of the year. Also, the atmosphere, the distribution of extraterrestrial radiation is not homogeneous because the earth orbits around the sun and that the inclination of the equatorial plane relative to the orbital plane varies with latitude and seasons [12]. Thus, the negative values of 6 am IGH_0 throughout the year are due to the timekeeping of the sun and that the extraterrestrial radiation at 6 am is skimming the surface. These negative IGH_0 values from 6 am have influenced the 6 am values

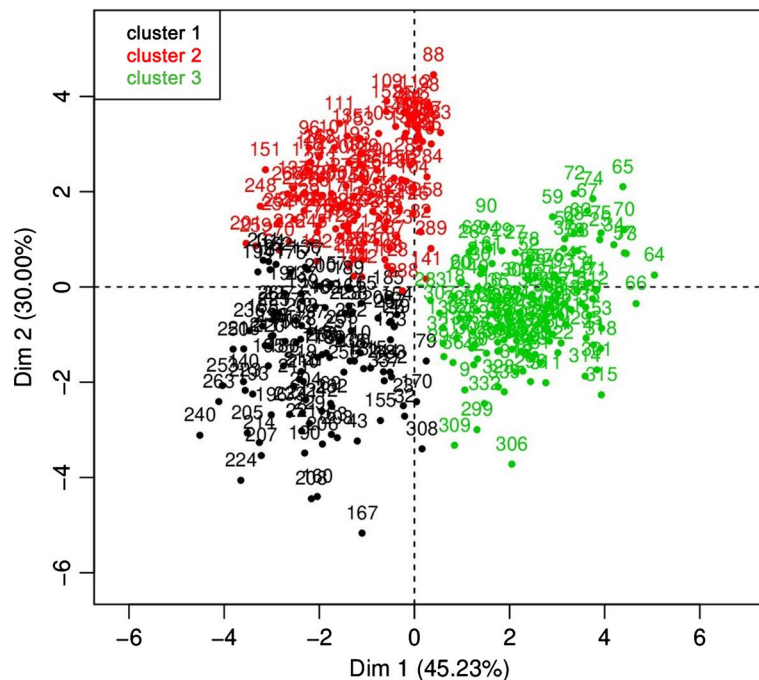


Figure 5. Map of individuals (days) colored according to their membership in each class.

of clearness index profiles, this prompted us to carry out the classification from 7 am to 5 pm by eliminating values from 6 am profile.

The PCA results showed three groups of days but does not distinguish accurately the intrinsic characteristics of the different combinations. Indeed, the importance of this step (PCA) is to rigorously perform and display projections in orthogonal planes [21].

The hierarchical clustering provided three clusters with the level of the proposed cuts dendrogram. The dendrogram resulting from this classification allows to examine successive aggregations of all individuals and visualize the connections between them [21]. Looking closely at the 3D projection of the dendrogram on the dimensions of the PCA, we see that the cluster 2 and 3 contain tiny parts of the cluster 1. This classification noise is due to the hierarchy between individuals.

The projection clusters on PCA's size allows us to accurately distinguish the membership of individuals (days) for each class. In fact the K-means method disregarding the individuals level of aggregation to better visualize and emphasize the similarity between individuals [12].

4. Conclusions

This study has classified solar radiation in the district of Yamoussoukro in 2017 with the hourly clearness index. In this study, this index was established as the ratio of hourly global solar radiation (IGH) measured and hourly global extra-terrestrial irradiation (IGH_0) calculated for 2017.

Thus, it is clear from this study that the 365 days of the year 2017, 346 days were classified (95%). In this classification we could conclude that the district of Yamoussoukro is characterized by three times of clusters, namely cloudy, partly cloudy and clear sky with clear weather dominance (39%). The cloudy and partly cloudy sky occupy respectively 29% and 32% of the time during the year.

It is necessary to know the state of the sky in a region. Indeed this knowledge contributes in an important part of the solar system design process. The results obtained show that the district of Yamoussoukro could be a favorable area to the exploitation of solar energy systems.

In order to have an excellent forecasting tool at the national level, both in the energy and climatological fields, we are in the process of extending this classification to all weather and climate stations in network stations and solar monitoring in Côte d'Ivoire (ROSSCI).

Acknowledgements

We would like to thank the State of Côte d'Ivoire for financing by the PreSeD-CI and AMRUGE-CI projects through the debt reduction contract. We also thank the IRD and the France embassy in Côte d'Ivoire for the management of the financing obtained. At the same time, we would like to thank the Polytechnic National Institute Felix Houphouet-Boigny (INP-HB) Yamoussoukro, Côte d'Ivoire for

the working site.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] N'goran, Y. (2011) Establishing of West Africa Solar Radiation Atlas: 1-Preliminary Study of Solar Radiation Climatology in Ivory Coast. *ICPSR Journal ISESCO Science and Technology Vision*, **11**, 11-19.
https://www.isesco.org.ma/ISESCO_Technology_Vision/NUM11/doc/2.pdf
https://www.isesco.org.ma/ISESCO_Technology_Vision/NUM11/content.html
- [2] Woyte, A., Belmans, R. and Nijs, J. (2007) Fluctuations in Instantaneous Clearness Index: Analysis and Statistics. *Solar Energy*, **2**, 195-206.
<https://www.sciencedirect.com/science/article/pii/S0038092X0600079X>
<https://doi.org/10.1016/j.solener.2006.03.001>
- [3] Woyte, A., Thong, V.V., Belmans, R. and Nijs, J.C. (2006) Voltage Fluctuations on Distribution Level Introduced by Photovoltaic Systems. *IEEE Transactions on Energy Conversion*, **21**, 202-209. <https://doi.org/10.1109/TEC.2005.845454>
<https://ieeexplore.ieee.org/abstract/document/1597338>
- [4] M'Raoui, A., Mouhous, S., Malek, A. and Benyoucef, B. (2011) Etude statistique du rayonnement solaire à Alger. *Revue des Energies Renouvelables*, **14**, 637-648.
http://www.cder.dz/download/Art14-4_7.pdf
- [5] Li, D.H. and Lam, J.C. (2001) An Analysis of Climatic Parameters and Sky Condition Classification. *Building and Environment*, **36**, 435-445.
<https://www.sciencedirect.com/science/article/pii/S036013230000275>
[https://doi.org/10.1016/S0360-1323\(00\)00027-5](https://doi.org/10.1016/S0360-1323(00)00027-5)
- [6] Harrouni, S., Guessoum, A. and Maafi, A. (2005) Classification of Daily Solar Irradiation by Fractional Analysis of 10-Min-Means of Solar Irradiance. *Theoretical and Applied Climatology*, **80**, 27-36. <https://doi.org/10.1007/s00704-004-0085-0>
<https://link.springer.com/article/10.1007/s00704-004-0085-0>
- [7] Soubdhan, T., Emilion, R. and Calif, R. (2009) Classification of Daily Solar Radiation Distributions Using a Mixture of Dirichlet Distributions. *Solar Energy*, **83**, 1056-1063. <https://www.sciencedirect.com/science/article/pii/S0038092X09000073>
<https://doi.org/10.1016/j.solener.2009.01.010>
- [8] Muselli, M., Poggi, P., Notton, G. and Louche, A. (2000) Classification of Typical Meteorological Days from Global Irradiation Records and Comparison between Two Mediterranean Coastal Sites in Corsica Island. *Energy Conversion and Management*, **41**, 1043-1063. [https://doi.org/10.1016/S0196-8904\(99\)00139-9](https://doi.org/10.1016/S0196-8904(99)00139-9)
<https://www.sciencedirect.com/science/article/pii/S0196890499001399>
- [9] Tejera, S.M., Pérez, M.A.S., Santigosa, L.R. and Bravo, I.L. (2017) Classification of Days According to DNI Profiles Using Clustering Techniques. *Solar Energy*, **146**, 319-333. <https://www.sciencedirect.com/science/article/pii/S0038092X1730124X>
<https://doi.org/10.1016/j.solener.2017.02.031>
- [10] Voyant, C. (2011) Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels. Thèse de Doctorat, Université Pascal Paoli, Corse.
<https://tel.archives-ouvertes.fr/tel-00635298>

- [11] Dambreville, R. (2014) Prédiction du rayonnement solaire global par télédétection pour la gestion de la production d'énergie. Thèse de Doctorat, Université de Grenoble, Grenoble. <https://tel.archives-ouvertes.fr/tel-01130251>
- [12] Oumbe, A. (2009) Exploitation des nouvelles capacités d'observation de la Terre pour évaluer le rayonnement solaire incident au sol. Thèse de doctorat, Ecole Nationale Supérieure des Mines de Paris, Paris. https://pastel.archives-ouvertes.fr/file/index/docid/501384/filename/these_oumbe.pdf
- [13] Duffie, J.A. and Beckman, W.A. (2013) Solar Engineering of Thermal Processes. John Wiley & Sons, Inc., Hoboken. <https://books.google.fr/books>
<https://doi.org/10.1002/9781118671603>
- [14] Ricaud, A. (2011) Gisement solaire et transferts énergétiques. Thèse de doctorat, Université de Cergy-Pontoise, Cergy-Pontoise.
- [15] Mouhous-Chaouchi, S. (2012) Etude statistique du rayonnement solaire sur un plan incliné. Thèse de Doctorat, Université Abou Berk Bouzaréah, Algérie. <http://dSPACE.univ-tlemcen.dz/handle/112/1994>
- [16] Mellit, A., Kalogirou, S.A., Shaari, S., Salhi, H. and Arab, A.H. (2008) Methodology for Predicting Sequences of Mean Monthly Clearness Index and Daily Solar Radiation Data in Remote Areas: Application for Sizing a Stand-Alone PV System. *Renewable Energy*, **7**, 1570-1590. <https://doi.org/10.1016/j.renene.2007.08.006>
<https://www.sciencedirect.com/science/article/pii/S0960148107002662>
- [17] Linguet, L. (2016) De la modélisation du rayonnement solaire à la production d'énergie: Recherches sur l'optimisation de la production photovoltaïque en contexte amazonien. Thèse de Doctorat, Université de Guyane, Cayenne. <https://tel.archives-ouvertes.fr/tel-01359819>
- [18] Boubou, M. (2007) Contribution aux méthodes de classification non supervisée via des approches prétopologiques et d'agrégation d'opinion. Thèse de Doctorat, Université Claude Bernard-Lyon I, Lyon. <https://tel.archives-ouvertes.fr/tel-00195779>
- [19] Lê, S., Josse, J. and Husson, F. (2008) FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, **25**, 2-18. http://factominer.free.fr/more/article_FactoMineR.pdf
<https://doi.org/10.18637/jss.v025.i01>
- [20] Kudish, A.I. and Ianetz, A. (1996) Analysis of Daily Clearness Index, Global and Beam Radiation for Beer Sheva, Israel: Partition According to Day Type and Statistical Analysis. *Energy Conversion and Management*, **4**, 405-416. <https://hal.archives-ouvertes.fr/hal-00876254>
[https://doi.org/10.1016/0196-8904\(95\)00193-X](https://doi.org/10.1016/0196-8904(95)00193-X)
- [21] Timmerman, M.E., Ceulemans, E., Kiers, H.A. and Vichi, M. (2010) Factorial and Reduced K-Means Reconsidered. *Computational Statistics and Data Analysis*, **7**, 1858-1871. <https://www.sciencedirect.com/science/article/pii/S0167947310000745>
<https://doi.org/10.1016/j.csda.2010.02.009>